HABITAT AND DISTRIBUTION MODELS OF MARINE AND ESTUARINE SPECIES: ADVANCES FOR A SUSTAINABLE FUTURE

EDITED BY: Mary C. Fabrizio, Mark J. Henderson, Kenneth Alan Rose and Pierre Petitgas

PUBLISHED IN: Frontiers in Marine Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-692-9
DOI 10.3389/978-2-83250-692-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

HABITAT AND DISTRIBUTION MODELS OF MARINE AND ESTUARINE SPECIES: ADVANCES FOR A SUSTAINABLE FUTURE

Topic Editors:
Mary C. Fabrizio, Virginia Institute of Marine Science, William & Mary Gloucester Point, United States
Mark J. Henderson, U.S. Geological Survey, United States
Kenneth Alan Rose, University of Maryland, College Park, United States
Pierre Petitgas, Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), France

Citation: Fabrizio, M. C., Henderson, M. J., Rose, K. A., Petitgas, P., eds. (2022). Habitat and Distribution Models of Marine and Estuarine Species: Advances for a Sustainable Future. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-692-9

Table of Contents

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 20 October 2022
DOI 10.3389/fmars.2022.1050548

OPEN ACCESS

EDITED AND REVIEWED BY
Laura Airoldi,
University of Padova Chioggia Hydrobiological Station, Italy

*CORRESPONDENCE
Mary C. Fabrizio
mfabrizio@vims.edu

SPECIALTY SECTION
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

RECEIVED 21 September 2022
ACCEPTED 07 October 2022
PUBLISHED 20 October 2022

CITATION
Fabrizio MC, Henderson MJ, Rose K and Petitgas P (2022) Editorial: Habitat and distribution models of marine and estuarine species: Advances for a sustainable future.
Front. Mar. Sci. 9:1050548.
doi: 10.3389/fmars.2022.1050548

Editorial: Habitat and distribution models of marine and estuarine species: Advances for a sustainable future

Mary C. Fabrizio[1]*, Mark J. Henderson[2], Kenneth Rose[3] and Pierre Petitgas[4]

[1]Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, United States, [2]U.S. Geological Survey, Vermont Cooperative Fish and Wildlife Research Unit, Rubenstein School of the Environment and Natural Resources, University of Vermont, Burlington, VT, United States, [3]University of Maryland Center for Environmental Science, Horn Point Laboratory, Cambridge, MD, United States, [4]French Research Institute for Exploitation of the Sea (IFREMER), Department of Biological Resources and Environment (Dept. RBE), Centre Atlantique, rue de l'Île d'Yeu, Nantes, France

KEYWORDS

SDM, habitat, ecology, spatial, models, aquatic

Editorial on the Research Topic
Habitat and distribution models of marine and estuarine species: Advances for a sustainable future

The physical and biological characterization of suitable habitats and species-specific models to estimate their extent are valuable for conservation and fisheries management. As exploited species and habitats face challenges from anthropogenic influences, such as fishing and climate change, the identification and protection of habitats becomes increasingly important. Most of the papers within this special topic issue used some form of species distribution model (SDM) to identify habitats used by fishes (Asch et al.; Crear et al.; Fabrizio et al.; Freidland et al.; Zydlewski et al.), marine mammals (Astarloa et al.), nearshore invertebrates (Cristiani et al.; Behan et al.), or deep-sea communities (Bowden et al.; Saunders et al.). A few papers focused on developing methods to better describe habitats (Griffin et al.; Henderson et al.; Cecino et al.), while other papers investigated model performance and incorporation of new statistical methods to improve model accuracy (Asch et al.; Behan et al.; Bowden et al.). Below we provide a synthesis of these papers under the topics of data sources used for analyses, statistical methods, stationarity and model performance, connectivity, and management implications; we conclude with a consideration of opportunities for advancing this field of study.

Data sources used for analyses

Most SDMs used presence/absence information to describe relationships between taxa and habitat features; only a few SDMs were informed by estimates of relative

abundance, density, or biomass (Astarloa et al.; Behan et al.; Fabrizio et al.; Freidland et al.). Occurrence and abundance information were commonly derived from stratified random surveys using bottom trawls, dredges, or epibenthic sleds. Transect sampling with ROV video and still cameras was used in three papers, suggesting that new methods to automate image analysis allow researchers to obtain rich datasets from habitats that are otherwise poorly sampled. Two papers in this issue used acoustic telemetry (Crear et al.; Griffin et al.) to develop habitat models for marine fishes.

Data sources to describe spatio-temporal variation in environmental conditions comprised outputs from a variety of numerical models, including ocean circulation models, coupled physical-biogeochemical models, and earth system models (Asch et al.; Behan et al.; Cecino et al.; Crear et al.; Cristiani et al.; Fabrizio et al.; Henderson et al.). A few of these authors noted the need to rescale (simplify or summarize) biotic data to be consistent with the spatial resolution of oceanographic models (Asch et al.; Cecino et al.; Cristiani et al.). Physico-chemical models were also used to project habitat conditions under one or more future climate scenarios that typically included the 'status quo.' Resource-specific attributes (e.g., life history, distribution, length of time series, and so forth) appeared to influence the choice of projection years, as these varied among studies. Remotely sensed environmental conditions and chlorophyll-*a* concentrations were also incorporated into habitat models (Freidland et al.). In most papers, however, SDMs were often informed by readily available habitat descriptors (e.g., depth, temperature, salinity, dissolved oxygen) that were typically collected at the time of sampling of the biota. Biotic variables, such as prey abundance and primary productivity, were sometimes used to describe habitats (Asch et al.; Astarloa et al.; Freidland et al.). When relatively long time series (> 35 years) of observations of marine biota were available, atmospheric indices such as the PDO, AMO, and NAO were considered as climate indicators (Asch et al.; Astarloa et al.).

Statistical methods

Statistical modelling corresponds to an ensemble of steps that considers the sampling design, the covariates and their quality, the model type, and the fitting procedure. All these aspects were addressed by the papers in this special topic issue. In particular, presence-only data are a problem for model fitting (Winship et al.) so analysts are often forced to use pseudo-absences generated by random resampling across the area (Griffin et al.). Zero-inflated data may require hurdle models to estimate the probability of occurrence and the conditional positive catch component (Astarloa et al.; Lowman et al.).

This collection of papers exemplifies the now widely accepted use of machine learning procedures that go beyond classical statistical methods. In particular, random forest approaches and regression trees (boosted by an iterative procedure) were used to identify covariates and their ranges of values and interactions that best explained observed species distributions (Cecino et al.; Fabrizio et al.; Freidland et al.; Griffin et al.; Henderson et al.; Saunders et al.). Suitable ranges for covariates can be envisaged using physiological principles (Crear et al.) or inferred from a histogram approach (Fabrizio et al.). Approaches to model species occurrence probability included logistic regression (Henderson et al.), resource selection functions (Griffin et al.), generalized linear or additive models, and generalized linear mixed-effects models (GLMM; Astarloa et al.; Lowman et al.). GLMMs were used to model space-time interactions and thus evidence of change in spatial distributions across time.

Multiple papers used recently developed numerical computing methods to efficiently incorporate spatial and temporal autocorrelation with a Bayesian approach. For example, two papers used Vector Autoregressive Spatio-Temporal (VAST) models to examine changes in animal densities and distributions through time (Astarloa et al.; Lowman et al.). VAST models treat the space-time correlation function as a time-varying spatial component so the remaining variance can be estimated as a fixed effect (Thorson and Barnett 2017). A similar approach was used to account for spatial autocorrelation in an analysis of associations between fish and deep-sea corals (Henderson et al.).

Stationarity and model performance

Because SDMs are based on correlative relationships between species occurrence and covariates, model projections outside the range of values, time frame, or area used to fit the SDM must assume that the correlative relationships are valid under conditions that were not considered in model fitting. Thus, stationarity of relationships is a critical assumption when making projections under future climate scenarios. Asch et al. questioned the temporal stationarity assumption by fitting SDMs for multiple time periods and Behan et al. address the spatial stationarity assumption by fitting models for multiple sub-areas. For some species, the assumption of stationarity is questionable and depends on the covariates.

Model performance refers to the ability of the model to reliably predict species distributions using values within the fitting conditions. In most of the papers in this special topic issue, data were split into a set for model fitting and another for model testing. Model performance is assessed by cross-validation using the test set. Yet, uncertainty in the covariates is seldom considered. Bowden et al. tackle this question and show that current cross-validation procedures may lead to overestimation of model performance. They also show that in general, presence/absence is estimated more reliably than abundance, and that model performance relies on the precision and spatio-temporal resolution of covariates.

Connectivity

The importance of considering connectivity when evaluating habitat occupancy has been long discussed (Bryan-Brown et al., 2017). The transport and movement abilities of individuals, combined with the spatial distribution of habitats, determine the dispersal patterns of key life stages of many coastal and marine species (Cowen and Sponaugle, 2009). Closing the life cycle by linking post-settlement larvae to juveniles and adults is valuable and informs a species' spatial distribution (Beck et al., 2001). Connectivity also plays a critical role in the theme of "seascapes" (Pittman, 2017). Two papers in this special topic issue (Cecino et al., Cristiani et al.) specifically focused on connectivity.

The effect of connectivity on predictions of species distributions can be quantified explicitly by incorporating connectivity metrics as covariates in SDMs. Cecino et al. offered a new approach to bridge standard SDMs with methods that assess connectivity. They found that centrality measures, which characterize connectivity, influenced the geographic structure of predicted habitat quality. Using a particle-tracking approach and network analysis, Cristiani et al. reported that topography acted to limit dispersal of benthic larvae among seagrass beds, and that subregions with limited exchange could be identified.

Management implications

Effective management and conservation require knowledge of species habitat use and the ecosystem effects of anthropogenic change. Multiple papers raised the issue of using SDMs to identify habitats that could be targeted for protection and species conservation, particularly with the potential for habitat degradation as the climate changes (Griffin et al.; Fabrizio et al.; Cecino et al.). In addition to using SDMs to identify suitable habitats, other papers within this issue used SDMs to assess anthropogenic impacts from dams, fisheries, and wind farms (Zydlewski et al.; Lowman et al.; Freidland et al.). A common message across these papers was the importance of reporting accurate measures of uncertainty alongside predictions to ensure proper interpretation.

Considerations for the future

(1) A theme common to all papers in this special topic issue is the necessity of merging multiple datasets from unrelated sources to assemble a more complete understanding of the modeled system. Most habitat assessments involve the merging of information from physical oceanography models, biogeochemistry models, and marine ecological observations including information on trophic interactions and connectivity. To the extent practical, resolution of differences in spatial and temporal scales required to address ecological questions and those typical of hydrodynamic models would allow the exploration and application of habitat models to additional species and systems (e.g., Behan et al.).

(2) Habitat assessments, especially those for mobile life stages, should incorporate concepts and results from the rapidly expanding field of movement ecology (Nathan et al., 2008; Abrahms et al., 2021). In particular, estimates of connectivity can be improved by considering movements of organisms, which are often elicited in response to the key constituents of habitat quality. Movement and connectivity information can be valuable for assessing the actual use (i.e., realized habitat) of potential habitats predicted by SDMs.

(3) Increasing refinement of the biological characteristics of individuals represented in particle-tracking and movement models would improve predictions and allow for species and regional differences to emerge. For example, Cristiani et al. defaulted to using a single mortality rate for their community-level analysis because species-specific information was lacking.

(4) Studies that consider habitat and connectivity effects at the population and ecosystem levels and over the entire life cycle of organisms can provide insight on the cumulative life-time effects of directional changes in habitat such as loss of spawning and rearing areas (e.g., Zydlewski et al.). Agent-based modeling approaches offer a viable approach to explore such effects.

(5) Results from habitat models, and SDMs in particular, should be expressed in a manner to ensure proper interpretation. Two major considerations are the proper reporting of the confidence in model predictions (e.g., confidence intervals), and distinguishing between habitat capacity, potential habitat, and realized habitat when interpreting model results. Additionally, all data and code from these analyses should be made available to aid in reproducibility.

(6) Habitat models could be improved by considering stakeholder and local knowledge (e.g., Lowman et al.). Such information could increase the understanding of past and present conditions, movement and migration patterns of organisms, and locations and environmental conditions under which organisms were previously found.

(7) Finally, practitioners should identify and examine key assumptions of models and data used to describe relationships between habitats and biota, including stationarity (e.g., Asch et al.). Additional comparison of methods and sensitivities of models to input data would also be valuable (e.g., Bowden et al.).

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abrahms, B., Aikens, E. O., Armstrong, J. B., Deacy, W. W., Kauffman, M. J., and Merkle, J. A. (2021). Emerging perspectives on resource tracking and animal movement ecology. *Trends Ecol. Evol.* 36 (4), 308–320. doi: 10.1016/j.tree.2020.10.018

Beck, M. W., Heck, K. L., Able, K. W., Childers, D. L., Eggleston, D. B., Gillanders, B. M., et al. (2001). The identification, conservation, and management of estuarine and marine nurseries for fish and invertebrates: a better understanding of the habitats that serve as nurseries for marine species and the factors that create site-specific variability in nursery quality will improve conservation and management of these areas. *Bioscience* 51 (8), 633–641. doi: 10.1641/0006-3568(2001)051[0633:TICAMO]2.0.CO;2

Bryan-Brown, D. N., Brown, C. J., Hughes, J. M., and Connolly, R. M. (2017). Patterns and trends in marine population connectivity research. *Mar. Ecol. Prog. Ser.* 585, 243–256. doi: 10.3354/meps12418

Cowen, R. K., and Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.* 1 (1), 443–466. doi: 10.1146/annurev.marine.010908.163757

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., et al. (2008). A movement ecology paradigm for unifying organismal movement research. *Proc. Natl. Acad. Sci.* 105 (49), 19052–19059. doi: 10.1073/pnas.0800375105

Pittman, S. J. (2017). *Seascape ecology* (Hoboken, NJ: John Wiley & Sons).

Thorson, J. T., and Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES J. Mar.* 74, 1311–1321.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 11 November 2020
doi: 10.3389/fmars.2020.579135

Estimating Shifts in Phenology and Habitat Use of Cobia in Chesapeake Bay Under Climate Change

Daniel P. Crear, Brian E. Watkins, Marjorie A. M. Friedrichs, Pierre St-Laurent and Kevin C. Weng*

Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, United States

Cobia (*Rachycentron canadum*) is a large coastal pelagic fish species that represents an important fishery in many coastal Atlantic states of the U.S. They are heavily fished in Virginia when they migrate into Chesapeake Bay during the summer to spawn and feed. These coastal habitats have been subjected to warming and increased hypoxia which in turn could impact the timing of migration and the habitat suitability of Chesapeake Bay. With conditions expected to worsen, we project current and future habitat suitability of Chesapeake Bay for cobia and predict changes in their arrival and departure times as conditions shift. To do this we developed a depth integrated habitat model from archival tagging and physiology data from cobia that used Chesapeake Bay, and applied the model to contemporary and future temperature and oxygen output from a coupled hydrodynamic-biogeochemical model of Chesapeake Bay. We found that estimated arrival occurs earlier and estimated departure time occurs later when temperatures are warmer and that by mid- and end-of-century cobia may spend on average up to 30 and 65 more days, respectively, in Chesapeake Bay. By mid-century we do not expect habitat suitability to change substantially for cobia, but by end-of-century we project it will significantly decline and shift closer to the mouth of Chesapeake Bay. Our study provides evidence that cobia will have the capacity to withstand near term impacts of climate change, but that their migration phenology varies from year to year with changing temperatures. These findings emphasize the need to incorporate the relationship between fishes and their environment into how fisheries are managed. This information can also help guide managers when deciding the timing and allocation of a fishery.

Keywords: archival tags, fisheries management, habitat modeling, recreational fishery, warming, hypoxia

OPEN ACCESS

Edited by:
Kenneth Alan Rose, University of Maryland Center for Environmental Science (UMCES), United States

Reviewed by:
Elliot John Brown, Technical University of Denmark, Denmark
Eliza C. Heery, Smithsonian Institution, United States

****Correspondence:***
Daniel P. Crear
dcrear8@gmail.com

Specialty section:
This article was submitted to Habitat and Distribution Models of Marine and Estuarine Species: Advances for a Sustainable Future, a section of the journal Frontiers in Marine Science

Received: *01 July 2020*
Accepted: *15 October 2020*
Published: *11 November 2020*

Citation:
Crear DP, Watkins BE, Friedrichs MAM, St-Laurent P and Weng KC (2020) Estimating Shifts in Phenology and Habitat Use of Cobia in Chesapeake Bay Under Climate Change. Front. Mar. Sci. 7:579135. doi: 10.3389/fmars.2020.579135

INTRODUCTION

Cobia (*Rachycentron canadum*) is a large coastal pelagic fish species that uses waters along the mid- and south- Atlantic regions of the U.S. east coast throughout the year. Along the east coast of the U.S., cobia migrate into bays and estuaries, such as Chesapeake Bay, in late spring/early summer to spawn and feed (Joseph et al., 1964; Smith, 1995; Perkinson et al., 2019). They remain in these habitats until late summer/early fall when they migrate primarily offshore to the shelf waters ranging from North Carolina to Florida (Crear et al., 2020b). The exact timing of both inshore and

offshore migrations fluctuate each year and are thought to be driven by temperature cues (Smith, 1995; Lefebvre and Denson, 2012). Anecdotal evidence from fishermen suggests that cobia have been entering Chesapeake Bay earlier in recent years, consistent with habitat suitability models suggesting that future climate warming will result in arrival into inshore habitats, like Chesapeake Bay, earlier in the spring (Crear et al., 2020b).

Cobia support a valuable recreational fishery on the U.S. east coast from Florida to Virginia. Estimated cobia landings from the recreational fishery occur primarily in Virginia or North Carolina state waters (SEDAR, 2020). With an average of approximately 225,000 cobia trips occurring annually in Virginia alone, valued between $488–$685 per trip (Scheld et al., 2020), the cobia fishery is extremely important for coastal states like Virginia. In recent years, estimated landings exceeded the Atlantic cobia allowable catch limits, which led the National Marine Fisheries Service (NMFS) to close the fishery in federal waters (NCDENR, 2016; NMFS, 2017). Despite the closure in federal waters, the cobia fishery remained open in state waters (within 3 nautical miles of the coast) because of the importance of the cobia fishery to many coastal states.

Warming within these ecologically and economically important inshore habitats has been occurring and is expected to intensify in the future with climate change (Najjar et al., 2010). As a result of atmospheric warming we expect to see an approximately 2°C increase by mid-century and a 5°C increase by end-of-century in Chesapeake Bay inferred from Saba et al. (2016) and Muhling et al. (2018).

Being adjacent to human populations, coastal habitats like Chesapeake Bay are often impacted by anthropogenic inputs (Brown et al., 2018). Specifically, anthropogenic nutrient inputs combined with warming waters has led to an increase in the extent and severity of hypoxic regions within Chesapeake Bay (Hagy et al., 2004; Rabalais et al., 2009; Najjar et al., 2010). We expect that as climate change continues these impacts will be exacerbated. Irby et al. (2018) project that the largest increase in cumulative hypoxic volume in Chesapeake Bay will occur between oxygen concentrations of 2–5 mg l^{-1}. With an increase in 2 and 5°C and corresponding solubility changes, phytoplankton growth rates, and organic matter remineralization, Chesapeake Bay is expected to see estimated reductions in dissolved oxygen of 0.5 and 1.5 mg l^{-1} by mid-century and end-of-century, respectively (Irby et al., 2018). These environmental changes may impact the suitability of Chesapeake Bay for cobia and could affect their arrival and departure time, a trend that has been seen in other migratory species (Sims et al., 2004; Jansen and Gislason, 2011).

The relationship between fish physiology and the environment is one way to understand the impacts of climate change on fish. A recent physiology study found that cobia are able to withstand temperatures as warm as 32°C; however, when exercised to exhaustion in these conditions, 30% of individuals suffered mortality (Crear et al., 2020a). Furthermore, this study showed cobia had a very high hypoxia tolerance, where individuals could tolerate oxygen levels as low as 1.7–2.4 mg l^{-1} at temperatures between 24 and 32°C (Crear et al., 2020a). Based on these results, it appears cobia are more hypoxia tolerant than many active predatory species and therefore might be less impacted by future decreases in dissolved oxygen concentration.

Habitat modeling has been used to assess the impacts of climate change on a number of marine species (Pinsky et al., 2013; Muhling et al., 2016; Kleisner et al., 2017; Morley et al., 2018; McHenry et al., 2019; Crear et al., 2020c). These studies have been used to identify both habitat reductions and range shifts. Although a recent study assessed climate impacts on cobia distribution along the U.S. east coast (Crear et al., 2020b), the spatial resolution of the analysis was too coarse to assess the changes in the habitat quality of Chesapeake Bay.

To predict future changes in phenology and habitat suitability for cobia within the Chesapeake Bay, we developed a habitat model parameterized with our physiology data (Crear et al., 2020a) and archival tagging data. This model was used to project the current arrival and departure times of cobia into Chesapeake Bay and the changes to this phenology in the future under climate change. In addition, our model was used to project changes in habitat suitability in Chesapeake Bay as a function of temperature and oxygen concentration.

MATERIALS AND METHODS

Tagging

Cobia were caught on rod and reel using typical recreational methods in Chesapeake Bay during the 2017–2018 summer months. Cobia were placed upside down in a V-board, and a hose with water pumping through it was inserted into the mouth. Cobia were measured and tagged by making a 2 cm incision in the abdominal wall, and inserting two tags. The first tag was an acoustic transmitter (V16-4L/4H coded transmitter, 16 mm diameter x 68 mm long, pulse interval 30–120 s, estimated battery life 1,613–3,650 days, 152–158 dB, 24 g in air, Vemco Inc., herein referred to as an "acoustic tag"). The second tag was a data storage tag (G5 data storage tag, 8 mm diameter x 31 mm long, 2.7 g in air, Cefas Technology Limited, herein referred to as a "data logger"), which was programmed to record temperature every 20 min and depth every 1 min for 2 years. A conventional tag was fixed to the data logger and designed to protrude from the incision to alert fishers that caught a tagged fish that a data logger was present inside the fish and that a monetary reward would be given if the tag was returned. The incision was closed with 3 interrupted sutures (PDS II) or 5–8 staples (Conmed Reflex One Skin). An external dart tag was inserted at the base of the dorsal fin. Fish were immediately released following tagging unless the fish appeared lethargic. When this occurred, we held the fish underwater as the boat moved forward slowly to irrigate the gills until the fish was able to swim off on its own. All fish capture, handling, and surgical procedures were approved by the College of William & Mary Institutional Animal Care and Use Committee (protocol no. IACUC-2017-05-26 133-kcweng).

Habitat Model

The habitat model followed similar methods described in Eveson et al. (2015) and Crear et al. (2020b), which uses the ratio between habitat use and habitat availability to

determine habitat suitability of the fish species. Habitat use was characterized by the temperatures utilized by tagged cobia (section "Habitat Use Densities" below) and habitat availability was the thermal distribution of the environment, as predicted via biogeochemical modeling (section "Habitat Availability Densities" below). A value greater than 1 indicates suitable conditions (i.e., the conditions the fish occupied occurred in a greater proportion than those conditions in the available habitat data), below 1 indicates unsuitable conditions, and equal to 1 represents no difference than random. In addition to the data from the data loggers, we used the environmental conditions simulated by the three-dimensional (3D) ChesROMS-Estuarine-Carbon-Biogeochemistry (ECB) model. This coupled hydrodynamic-biogeochemical model had a horizontal resolution of approximately 1 km × 1 km and 20 terrain-following vertical levels (i.e., depth levels that follow the contour of the bottom) (Shchepetkin and McWilliams, 2005) that have a higher vertical resolution near the surface and bottom of the water column (Feng et al., 2015; Da et al., 2018; Irby et al., 2018). The results from the ChesROMS-ECB model had three uses: estimates of habitat availability (daily outputs), for predictions of arrival and departure time of cobia to and from Chesapeake Bay over contemporary and future time periods (daily outputs), and for habitat suitability predictions over contemporary and future time periods (across summer averages). Details of the complete habitat model are described below in six steps (**Figure 1**).

1. Habitat Use Densities

Habitat use data came from the data loggers and was defined as the temperatures occupied by tagged cobia when the fish were in Chesapeake Bay. Presence inside and outside Chesapeake Bay was determined using the acoustic detections from these fish that were detected on acoustic receiver stations at the mouth of Chesapeake Bay (75.98°W). A fish was deemed inside Chesapeake Bay from when the fish was first detected west of 75.98°W to the last time the fish was detected west of 75.98°W. This method was selected because there was not an acoustic array with a fine enough resolution to determine when the fish was outside Chesapeake Bay. Data within the first 24 h of tagging and during the day of recapture were removed from each fish's dataset to disregard handling and tagging stress behaviors. Temperature and depth data were summarized by hour for each fish over a specified time range. Densities were extracted from temperature histograms with 0.5°C bins ranging from 1.5 to 33.5°C for each fish. The densities for each temperature were averaged over all fish present in Chesapeake Bay over a specified time range. These histograms and densities were generated for the months cobia arrive to (May and June) and depart from (August and September) Chesapeake Bay, as well as over all 5 months of Chesapeake Bay occupancy (May–September) combined. These densities were considered habitat use for cobia in Chesapeake Bay.

2. Habitat Availability Densities

Habitat availability information for Chesapeake Bay were temperatures and oxygen derived daily from the ChesROMS-ECB model for the time cobia are typically found in Chesapeake Bay (May 15–September 30) over the summers tagged cobia were at-liberty (2017–2019). We did not want to include all of May because the available temperatures would be skewed lower than what is actually available to cobia during the second half of May. Because the vertical levels in the model are not equally spaced, we generated eight depth bins at 3 m intervals (0–3 m, 3–6 m, 6–9 m, 9–12 m, 12–15 m, 15–18 m, 18–21 m, 21+ m). To allow each depth to be treated equally, all temperatures for a given latitude and longitude from levels within a depth bin were averaged over each day. To integrate cobia hypoxia tolerance quantified in Crear et al. (2020a), we removed those portions of the dataset corresponding to physiologically uninhabitable waters. These experiments showed that hypoxia tolerance declines in warmer waters (Crear et al., 2020a). To remove those portions where habitats were physiologically unavailable to cobia, we adjusted cells from depth bins to not available values (NAs) where temperatures were between 24 and 28°C and dissolved oxygen levels were less than or equal to 1.7 mg l^{-1}, where temperatures were greater than 28°C and less than 32°C and dissolved oxygen levels were less than or equal to 2 mg l^{-1}, and where temperatures exceeded 32°C and dissolved oxygen levels were less than 2.4 mg l^{-1} (Crear et al., 2020a). Because salinity preference is unknown for adult cobia while inhabiting Chesapeake Bay, we generated an area based on where cobia are caught while in Chesapeake Bay. This area extended slightly north of the mouth of the Potomac River (38.10°N) and excluded all areas in Chesapeake Bay tributaries (James, York, Rappahannock, and Potomac Rivers). We also excluded ocean waters, i.e., those east of the Chesapeake Bay mouth at 75.98°W. From here on, this region will be referred to as "Chesapeake Bay." The accuracy of the ChesROMS-ECB model has not been well-evaluated in shallow depths; therefore, any cells where bottom depths were less than 3 m were not included in these data. All temperatures over all eight depth bins for a specified time period were combined and a histogram and accompanying densities were created from 1.5 to 33.5°C with 0.5°C bins. These densities were generated for the months cobia arrive to (May and June) and depart from (August and September) Chesapeake Bay, as well as over all 5 months (May 15–September) combined. These densities were considered habitat availability for cobia in Chesapeake Bay.

3. Create Ratios

Ratios were calculated for each arrival month (May and June) and departure month (August and September) by dividing the corresponding habitat use densities by the habitat availability densities for those months. Ratios were also calculated from habitat use and habitat availability densities for all 5 months combined. Together this resulted in five sets of ratios (May, June, August, September, and all months combined).

4. Apply Ratios to 3D Habitat

The contemporary and future Chesapeake Bay habitats to predict over were derived from the ChesROMS-ECB model simulation. Daily 3D gridded arrays of temperature and oxygen over a 20-year time period (2000–2019) were considered to represent the contemporary habitat. We generated two future habitat

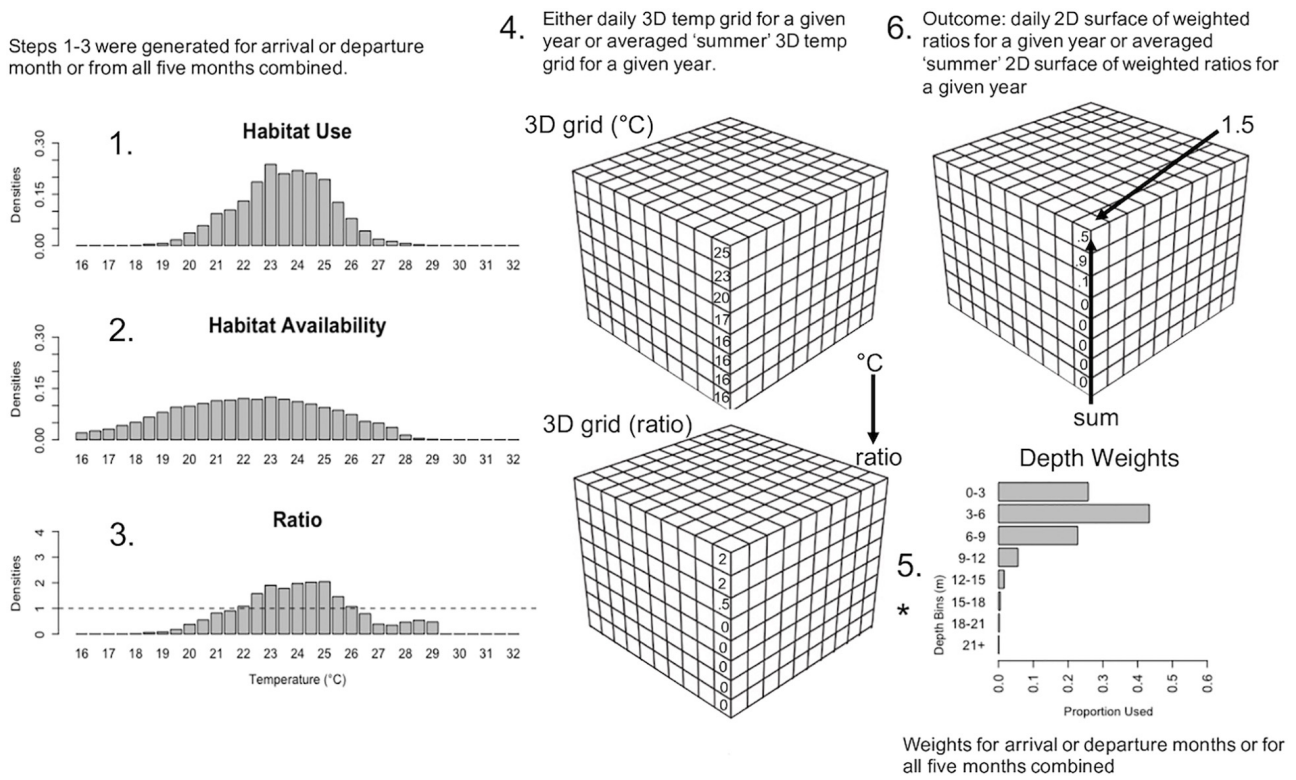

FIGURE 1 | Schematic of the habitat model and how Chesapeake Bay temperature data for specific time periods were run through it. **Step 1:** Histogram of habitat use densities from data loggers of tagged fish in Chesapeake Bay averaged over a specified time period; 2 arrival months (May, Jun.), 2 departure months (Aug., Sept.), and all 5 summer months combined (May–Sept.). **Step 2:** Histogram of habitat availability densities generated from daily 3D temperature arrays from Chesapeake Bay (ChesROMS-ECB model), summarized vertically into eight depth bins combined for each of the aforementioned time periods. **Step 3:** Ratios generated by dividing habitat use densities by habitat availability densities for each time period. **Step 4:** Either daily 3D arrays (for arrival and departure months) or 3D arrays averaged across two summer periods (May 15–Sept. 30 and Jun. 1–Aug. 31) were extracted from the ChesROMS-ECB model and ratios from Step 3 were assigned to each grid cell at each depth bin based on temperature in that grid cell. **Step 5:** The vertical habitat distribution of fish in Chesapeake Bay from data loggers was used to generate a depth weighting factor for the above eight depth bins, for each time period. The ratios generated in Step 4 were then multiplied (*) by the appropriate depth weighting factor based on each ratio's depth in each grid cell for each time period. **Step 6:** Sum the weighted ratios through the water column to get daily 2D surfaces of weighted ratios for arrival months (May, Jun.) and departure months (Aug., Sept.) for each year and yearly 2D surfaces of weighted ratios for the two summer periods (May 15–Sept. 30 and Jun. 1–Aug. 31).

scenarios predicted to occur by mid-century and end-of-century within Chesapeake Bay by adding deltas to the contemporary habitat data. We selected the mid-century deltas to be +2°C and −0.5 mg/l and the end-of-century deltas to be +5°C and −1.5 mg/l (based on Irby et al., 2018). It is important to mention that, similar to Irby et al. (2018), deltas were not selected to reflect any particular Representative Concentration Pathway (RCP) scenario or global climate model (GCM), but to more generally represent what is believed will occur by mid- and end-of-century and thus understand cobia's sensitivity to these changes. Deltas were applied to all 20 years and evenly over horizontal space and throughout the entire water column since observations suggest that climate change impacts temperatures along the north/south gradient and the temperature of the surface and bottom waters of Chesapeake Bay similarly (Preston, 2004; Irby et al., 2018; Hinson et al., in review). These 3D arrays were then summarized into the eight predefined depth bins and the other adjustments to the arrays described above (section "Habitat Availability Densities") were also done here. This resulted in daily 3D gridded arrays for three different 20-year timeseries, for contemporary, mid-century, and end-of-century.

To represent arrival (May and June) and departure months (August and September) daily 3D temperature arrays for each year were used. To represent the summer in Chesapeake Bay, the temperature arrays were averaged across days for all 5 months (May 15–September 30) and averaged across days for June 1–August 31 (months when cobia most heavily occupy Chesapeake Bay) for each year. Ratios for the arrival and departure months were then assigned to each grid cell at each depth bin based on the daily temperature in that grid cell and given month for all 20 years. Ratios for the 5 months combined were applied to the two average temperature arrays (5 months combined and June–August combined) based on the temperature in that grid cell at each depth bin for each year.

5. Weight Ratios by Depth

To produce a single ratio value for each latitude and longitude, depth weighting factors were generated for the arrival and

departure months and for all 5 months when cobia were present in Chesapeake Bay. The depth weighting factor was calculated by taking the proportion of hourly depth observations from the data loggers at each of the eight depth bins for the arrival and departure months and for all 5 months combined. Based on specified time period and the depth bin the ratio was in, the ratio was multiplied by the appropriate depth weighting factor. For example, if the ratio at a specific latitude and longitude was 2 at the 3–6 m depth bin in June and the depth weighting factor at 3–6 m was 0.5 in June, then the new weighted ratio would be 1.0 at the 3–6 m depth bin in June.

6. Sum Ratios Through Water Column

Once all ratios were weighted, the eight weighted ratios were summed through the water column at each grid cell for each month (May, June, August, and September) and the two arrays of combined months. This resulted in daily 2D surfaces of weighted ratios for May, June, August, September for each year and yearly 2D surfaces of weighted ratios for the May 15–September 30 time period and June 1–August 31 time period. Suitable habitat was considered to be any cell within the Chesapeake Bay habitat where the predicted ratio was greater than 1. Any predicted ratios below 1 were considered unsuitable habitat and any equal to 1 were considered no preference.

Arrival/Departure Analysis

To determine arrival and departure time of cobia in Chesapeake Bay we calculated available suitable habitat in Chesapeake Bay each day from May 1 to June 30 (for arrival) and August 1 to September 30 (for departure) for each year (2000–2019). To do this, the number of grid cells in the Chesapeake Bay area with predicted ratio values greater than 1 were tallied. Arrival day was considered the first date in May or June where greater than 50% of the cells were deemed suitable (>1). Departure day was considered the first date in August or September where less than 50% of the habitat was deemed suitable. We selected a 50% threshold because it estimated dates that fell within one standard deviation of the mean arrival and departure dates for cobia that were acoustically tagged. Specifically, we focused on departures in 2018 and arrivals in 2019 when there were 33 and 32 acoustically tagged cobia that left and entered Chesapeake Bay, respectively. To accommodate expected warming in our future scenarios, we extended Chesapeake Bay cobia habitat projections into April (for arrival) and October (for departure). There were very little or no contemporary habitat use data for cobia in Chesapeake Bay for the months of April and October; therefore ratios and depth weighting factors from May and September were used to predict over mid-century and end-of-century habitat in April and October, respectively.

To assess if arrival or departure day significantly changed over the current 20 year period (2000–2019) or over temperature we ran two linear models. The response variables were estimated yearly arrival day relative to May 1st and estimated yearly departure day relative to September 1st, for the arrival and departure model, respectively. The fixed effects for the arrival model were overall mean May water temperature in Chesapeake Bay each year and year, while the fixed effects for the departure model were overall mean September water temperature in Chesapeake Bay each year and year. Linear mixed effects models were run to determine if arrival and departure dates differed over the contemporary, mid-century, and end-of-century time periods using the nlme package (Pinheiro et al., 2013). For these models, the response variables were again estimated arrival day and estimated departure day for each year, for the arrival and departure model, respectively. The fixed effect was time period, while the random effect was year (2000–2019) in these models. Tukey's *post-hoc* tests were run to determine differences among the time periods. All statistics were evaluated at a significance level of $\alpha = 0.05$.

Habitat Suitability

The yearly 2D predicted ratio surfaces generated from the two summer periods (May 15–September 30; June 1–August 31) for each of the three time periods (contemporary, mid-century, and end-of-century) were used to calculate habitat suitability values for Chesapeake Bay. The predicted ratio values greater than 1 were summed over Chesapeake Bay for each year of each time period for each summer period to get yearly total habitat suitability index values. A linear mixed effects model was used to determine whether total habitat suitability index changed through each long term time period (contemporary, mid-century, and end-of-century) for the two summer periods (May 15–September 30; June 1–August 31). An interaction was used between these two fixed effects, long term time period and the two summer time periods, while the response variable was total habitat suitability index each year. Year was a random effect in this model. All R code for modeling and statistical analyses can be found in the **Supplementary Material**.

RESULTS

Data Retrieval

We received eight data loggers from fishermen. These eight fish ranged from 78.7 to 139.7 cm total length (mean ± SD: 106.0 ± 18 cm) (**Table 1**). Days-at-liberty within Chesapeake Bay ranged from 26 to 151 days (92 ± 46) days, yielding a total of 736 days of data.

TABLE 1 | Tag information for cobia tagged with a G5 data storage tag (Cefas), including total length when tagged, tagging date, and days-at-liberty in Chesapeake Bay.

Animal #	Total length (cm)	Date tagged	Days-at-liberty
A14128	78.7	2017-07-08	57
A14158	139.7	2018-06-01	26
A14135	100	2018-06-28	151
A14144	120.1*	2018-08-05	104
A14148	104.1	2018-08-05	114
A14149	108	2018-08-05	98
A14152	110	2018-08-15	143
A14164	101.6	2018-09-26	43

**Calculated from measured fork length using unpublished total length-fork length conversion equation.*

Habitat Model

Habitat suitability ratios were generated for each arrival month (May and June), for each departure month (August and September), and for all summer months combined. During arrival and departure months, cobia preferred temperatures from 21.5 to 27°C and 24.5 to 31°C, respectively. Over the entirety of the summer cobia preferred 22.5–28°C (**Figure 2**). Depth weighting factors were generated for each arrival and departure month and for all summer months combined. During early arrival (May) cobia preferred 0–6 m, but later into June cobia selected 0–9 m. During early departures (August) cobia were observed between 0 and 9 m, but during September cobia were most common at slightly deeper depths (0–12 m). When all summer months were combined, cobia were observed most frequently at depths between 0 and 9 m (**Figure 3**).

Arrival/Departure

Estimated cobia arrival time fluctuated over the last 20 years. Although there was no significant trend ($t = -0.52$, $p > 0.05$), in more recent years it appears that cobia have been arriving earlier in the year. For example, the mean estimated arrival time between earlier years (2000–2004) was 29.2 days since May 1st (all arrival values from here on are relative to May 1st), but 23.0 days for later years (2015–2019) (**Figure 4A**). Estimated arrival time significantly decreased ($t = -5.4$, $p < 0.001$) as average May temperature increased. Specifically, for every °C increase, arrival time occurred 8.6 days earlier in the Spring (**Figure 4B**). Arrival time significantly differed [$F_{(2,\ 38)} = 106.6$, $p < 0.001$] among time periods (contemporary, mid-century, end-of-century; **Figure 4C**), where contemporary mean arrival time (mean ± SD; 27.8 ± 9.0 days) significantly differed from mid-century arrival time (16.0 ± 7.9 days; $p < 0.05$) and end-of-century arrival time (−1.5 ± 7.0 days; $p < 0.05$). Arrival times for mid-century and end-of-century also significantly differed ($p < 0.05$).

Similar to arrival time, estimated departure time relative to September 1st (all departure values from here on are relative to September 1st) also varied over the last 20 years, and there was no significant trend ($t = 0.23$, $p > 0.05$). Despite this, the mean estimated departure time between earlier years was 3.0 days since September 1st, but 15.4 days for later years (**Figure 5A**). As average September temperature increased, estimated departure time significantly increased ($t = 6.0$, $p < 0.001$), where for every °C increase, departure time occurred 9.4 days later in the Fall (**Figure 5B**). Estimated departure time also significantly differed among time periods [$F_{(2,\ 38)} = 154.6$, $p < 0.001$; **Figure 5C**]. Specifically, contemporary mean departure time (10.1 ± 8.3 days) significantly differed from mid-century (27.7 ± 10.6 days; $p < 0.05$) and end-of-century (45.2 ± 6.7 days; $p < 0.05$). Departure times for mid-century and end-of-century significantly differed as well ($p < 0.05$).

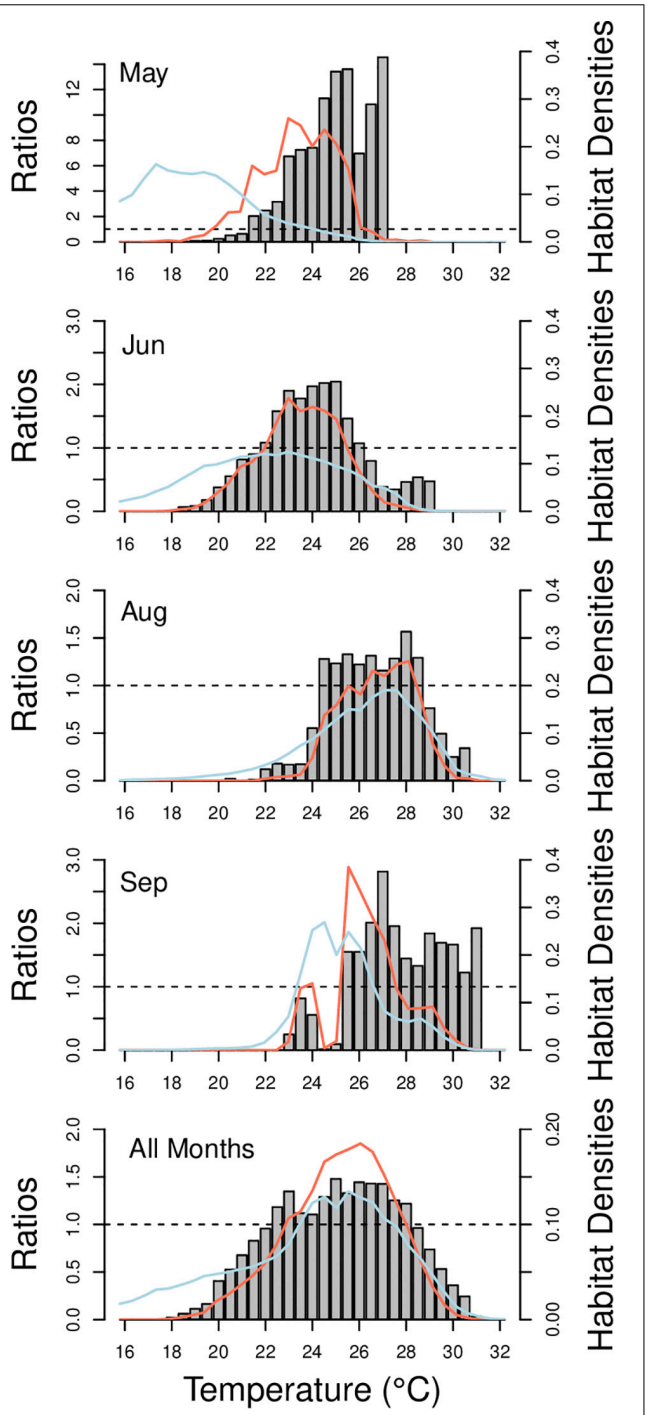

FIGURE 2 | Habitat suitability ratios from 16 to 32°C for each arrival month (May and June), departure month (August–September), and all summer months combined for when cobia were inside Chesapeake Bay (May–September). The ratios were developed from dividing the habitat use densities (red lines) by the habitat availability densities (blue lines) at each temperature during times when cobia were inside Chesapeake Bay. Dashed lines is at a ratio of 1.0.

Habitat Suitability

An interaction between long term time period (contemporary, mid-century, end-of-century) and the two summer time periods (May 15–September 30; June 1–August 31) significantly affected total habitat suitability index [$F_{(2,\ 95)} = 25.1$, $p < 0.001$].

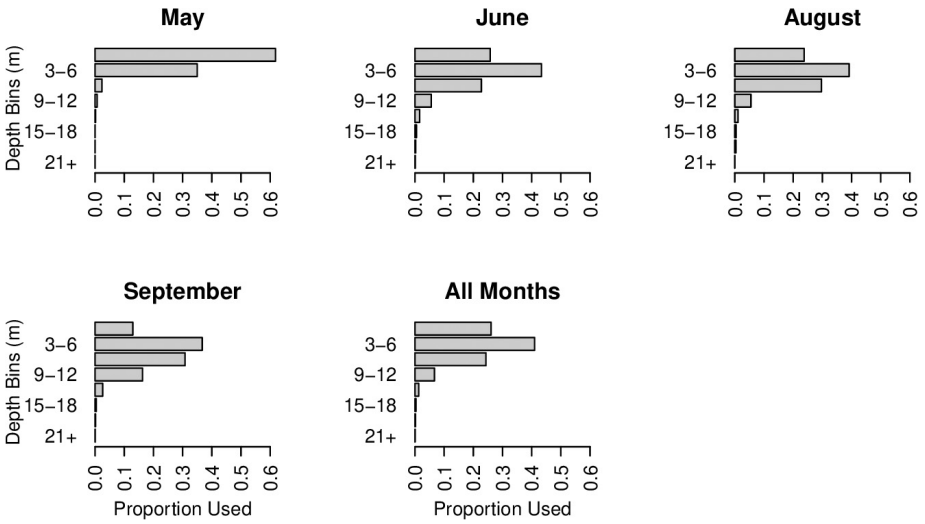

FIGURE 3 | Depth weighting factors for each 3 m depth bin for each arrival month (May and June), departure month (August and September), and all summer months combined (May–September).

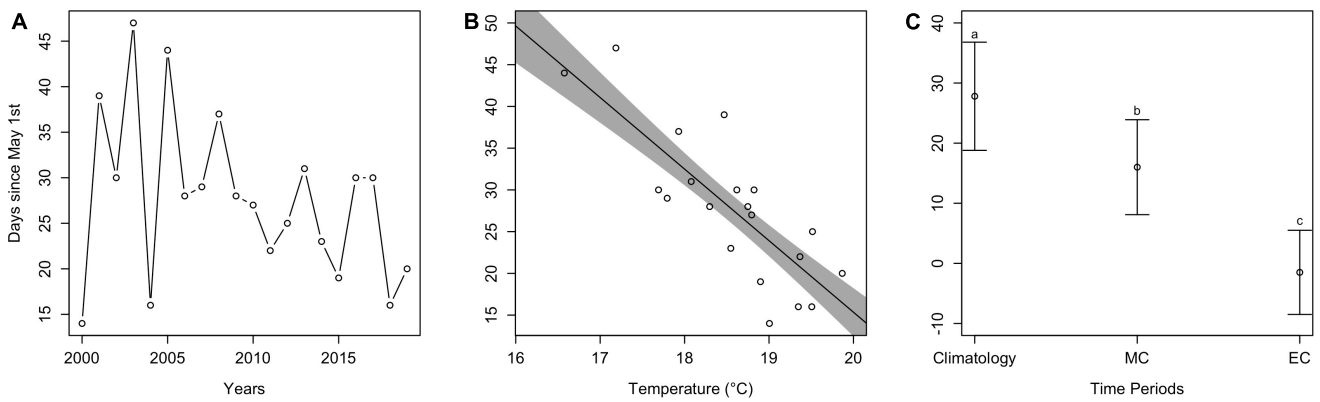

FIGURE 4 | Estimated arrival plots. **(A)** Estimated arrival time of cobia each year from 2000 to 2019. **(B)** The dots represent estimated cobia arrival times over mean May temperatures in Chesapeake Bay, while the line and shaded region represent model output and uncertainty. **(C)** Mean estimated cobia arrival time among climatology (i.e., contemporary), mid-century (MC), and end-of-century (EC) time periods. Different lower case letters indicate a statistical difference between time periods.

The most suitable habitat during June-August for cobia in Chesapeake Bay spans from north of the James River all the way to the northern extent of the study region (north of the Potomac River) for the contemporary time period (**Figure 6A**). Despite the lack of a significant difference between the total habitat suitability index for the two summer time periods ($p > 0.05$; **Figure 7**), mean habitat suitability does appear to decline slightly throughout most of the Chesapeake Bay cobia region when we incorporated days in May and all of September (**Figure 6D**). Mean habitat suitability from June-August, during mid-century shifted further south, closer to the mouth compared to the contemporary period (**Figure 6B**). In addition, the total habitat suitability index significantly decreased between the contemporary and mid-century periods for June-August ($p < 0.05$; **Figure 7**). However, when assessing May 15–September 30 during mid-century, total habitat suitability index did not decline relative to the contemporary period. Although there is no significant difference, it does appear that total habitat suitability index increased slightly by mid-century ($p > 0.05$; **Figure 7**). This is also reflected in habitat improvements over much of Chesapeake Bay (**Figure 6E**). Total habitat suitability index also significantly differed between the two summer periods during mid-century ($p < 0.05$; **Figure 7**). For end-of-century, we project a significant decrease in suitable cobia relative to mid-century for June-August ($p > 0.05$) and May 15–September 30 ($p > 0.05$; **Figure 7**). This is reflected in habitat loss throughout most of Chesapeake Bay and a shift toward the bay's mouth (**Figures 6C,F**). Total habitat suitability index was also significantly lower for June-August compared to May 15–September 30 ($p < 0.05$; **Figure 7**).

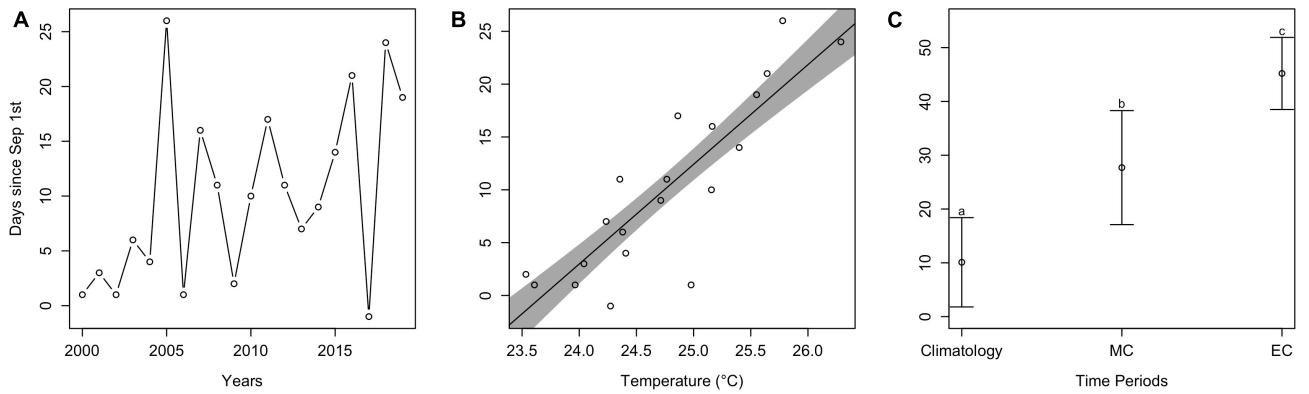

FIGURE 5 | Estimated departure plots. **(A)** Estimated departure time of cobia each year from 2000 to 2019. **(B)** The dots represent estimated cobia departure times over mean September temperatures in Chesapeake Bay, while the line and shaded region represent model output and uncertainty. **(C)** Mean estimated cobia departure time among climatology (i.e., contemporary), mid-century (MC), and end-of-century (EC) time periods. Different lower case letters indicate a statistical difference between time periods.

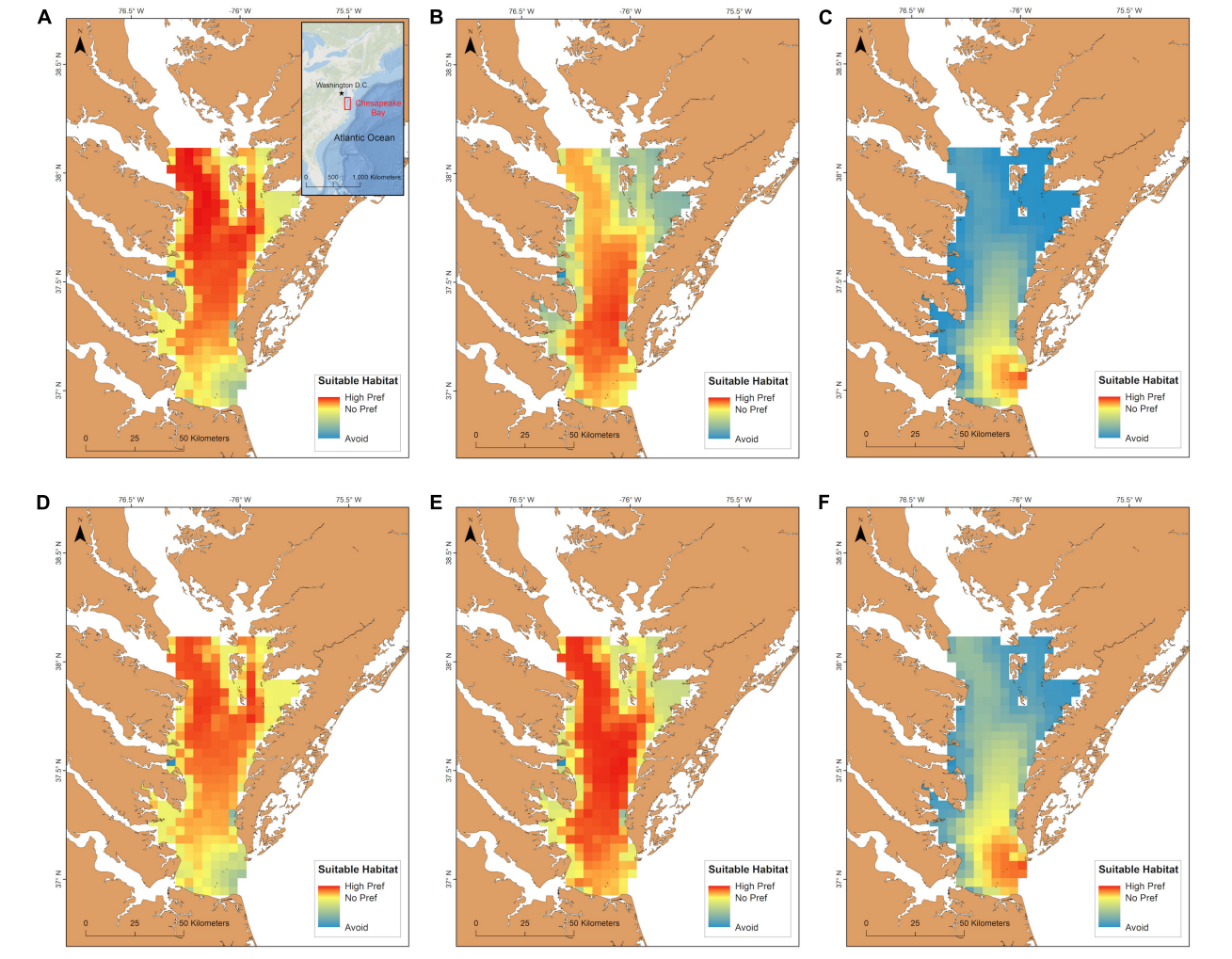

FIGURE 6 | Cobia habitat suitability within Chesapeake Bay averaged over years for the contemporary **(A,D)**, mid-century **(B,E)**, and end-of-century **(C,F)** time periods. The first row is mean habitat suitability for June-August (a-c) and the second row is mean habitat suitability for May 15–September 30 **(D–F)**.

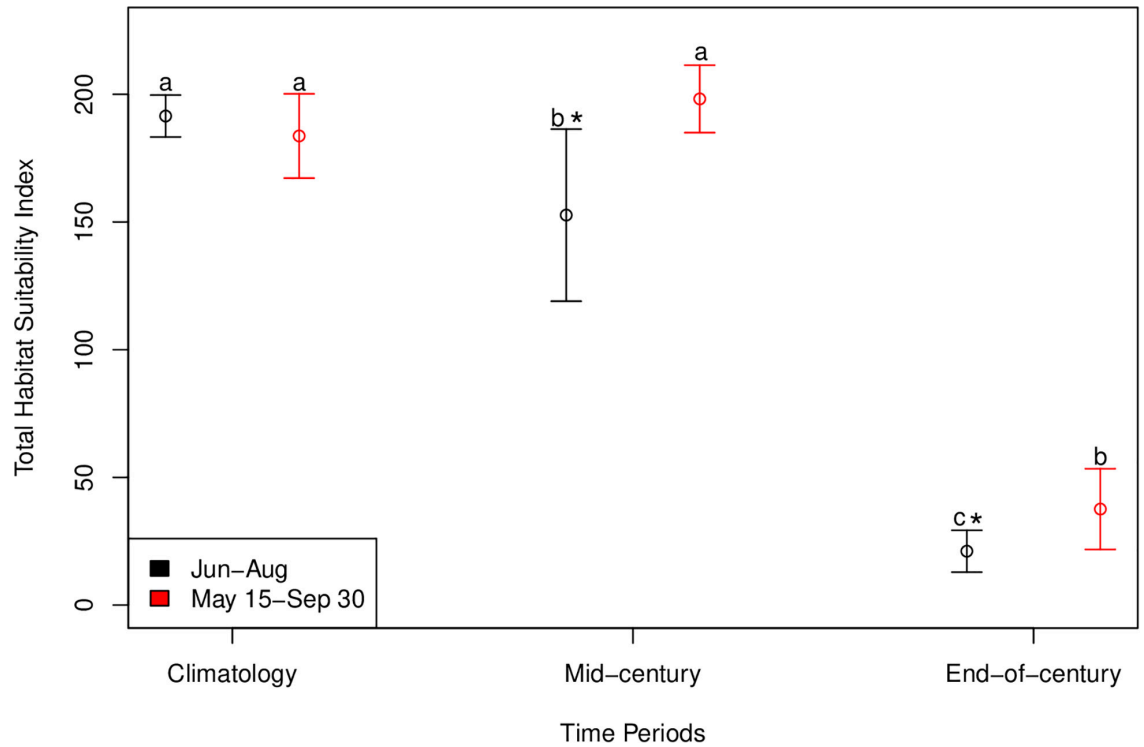

FIGURE 7 | Total habitat suitability index for cobia in Chesapeake Bay during contemporary, mid-century, and end-of-century time periods for each summer period (Jun–Aug and May 15–Sep 30). Error bars represent standard deviation. Black symbols represent the mean total habitat suitability index values for June–August and red symbols represent the mean total habitat suitability index values for May 15–September 30. Different lower case letters indicate a statistical difference among time periods within a summer period. For example, in the Jun-Aug summer period all three times periods are different from one another. An * indicates a statistical difference between summer periods within a time period.

DISCUSSION

This study presents the first attempt at describing the distribution of cobia within Chesapeake Bay. We generated a depth integrated habitat model to predict contemporary and future distributions of cobia within Chesapeake Bay using temperature, dissolved oxygen, and depth. By developing a novel model incorporating 3D habitat and physiology, we limit our model variables to those that are available in 3D, but we felt it was more important to incorporate depth than include other variables (many of which are not available at a fine enough resolution) because cobia use the entire water column. Although weighting and summarizing by depth has its benefits (e.g., two-dimensional output; patterns more easily discernable), the approach does have some limitations. For example, there may be small pockets of more suitable habitat at various depths (sub-gridscale) that are not expressed in our results and thus could potentially lead to an underestimation in some suitable habitat predictions. While inhabiting Chesapeake Bay, cobia are highly driven by biotic factors, like spawning and feeding, which were not included in our model; however, we believe environmental variables constrain cobia to certain areas in Chesapeake Bay, which are expressed in our model output. Because of this, we only assessed habitat suitability for the entire summer as a whole. The phenology of cobia arrival and departure to Chesapeake Bay appears to be cued by temperature, which then leads to inshore spawning and foraging. Therefore, we believe our temperature driven habitat model is justified in describing cobia phenology. It is important to note that another limitation of this study is the low sample size of cobia used in the model and that individuals used in our model may not be a full representation of cobia that summer Chesapeake Bay. An increase in sample size may lead to shifts in estimated phenology and habitat suitability. Despite this, our phenology estimates fell within one standard deviation of actual departure and arrival days based on over 30 acoustic tagged cobia. Lastly, we also would like to reiterate that the trends estimated from climate change projections are not intended to represent shifts under any RCP scenario or GCM, but more generally demonstrate cobia's sensitivity to future oxygen and temperature conditions likely to occur around mid- and end-of-century.

Contemporary Trends

It is clear temperature is a major driver of cobia arrival to and departure from Chesapeake Bay. Over the last 20 years, when temperatures were warmer in May, cobia arrived earlier. Tag pop off locations and modeling suggest that cobia overwinter offshore along the U.S. shelf from North Carolina to Florida (Crear et al., 2020b; Jensen and Graves, 2020). Although cobia are unaware of the temperature in Chesapeake Bay when they

are in their overwintering offshore waters, warm temperature cues on the shelf are most likely reflected in Chesapeake Bay as well. This has been observed in mackerel, which arrived to their spawning grounds earlier when sea surface temperature was warmer at a rate of -15 ± 12.1 days/°C (Jansen and Gislason, 2011). Although the trend was not significant, it does appear that when comparing estimated arrival time approximately 20 years ago to today, cobia may be migrating into Chesapeake Bay earlier in recent years. Earlier migrations have been recorded in various tuna species as well, which have migrated to feeding grounds up to 14 days earlier over a 25 year period (Dufour et al., 2010).

Once cobia enter Chesapeake Bay in May, the occurrence of high ratio values generated from the habitat use and habitat availability densities and the use of shallower habitats suggest that cobia are likely seeking out the warm shallow habitats until some of the deeper areas (>6 m) warm up. During the main summer months in Chesapeake Bay (June–August), most of the areas in southern Chesapeake Bay appear to be suitable for cobia. The suitability of most of these areas allow cobia to spawn and feed freely without being restricted by less optimal conditions, except for areas that are excluded as a result of low oxygen. However, because cobia have a high hypoxia tolerance the negative impacts of low oxygen is likely minimal. These favorable conditions are ideal for cobia, which are indeterminate batch spawners and are capable of spawning multiple times over the spawning season (Brown-Peterson et al., 2001; Lefebvre and Denson, 2012). To further define cobia habitat use in estuaries, it would be useful for future studies to examine the relationships between cobia and the location of bathymetric features, manmade structure (e.g., buoys and pilings), salinity, tidal currents, and bait schools (e.g., menhaden) all of which are thought (based on anecdotal evidence) to influence cobia movements while inshore.

Typically, once spawning is complete, individuals have foraged, and temperatures cool in Chesapeake Bay, cobia begin their migration out onto the shelf. However, when temperatures in September are warmer than usual, cobia remain in Chesapeake Bay longer. Similar to arrivals, despite no significant trend over time, it appears that in recent years cobia are leaving Chesapeake Bay later compared to 20 years ago. Although we did not directly look at changes in dissolved oxygen levels on cobia phenology, it is likely not a major driver because of cobia's hypoxia tolerance and the lack of large hypoxic zones during the months they arrive and depart to and from Chesapeake Bay. Overall, these results suggest that cobia phenology has already been impacted by climate change over the last 20 years.

Future Trends

Phenology trends observed over the last 20 years are projected to extend more rapidly in the future as climate change contributes to even warmer conditions. By mid-century and end-of-century, conditions in Chesapeake Bay may allow cobia to arrive by mid-May and late April/early May on average, respectively. Furthermore, departure time is predicted to extend to the end of September and mid-October by mid-century and end-of-century, respectively. When combining the estimated earlier arrival and later departure, our results indicate that cobia may increase their residence time in Chesapeake Bay by an extra 30 days by mid-century and 65 days by end-of-century. Despite this large increase in the number of days, cobia may be faced with more unsuitable habitat during the months when temperatures are the warmest. When combining more favorable conditions during the last 2 weeks of May and all of September, suitable habitat does not change much by mid-century. If climate change continues at its current rate, suitable habitat is expected to decrease significantly and shift closer to the Chesapeake Bay mouth by the end-of-century, even when incorporating the second half of May and all of September. Further, these trends should be interpreted as the average summer cobia distribution, which in turn, could potentially hide periodic marine heatwave events that could result in displacement and further habitat reduction. Future decline in suitable habitat has similarly been projected for many other coastal species (Albouy et al., 2013; Brown et al., 2016). For example, an increase in sublethal temperatures in the San Francisco Estuary as a result of climate change will likely cause behavioral avoidance of these temperatures and considerable habitat reduction for the Delta smelt (*Hypomesus transpacificus*) (Brown et al., 2016).

Although habitat shifts and community composition has favored warm-adapted species (Howell and Auster, 2012), the predicted occurrence of more extreme temperatures has the capability to negatively impact warm-adapted species like cobia. For example, if cobia migrate into Chesapeake Bay earlier, spawning may occur earlier. This could impact the survival of eggs and larvae, which depend on the timing of specific temperatures and favorable primary production conditions (Durant et al., 2007). On the other hand, if spawning duration is extended and phytoplankton blooms align, larval survival may improve (Kristiansen et al., 2011). If substantial spawning habitat is lost for estuarine species like cobia, we may see populations decline. We may also see species shift their spawning habitat to more poleward estuaries or offshore habitat where conditions are more favorable for spawning adults and larvae. Recent genetic studies suggest that cobia already have a separate offshore spawning group (Darden et al., 2014; Perkinson et al., 2019), meaning cobia have the ability to spawn in offshore waters. Furthermore, Crear et al. (2020b) found that over the next 60–80 years, there will continue to be an increase in the proportion of suitable cobia habitat in state waters (within 3 nautical miles of shore) from Maryland to Massachusetts during the summer spawning months. Likewise, non-warm-adapted species like, Northeast Artic cod (*Gadus morhua*) have already shifted their spawning habitat further north over the last half a century, a behavior likely linked to climate change (Sandø et al., 2020). If cobia shift their spawning habitat further north or extend their time inshore, they may subsequently shift their overwintering grounds to be closer to their spawning habitat. Because cobia offshore migrations are driven by temperature, we hypothesize that their overwintering grounds are likely plastic. Therefore, although suitable habitat may still be available further south in the winter (Crear et al., 2020b), it may be less energetically costly to migrate off the shelf toward the Gulf Stream instead of migrating to shelf waters

between North Carolina and Florida. If further migrations are made, females may be required to divert energy away from egg production to compensate. Although we talk about these impacts being decades away, some of these hypotheses can be tested today as marine heatwaves become more prevalent along the Northeast Shelf.

Cobia may have the ability to behaviorally adapt to climate change within Chesapeake Bay. The fact that cobia could withstand water temperatures as warm as 32°C (Crear et al., 2020a), suggests that if waters warmed throughout Chesapeake Bay, areas with water temperatures up to 32°C could still be habitable or maybe even suitable. Meaning, temperatures between 22.5 and 28°C may be preferred, but if unavailable, cobia could still inhabit warmer temperatures. If this is the case, our projected future habitat suitability maps may underestimate the amount of suitable habitat in Chesapeake Bay. If this is possible, it is still unknown whether other essential functions like growth or reproduction could be compromised.

Management Implications

Hundreds of thousands of recreational fishermen enjoy fishing for cobia each year in Virginia alone and it appears this number has increased in recent years (B. Watkins pers. comm.). As the amount of time cobia spend in Chesapeake Bay increases with climate change, management will need to be prepared for catch increases. In recent years, the fishery in Virginia has been open from June 1 to various dates in September. If the fishing season dates remain the same, we may expect to see an increase in the catch and release of more cobia in May and more cobia retained later in the season. Our study and a previous study (Crear et al., 2020a) suggest cobia have the capacity to withstand near term (+30 years) impacts of climate change, which is a good sign for a fishery that has grown over the last decade.

A dynamic approach to management may prepare managers for the early migrations to or late departures from Chesapeake Bay. Dynamic management provides managers with the opportunity to adjust managed areas temporally and spatially in time when our coastal waters are changing faster than we are accustomed to (Maxwell et al., 2015; Dunn et al., 2016; Welch et al., 2019). Specifically, as the predictability of coastal ocean models improve, we will have the capacity to couple them with our cobia habitat model to project the timing of cobia migrations months to seasons in advance. This information could be used to guide the timing of the fishing season in Virginia and also influence allocation of cobia among states on a broader scale. As fish behaviorally adapt to changing water conditions, it is critical that management be prepared to adapt as well.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The animal study was reviewed and approved by the College of William & Mary Institutional Animal Care and Use Committee (protocol no. IACUC-2017-05-26 133-kcweng).

AUTHOR CONTRIBUTIONS

DC substantially contributed to project conception and design, animal tagging, analysis and interpretation of data, and drafting the manuscript. BW contributed to animal tagging and assisted in manuscript revision. MF and PS provided funding support, generated environmental data, and provided manuscript revisions. KW contributed funds to the project, participated in project conception and design, animal tagging, and assisted in manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

Financial support was provided by the VIMS startup funds, Virginia Sea Grant (award # V721500), the National Oceanic and Atmospheric Administration (NOAA) Saltonstall Kennedy Program (award # NA18NMF4270199), the NOAA's National Centers for Coastal Ocean Science (award # NA16NOS4780207), and the National Science Foundation (award # ACI-1548562).

ACKNOWLEDGMENTS

We thank many VIMS undergraduates and VIMS graduate students and staff for their help in tagging cobia, including S. Kilmer, B. Gallagher, M. Oliver, J. Martinez, E. Biesack, and B. Hamilton. We especially thank all of the cobia fishermen that helped us catch and tag cobia and for those that returned tags to us. We want to thank R. Gallagher at NCSU for providing some acoustic tagging data for fish that shared our archival tag and their acoustic transmitter. We also thank J. Hartog and T. Patterson for their advice in model design. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) supercomputer Comet at SDSC through allocation OCE-160013. We also acknowledge William & Mary Research Computing (https://www.wm.edu/it/rc) for providing computational resources and/or technical support that have contributed to the results reported within this manuscript. Lastly, we appreciate the input and advice on research contributing to this work from Alistair Hobday and Rich Brill, and comments on the manuscript from the two reviewers. This is Virginia Institute of Marine Science contribution # 3959.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2020.579135/full#supplementary-material

REFERENCES

Albouy, C., Guilhaumon, F., Leprieur, F., Lasram, F. B. R., Somot, S., Aznar, R., et al. (2013). Projected climate change and the changing biogeography of coastal Mediterranean fishes. *J. Biogeogr.* 40, 534–547. doi: 10.1111/jbi.12013

Brown, E. J., Vasconcelos, R. P., Wennhage, H., Bergström, U., Støttrup, J. G., van de Wolfshaar, K., et al. (2018). Conflicts in the coastal zone: human impacts on commercially important fish species utilizing coastal habitat. *ICES J. Mar. Sci.* 75, 1203–1213. doi: 10.1093/icesjms/fsx237

Brown, L. R., Komoroske, L. M., Wagner, R. W., Morgan-King, T., May, J. T., Connon, R. E., et al. (2016). Coupled downscaled climate models and ecophysiological metrics forecast habitat compression for an endangered estuarine fish. *PLoS One* 11:e0146724. doi: 10.1371/journal.pone.0146724

Brown-Peterson, N. J., Overstreet, R. M., Lotz, J. M., Franks, J. S., and Burns, K. M. (2001). Reproductive biology of cobia. Rachycentron canadum, from coastal waters of the southern United States. *Fish. Bull.* 99, 15–28.

Crear, D. P., Brill, R. W., Averilla, L. M., Meakem, S. C., and Weng, K. C. (2020a). In the face of climate change and exhaustive exercise: the physiological response of an important recreational fish species. *R. Soc. Open Sci.* 7, 1–13. doi: 10.1098/rsos.200049

Crear, D. P., Watkins, B. E., Saba, V. S., Graves, J. E., Jensen, D. R., Hobday, A. J., et al. (2020b). Contemporary and future distributions of cobia. *Rachycentron canadum*. *Div. Distrib.* 26, 1002–1015. doi: 10.1111/ddi.13079

Crear, D. P., Latour, R. J., Friedrichs, M. A. M., St-Laurent, P., Weng, K. C. (2020c). Sensitivity of a shark nursery habitat to a changing climate. *Mar. Ecol. Prog. Ser.* 652, 123–136. doi: 10.3354/meps13483

Da, F., Friedrichs, M. A. M., and St-Laurent, P. (2018). Impacts of atmospheric nitrogen deposition and coastal nitrogen fluxes on oxygen concentrations in Chesapeake Bay. *J. Geophys. Res. Oceans* 123, 5004–5025. doi: 10.1029/2018jc014009

Darden, T. L., Walker, M. J., Brenkert, K., Yost, J. R., and Denson, M. R. (2014). Population genetics of Cobia (Rachycentron canadum): implications for fishery management along the coast of the southeastern United States. *Fish. Bull.* 112, 24–35. doi: 10.7755/FB.112.1.2

Dufour, F., Arrizabalaga, H., Irigoien, X., and Santiago, J. (2010). Climate impacts on albacore and bluefin tunas migrations phenology and spatial distribution. *Prog. Oceanogr.* 86, 283–290. doi: 10.1016/j.pocean.2010.04.007

Dunn, D. C., Maxwell, S. M., Boustany, A. M., and Halpin, P. N. (2016). Dynamic ocean management increases the efficiency and efficacy of fisheries management. *Proc. Natl. Acad. Sci. U S A.* 113, 668–673. doi: 10.1073/pnas.1513626113

Durant, J. M., Hjermann, D. Ø, Ottersen, G., and Stenseth, N. C. (2007). Climate and the match or mismatch between predator requirements and resource availability. *Clim. Res.* 33, 271–283. doi: 10.3354/cr033271

Eveson, J. P., Hobday, A. J., Hartog, J. R., Spillman, C. M., and Rough, K. M. (2015). Seasonal forecasting of tuna habitat in the Great Australian Bight. *Fish. Res.* 170, 39–49. doi: 10.1016/j.fishres.2015.05.008

Feng, Y., Friedrichs, M. A. M., Wilkin, J., Tian, H., Yang, Q., Hofmann, E. E., et al. (2015). Chesapeake Bay nitrogen fluxes derived from a land-estuarine ocean biogeochemical modeling system: Model description, evaluation, and nitrogen budgets. *J. Geophys. Res.* 120, 1666–1695. doi: 10.1002/2015jg002931

Hagy, J. D., Boynton, W. R., Keefe, C. W., and Wood, K. V. (2004). Hypoxia in Chesapeake Bay, 1950–2001: long-term change in relation to nutrient loading and river flow. *Estuaries* 27, 634–658. doi: 10.1007/bf02907650

Howell, P., and Auster, P. J. (2012). Phase shift in an estuarine finfish community associated with warming temperatures. *Mar. Coastal Fisher.* 4, 481–495. doi: 10.1080/19425120.2012.685144

Irby, I. D., Friedrichs, M. A. M., Da, F., and Hinson, K. E. (2018). The competing impacts of climate change and nutrient reductions on dissolved oxygen in Chesapeake Bay. *Biogeosciences* 15, 2649–2668. doi: 10.5194/bg-15-2649-2018

Jansen, T., and Gislason, H. (2011). Temperature affects the timing of spawning and migration of North Sea mackerel. *Continental Shelf Res.* 31, 64–72. doi: 10.1016/j.csr.2010.11.003

Jensen, D. R., and Graves, J. E. (2020). Movements, habitat utilization, and post-release survival of cobia (Rachycentron canadum) that summer in Virginia waters assessed using pop-up satellite archival tags. *Anim. Biotele.* 8, 1–13.

Joseph, E. B., Norcross, J. J., and Massmann, W. H. (1964). Spawning of the cobia, Rachycentron canadum, in the Chesapeake Bay area, with observations of juvenile specimens. *Chesapeake Sci.* 5, 67–71. doi: 10.2307/1350791

Kleisner, K. M., Fogarty, M. J., McGee, S., Hare, J. A., Moret, S., Perretti, C. T., et al. (2017). Marine species distribution shifts on the US Northeast Continental Shelf under continued ocean warming. *Prog. Oceanogr.* 153, 24–36. doi: 10.1016/j.pocean.2017.04.001

Kristiansen, T., Drinkwater, K. F., Lough, R. G., and Sundby, S. (2011). Recruitment variability in North Atlantic cod and match-mismatch dynamics. *PLoS One* 6:e17456. doi: 10.1371/journal.pone.0017456

Lefebvre, L. S., and Denson, M. R. (2012). Inshore spawning of cobia (Rachycentron canadum) in South Carolina. *Fish. Bull.* 110, 397–412.

Maxwell, S. M., Hazen, E. L., Lewison, R. L., Dunn, D. C., Bailey, H., Bograd, S. J., et al. (2015). Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Mar. Policy* 58, 42–50. doi: 10.1016/j.marpol.2015.03.014

McHenry, J., Welch, H., Lester, S. E., and Saba, V. (2019). Projecting marine species range shifts from only temperature can mask climate vulnerability. *Glob. Change Biol.* 25, 4208-4221.

Morley, J. W., Selden, R. L., Latour, R. J., Frölicher, T. L., Seagraves, R. J., and Pinsky, M. L. (2018). Projecting shifts in thermal habitat for 686 species on the North American continental shelf. *PLoS One* 13, 1–28. doi: 10.1371/journal.pone.0196127

Muhling, B. A., Brill, R., Lamkin, J. T., Roffer, M. A., Lee, S.-K., Liu, Y., et al. (2016). Projections of future habitat use by Atlantic bluefin tuna: mechanistic vs. correlative distribution models. *ICES J. Mar. Sci.* 74, 698–716. doi: 10.1093/icesjms/fsw215

Muhling, B. A., Gaitán, C. F., Stock, C. A., Saba, V. S., Tommasi, D., and Dixon, K. W. (2018). Potential salinity and temperature futures for the Chesapeake Bay using a statistical downscaling spatial disaggregation framework. *Estuaries Coasts* 41, 349–372. doi: 10.1007/s12237-017-0280-288

Najjar, R. G., Pyke, C. R., Adams, M. B., Breitburg, D., Hershner, C., Kemp, M., et al. (2010). Potential climate-change impacts on the Chesapeake Bay. *Estuarine Coastal Shelf Sci.* 86, 1–20. doi: 10.1016/j.ecss.2009.09.026

NCDENR (2016). *NOAA Fisheries Announces the Atlantic Migratory Group (Georgia to New York) Cobia Recreational Fishing Season will close on June 20, 2016 [Online]. National Marine Fisheries Service, National Oceanic and Atmospheric Administration*. Available online at: http://portal.ncdenr.org/web/mf/fb16-018-cobia (accessed March 9, 2016).

NMFS (2017). *Atlantic Cobia (Georgia through New York) Recreational Fishing Season is Closed in Federal Waters [Online]. National Oceanic and Atmospheric Administration*. Available online at: https://www.fisheries.noaa.gov/bulletin/atlantic-cobia-georgia-through-new-york-recreational-fishing-season-closed-federal (accessed January 25,2017).

Perkinson, M., Darden, T., Jamison, M., Walker, M. J., Denson, M. R., Franks, J., et al. (2019). Evaluation of the stock structure of cobia (Rachycentron canadum) in the southeastern United States by using dart-tag and genetics data. *Fish. Bull.* 117:220. doi: 10.7755/fb.117.3.9

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2013). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-111*.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science* 341, 1239–1242. doi: 10.1126/science.1239352

Preston, B. L. (2004). Observed winter warming of the Chesapeake Bay estuary (1949–2002): implications for ecosystem management. *Environ. Manage.* 34, 125–139.

Rabalais, N. N., Turner, R. E., Díaz, R. J., and Justić, D. (2009). Global change and eutrophication of coastal waters. *ICES J. Mar. Sci.* 66, 1528–1537. doi: 10.1093/icesjms/fsp047

Saba, V. S., Griffies, S. M., Anderson, W. G., Winton, M., Alexander, M. A., Delworth, T. L., et al. (2016). Enhanced warming of the northwest Atlantic Ocean under climate change. *J. Geophys. Res. Oceans* 121, 118–132. doi: 10.1002/2015JC011346

Sandø, A. B., Johansen, G. O., Aglen, A., Stiansen, J. E., and Renner, A. H. (2020). Climate change and new potential spawning sites for Northeast Arctic cod. *Front. Mar. Sci.* 7:28. doi: 10.3389/fmars.2020.00028

Scheld, A. M., Goldsmith, W. M., White, S., Small, H. J., and Musick, S. (2020). Quantifying the behavioral and economic effects of regulatory change in a recreational cobia fishery. *Fish. Res.* 224:105469. doi: 10.1016/j.fishres.2019.105469

SEDAR (2020). *SEDAR 58 – Atlantic Cobia Stock Assessment Report*. North Charleston: SEDAR.

Shchepetkin, A. F., and McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean model.* 9, 347–404. doi: 10.1016/j.ocemod.2004.08.002

Sims, D. W., Wearmouth, V. J., Genner, M. J., Southward, A. J., and Hawkins, S. J. (2004). Low-temperature-driven early spawning migration of a temperate marine fish. *J. Anim. Ecol.* 73, 333–341. doi: 10.1111/j.0021-8790.2004.00810.x

Smith, J. W. (1995). Life history of cobia Rachycentron canadum (Osteichthyes: Rachycentridae), in North Carolina waters. *Brimleyana* 23, 1–23. doi: 10.21276/ijaq.2011.8.1.1

Welch, H., Hazen, E. L., Bograd, S. J., Jacox, M. G., Brodie, S., Robinson, D., et al. (2019). Practical considerations for operationalizing dynamic management tools. *J. Appl. Ecol.*56, 459–469. doi: 10.1111/1365-2664.13281

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fmars.2020.593844

Deep-Sea Coral and Sponge Taxa Increase Demersal Fish Diversity and the Probability of Fish Presence

Mark J. Henderson[1], David D. Huff[2] and Mary M. Yoklavich[3]*

[1] U.S. Geological Survey, California Cooperative Fish and Wildlife Research Unit, Humboldt State University, Arcata, CA, United States, [2] Fish Ecology Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Newport, OR, United States, [3] Fisheries Ecology Division, Southwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Santa Cruz, CA, United States

Fishes are known to use deep-sea coral and sponge (DSCS) species as habitat, but it is uncertain whether this relationship is facultative (circumstantial and not restricted to a particular function) or obligate (necessary to sustain fish populations). To explore whether DSCS provide essential habitats for demersal fishes, we analyzed 10 years of submersible survey video transect data, documenting the locations and abundance of DSCS and demersal fishes in the Southern California Bight (SCB). We first classified the different habitats in which fishes and DSCS taxa occurred using cluster analysis, which revealed four distinct DSCS assemblages based on depth and substratum. We then used logistic regression and gradient forest analysis to identify the ecological correlates most associated with the presence of rockfish taxa (*Sebastes* spp.) and biodiversity. After accounting for spatial autocorrelation, the factors most related to the presence of rockfishes were depth, coral height, and the abundance of a few key DSCS taxa. Of particular interest, we found that young-of-the-year rockfishes were more likely to be present in locations with taller coral and increased densities of *Plumarella longispina*, *Lophelia pertusa*, and two sponge taxa. This suggests these DSCS taxa may serve as important rearing habitat for rockfishes. Similarly, the gradient forest analysis found the most important ecological correlates for fish biodiversity were depth, coral cover, coral height, and a subset of DSCS taxa. Of the 10 top-ranked DSCS taxa in the gradient forest (out of 39 potential DSCS taxa), 6 also were associated with increased probability of fish presence in the logistic regression. The weight of evidence from these multiple analytical methods suggests that this subset of DSCS taxa are important fish habitats. In this paper we describe methods to characterize demersal communities and highlight which DSCS taxa provide habitat to demersal fishes, which is valuable information to fisheries agencies tasked to manage these fishes and their essential habitats.

Keywords: essential fish habitat, multivariate analysis, indicator species, submersible survey, rockfishes (*Sebastes*), spatial autocorrelation

OPEN ACCESS

Edited by:
Paul Snelgrove, Memorial University of Newfoundland, Canada

Reviewed by:
Evan Edinger, Memorial University of Newfoundland, Canada
Krista Dawn Baker, Department of Fisheries and Oceans, Canada

****Correspondence:***
Mark J. Henderson
mark.henderson@humboldt.edu

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: *11 August 2020*
Accepted: *30 October 2020*
Published: *23 November 2020*

Citation:
Henderson MJ, Huff DD and Yoklavich MM (2020) Deep-Sea Coral and Sponge Taxa Increase Demersal Fish Diversity and the Probability of Fish Presence. Front. Mar. Sci. 7:593844. doi: 10.3389/fmars.2020.593844

INTRODUCTION

It is well established that fishes co-occur with deep-sea corals and sponges (DSCS), but it is debated whether this relationship is facultative (circumstantial and not restricted to a particular function) or obligate (necessary for sustainability because fishes use them for spawning, breeding, feeding, or growth to maturity). If it is the latter, then DSCS meet the definition of essential fish habitat (Rosenberg et al., 2000), and these sensitive taxa would require protection from human activities that may cause them damage (e.g., benthic trawling). Some researchers have concluded that fishes are found among DSCS species in the same proportion as other structures, suggesting that DSCS are simply facultative habitat (Freese and Wing, 2003; Auster, 2005; Tissot et al., 2006; Edinger et al., 2007). In contrast, others have suggested that some DSCS species are essential fish habitat because they provide fishes nursery and rearing grounds (Stone, 2006; Harter et al., 2009; Baillon et al., 2012), trophic interactions (George et al., 2007; Quattrini et al., 2012), shelter (Du Preez and Tunnicliffe, 2011; Stone, 2014), and increased population growth (Foley et al., 2010). These conflicting conclusions have resulted in a call for more quantitative analyses designed to compare the associations between demersal fishes and DSCS species while controlling for important covariates such as depth and substratum type (Auster, 2005; Tissot et al., 2006).

Due to the difficulty in observing ecological interactions in deep-sea habitats, it is a challenge to examine associations between fishes and structure-forming invertebrates (i.e., coral and sponges) on the appropriate scale. Without the ability to observe *in situ* ecological interactions, a few studies have defined associations between fishes and DSCS as co-occurrence in trawl catches (Edinger et al., 2007; D'Onghia et al., 2010). This definition can be overly broad because trawls integrate catches through large areas, potentially with different substratum types, and do not provide any information on the proximity of the fishes and DSCS species. In addition, trawling focuses on low-relief mud and sand sediments, while most corals and sponges occur in high-relief rocky substrata. Other catch data, such as from long-lines and gillnets, can yield information on the distribution and abundance of fishes in deep-sea coral habitats (Husebø et al., 2002; D'Onghia et al., 2012). These capture techniques have the benefit of being more spatially explicit if the lines also coincidentally snag a piece of coral. However, these sampling methodologies are much more size selective for fishes and depend on species movement and foraging behaviors. As a result, catch data only reveal a limited sample of the fishes residing within rocky habitat.

More recently, scientists have gained the ability to observe deep-sea habitats *in situ* using video collected with occupied submersibles, remotely operated vehicles (ROVs), and other camera systems. Underwater video collected with these platforms is a vast improvement over previous collection methods, because we can observe the locations of fishes and DSCS relative to one another. These data generally are collected along line transects through habitats and, like many sampling methods, provide only a 'snapshot' of the associations between fishes and DSCS (Sward et al., 2019). These data required a new definition of what constitutes an association between fishes and structure-forming invertebrates. Because proximity is the most apparent evidence of association, many studies have defined the association between fishes and DSCS as being located within 1 m of each other (Krieger and Wing, 2002; Stone, 2006). This definition may be overly restrictive as fishes generally have home ranges thousands of times larger than 1 m. For example, blue (*Sebastes mystinus*) and black (*S. melanops*) rockfishes observed with acoustic telemetry had home ranges of approximately 0.2–0.25 km^2 (Green et al., 2014). Thus, we propose a broader definition of fish-invertebrate associations, which comprises fishes and DSCS found within the same patch of habitat, defined as having the same primary (>50% cover) and secondary (>20% cover) substratum type. This less restrictive definition assumes that fishes may be using the DSCS within the same habitat even if they were not observed in close proximity during the relatively brief observation period of the survey. Two potential explanations for why this definition may be more reasonable are: (1) some fishes may have a core area within their home range and use DSCS taxa only for a specific function (e.g., predator refuge or feeding) (Jorgensen et al., 2006), (2) fishes may be constantly moving throughout their home range looking for food resources (Reese, 1989), which makes the probability of observing them near an individual DSCS during a survey rather low.

Another challenge in examining associations between fishes and structure- forming invertebrates has been interpreting complex datasets that comprise multiple fish and invertebrate species. The need to reduce complexity has often led researchers to focus on individual species of interest (Fosså et al., 2002; Costello et al., 2005; Harter et al., 2009) or to ignore individual species and instead look at species assemblages and species diversity (Krieger and Wing, 2002; Auster, 2005; Ross and Quattrini, 2009). While both of these approaches provide valuable results, they can miss potentially important relationships between individual fish and invertebrate species. Focusing on individual "charismatic" coral species such as the reef-building *Lophelia pertusa* [(syn. *Desmophyllum pertussum*, Addamo et al., 2016); Fosså et al., 2002; Costello et al., 2005; Lessard-Pilon et al., 2010; Addamo et al., 2016] and *Oculina varicosa* (Harter et al., 2009) is intuitive because these are often the dominant structure-forming deep-sea invertebrate taxa resident in many habitats. However, restricting the analyses to these species results in overlooking many other structure-forming taxa such as sponges. Sponges can be the most abundant invertebrate megafauna in areas of the deep sea (Stone, 2006; Buhl-Mortensen et al., 2010; Baillon et al., 2012), and could therefore provide important habitat to various fish populations.

In this study, we used multiple analytical techniques to examine the associations between fishes and DSCS taxa, observed with *in situ* video collected by submersible in the Southern California Bight (SCB). Our first objective was to classify the SCB demersal habitat into different groups based on the predominant demersal DSCS communities. We used several multivariate methods to classify habitats into different DSCS assemblages. Classifying the different community assemblages is valuable to provide a measure of how much connectivity there is between different habitat types (Bowden et al., 2016). Our

second objective was to identify which DSCS taxa co-occurred with demersal fish taxa of management and conservation concern while controlling for potential confounding variables, such as depth and substratum type. To achieve this objective, we used two approaches: (1) logistic regression analysis that allowed us to look at individual relationships between fish taxa and DSCS taxa while controlling for spatial autocorrelation and (2) gradient forest analysis that allowed us to look at how DSCS taxa affected fish biodiversity. For both of these methods we also included additional ecological covariates to account for the effect of depth and substratum type. This approach provided a means to test the hypothesis that the relationship between DSCS and demersal fishes was either obligate or facultative. If the relationship was facultative, we would not expect any DSCS taxa to be associated with demersal fishes after controlling for the other ecological covariates. In contrast, if the relationship was obligate, this approach allowed us to identify the DSCS taxa that specific demersal fish taxa were associating with more than would be expected based on the observed depth and substratum type. Identifying which DSCS taxa might provide essential fish habitats makes it more feasible to locate those areas that are most vulnerable to potential damaging activities such as bottom trawling, petroleum exploitation, and cable laying (Sundahl et al., 2020), and to prioritize these areas for protection.

MATERIALS AND METHODS

Data Collection

The SCB is one of the most heavily exploited areas on the west coast of North America, having been commercially and recreationally fished over the past 100 years (Love, 2006). More than 5,000 benthic invertebrate species and approximately 500 fish species inhabit this region, likely because the SCB comprises ocean conditions representative of both the northern Oregonian and southern San Diegan zoogeographic provinces and because a wide variety of marine habitats are found in this area (Dailey et al., 1993). Much of the fish diversity within the SCB is dominated by rockfishes (Love et al., 2002), which are also heavily targeted by both recreational and commercial fishers. To protect this diversity of life, many areas within the SCB are now protected from some, or all, types of fishing.

We conducted 497 underwater video transect surveys of demersal communities throughout the SCB (**Figure 1**) using

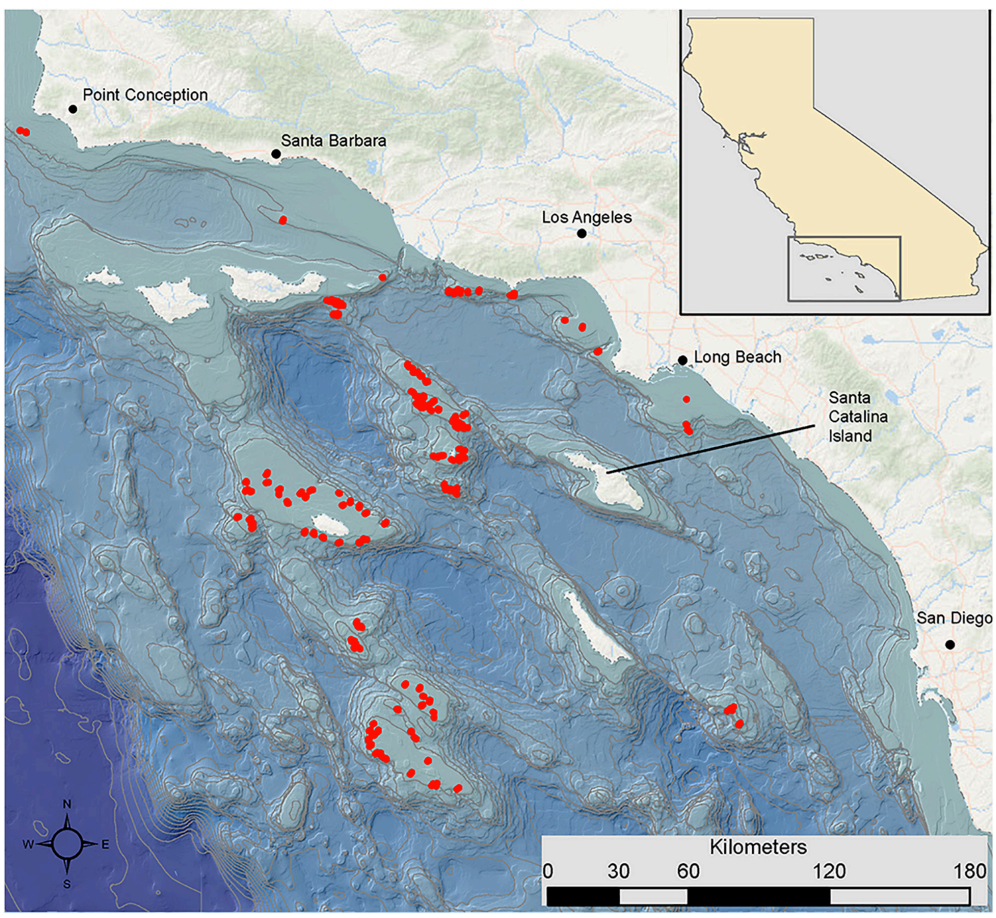

FIGURE 1 | Map of the study area of the Southern California Bight off the coast of California (United States), with dive locations (red points) and 200 m depth contours (gray lines).

human occupied submersibles (*Delta* and *Dual Deepworker*) during autumn (September, October, and November) between 2002 and 2011 (**Supplementary Table 1**; see Tissot et al., 2006; Love et al., 2009 for detailed methods). We used the *Delta* ($n = 398$ transects, depths: 22–342 m) and the *Dual Deepworker* ($n = 99$ transects, depths: 95–437 m) submersibles to conduct transects within 1 m of the seafloor at speeds of 0.5–1.0 knot. We restricted our analysis to transects deeper than 50 m, which is how Roberts et al. (2009) defined deep water. The scientific observer within the submersible verbally recorded the identification, number, and size of all fishes occurring within a strip 2–2.5 m along each transect in real time onto an externally mounted video camera oriented in the same direction as the observer. The video camera on the *Delta* submersible was a Sony TR-81 Hi-8 camcorder with 400 TVL resolution. The video camera on the Dual Deepworker was a Sony HVR Z1U digital camera with 1080i resolution. The video footage for each transect was later reviewed in the laboratory, and sponges and corals were identified, counted, and measured along the transects. Size of fishes (total length, cm) and DSCS (total height and width, cm) were estimated using reference light points from two parallel lasers installed 20 cm apart on either side of the externally mounted video camera positioned above the middle viewing-porthole on the starboard side of the submersible.

Seafloor habitat was characterized during the subsequent video analysis as the extent of substratum types and depth along each transect. Substratum types, comprised of pinnacle rock, boulder, rugose rock, cobble, gravel, pebble, flat rock, sand, and mud, were characterized along each transect by recording primary (>50% of the area) and secondary (>20% of the area) percent cover of each type based on review of the seafloor habitat visible in the video footage. We refer to each unique combination of primary and secondary substratum types along each transect as a patch. For the analyses, we also quantified the substratum types into a relative measure of "relief" based on the rough approximations used to classify each habitat. To do this, we extracted the minimum relief for each substratum type based on Greene et al. (1999; **Supplementary Table 2**) and summed the values of the primary and secondary habitat relief weighted by their percent cover (i.e., 0.5 for primary and 0.2 for secondary). Thus, we converted the categorical habitat classifications into a relative continuous substratum relief measurement.

Analysis Overview

To characterize associations between DSCS assemblages and demersal fishes, we conducted a series of multivariate analyses followed by fitting logistic regressions and gradient forest models to examine ecological correlates of fish presence. All analyses were conducted using the R programming language (R Core Team, 2019). We first used a cluster analysis to group habitats based on their DSCS assemblages. We then used both an indicator species analysis developed by Dufrene and Legendre (1997) and the "NbClust" package (Charrad et al., 2014) in R to determine the appropriate number of clusters necessary to describe the DSCS and fish assemblages. We then used a combination of random forest and logistic regression models to identify which physical and biological factors had the most influence on rockfishes (*Sebastes* spp.) density, distribution, and presence as well as to determine if any DSCS and fish taxa were statistically associated with each other while accounting for other physical (e.g., depth and substratum type) covariates. Finally, we used a gradient forest analysis to identify which of the ecological correlates used in the logistic regression analysis had the largest influence on fish biodiversity.

Cluster Analysis

To describe the DSCS assemblages, and improve the interpretability of the cluster analysis, we combined DSCS abundances for all patches of the same substratum type within 50 m depth bins. This unique combination of primary substratum (>50%), secondary substratum (>20%), and depth (hereafter referred to as a habitat unit) was the cluster analysis sample unit. For example, all patches that had boulder as a primary substratum type, rock as a secondary type, and were located at a depth between 100 and 150 m comprised the BR100 habitat unit. These habitat units were not spatially or temporally explicit (i.e., they were comprised of patches dispersed throughout the sample area and sampled anytime between 2002 and 2011), therefore we did not account for spatial or temporal autocorrelation at this stage in our analysis.

We then conducted a hierarchical cluster analysis on the Hellinger-transformed DSCS density data (Legendre and Gallagher, 2001). We calculated the density of DSCS by dividing the number of observed DSCS by patch area. Patch area was estimated as the length of each patch multiplied by transect width. Prior to conducting this cluster analysis, we removed four DSCS taxa that were not identified to a sufficient taxonomic level to contain any useful information (**Supplementary Table 3**). We used the Hellinger transformation for the cluster analysis because it had the best properties compared with other common multivariate transformations such as Wisconsin, frequency, range, and ubiquity (McCune et al., 2002). This conclusion was based on the "rankindex" function in the R package "vegan" using the Morista-Horn distance metric (Oksanen et al., 2017). Legendre and Gallagher (2001) showed that the Hellinger transformation had good statistical properties when compared to other common transformations used in multivariate transformation. The Morista-Horn distance metric is the recommended distance measure for ecological data due to its relative independence from sample size (Wolda, 1981). We conducted the cluster analysis using the "gaverage" agglomerative clustering algorithm in the R package "cluster" (Belbin et al., 1992; Maechler et al., 2017). The "gaverage" algorithm was referred to by Belbin et al. (1992) as the "flexible beta" and uses the Lance-Williams formula to specify how dissimilarities are computed. We used the default beta value of −0.1 as recommended by Belbin et al. (1992) for a general agglomerative hierarchical clustering strategy.

We next employed both the indicator species method and the "NbClust" package (Charrad et al., 2014) to determine how many clusters most appropriately described our observed data. We used the indicator species method of Dufrene and Legendre (1997) to determine which species were more often associated with a given cluster of habitat units then would be expected

by chance. From the indicator species method, we were able to identify species found primarily in one cluster and present within the majority of habitat units of that cluster. Although taxa could be present in multiple clusters, they could only be an indicator species for a single cluster. This method can be used to select an appropriate number of clusters by sequentially increasing the number of clusters and quantifying the total number of indicator species that have a significant ($p < 0.05$, via Monte Carlo) association with any single cluster (McCune et al., 2002). Because our primary interest was identifying associations between fishes and DSCS assemblages, we quantified the number of fish taxa that were indicator species for a cluster of habitat units defined by the DSCS assemblages. As with the DSCS, prior to the analysis we removed fish taxa that were not demersal or were not identified to a sufficient taxonomic level to contain any useful information ($n = 14$, **Supplementary Table 3**). We then selected the number of habitat unit clusters that had the most significant indicator species (McCune et al., 2002). This analysis was conducted using the "multipatt" function in the R package "indicspecies" (De Caceres and Legendre, 2009).

We also used the "NbClust" function (Charrad et al., 2014) to explore the number of clusters recommended by other indices. The "NbClust" function uses 30 indices for determining the number of clusters by varying different combinations of clusters, distance measures, and clustering methods. We used the "kmeans" cluster method because (1) "gaverage" method was not available in "NbClust" and (2) it is an iterative approach to forming clusters and therefore less susceptible to chaining, or forming large clusters from poorly separated groups.

Logistic Regression

Our next objective was to identify which DSCS and rockfish taxa were associated with each other, and our first step was to use logistic regression models to estimate the probability of presence for individual fish taxa as a function of biotic and abiotic covariates in each habitat patch. In contrast to the cluster analysis, where our goal was to identify broad-scale assemblage associations, we used each individual patch as the sample unit for this analysis to ensure that any observed relationships were among fishes and DSCS in relatively close spatial proximity. When examining fish-habitat associations, both the spatial scale of the experimental unit and the choice of statistical model are important in determining the outcome (Sharma et al., 2012).

Because individual patches were our sample unit, we wanted to account for any potential spatial autocorrelation in fish distributions. Therefore, we fit our models using a hierarchical Bayesian framework that easily allowed us to add complexity and determine if the added complexity improved model fit. We fit our logistic regression models using the integrated nested Laplace approximation implemented with the R-INLA package (Lindgren and Rue, 2015). The R-INLA package uses the Matérn correlation function to estimate a spatial covariance matrix based on the distance between two sample locations and two estimated parameters (Zuur et al., 2017). The two estimated parameters are k, which is related to the range of spatial dependency, and s, which is a spatial variance parameter. To determine if spatial autocorrelation improved model fit, we compared the global model (the full model with all potential ecological covariates) with and without spatial autocorrelation using Watanabe's information criterion (WAIC; Watanabe, 2013). The WAIC value of the model that included spatial autocorrelation had to be more than 4 units lower than the non-spatial model for the spatial model to be selected as the most parsimonious.

Potential ecological covariates included in the model were specific to each patch and included depth, temperature, salinity, substratum, percent DSCS cover, DSCS density, mean DSCS height, and the density of each DSCS taxon. Each of these covariates was selected *a priori* based on their hypothesized influence on fish presence. Percent DSCS cover was calculated as the total width of all DSCS species observed within a patch divided by the patch area. Although it is common to include measures of bathymetry to derive seafloor characteristics, such as the bathymetric position index (BPI), we chose to only use observations collected *in situ* and, thus, used our estimate of "relief" from the primary and secondary substratum type. Prior to fitting models, we examined the densities of all DSCS within the patches to determine if there was a minimum habitat patch size where DSCS densities might be biased. Based on this analysis, we removed all patches smaller than 3 m^2 (**Supplementary Figure 1**). Also prior to fitting models, we used pairwise Pearson correlations to quantify collinearity among variables and selected a single variable from any pair with a correlation over 0.7. Based on this analysis, we excluded temperature and salinity as covariates because they were collinear with depth. We also excluded any DSCS taxa that were observed in less than 1% of patches to avoid potential analysis issues that could be caused by small sample sizes.

Due to the large number of potential DSCS taxa that were candidates, we decided to use a random forest analysis as an initial screening method to eliminate ecological covariates that had a low likelihood of association with fish presence. To quote from Ellis et al. (2012), "a random forest (Breiman, 2001) is an ensemble of a large number of regression (or classification) trees, in which each tree is fit to a bootstrap sample (i.e., with replacement) of the observations, and each partition within a tree is split on the best of a random subsample of the predictor variables." Random forests have generally performed better than other approaches to examine species distributions (Prasad et al., 2006) as well as fish-habitat relationships (Knudby et al., 2010). To account for the spatial nature of our analysis, we used a recently developed spatial extension of the random forest that accounts for the spatial dependency and heterogeneity in the data (Georganos et al., 2019). Although random forest is a valuable method for ranking relative variable importance, it is difficult to identify individual relationships between taxa. Because our ultimate goal was to provide managers with a prioritized list of DSCS taxa that were associated with fish taxa, we decided to use the random forest as an initial screening and use the logistic regression to identify the specific taxa that were most associated with the presence of individual fish taxa. After some preliminary examination of the data, we arbitrarily used two criteria to screen variables based on the spatial random forest results: (1) the maximum increase in mean squared error (MSE) was greater than 20 and (2) the percent increase in MSE (calculated as the

increase in MSE multiplied by 100 divided by the MSE standard deviation) was greater than 15%.

The response of the logistic regression model was whether or not an individual fish taxon was present or absent within that patch. Thus, for each taxon (i) in each patch (j) our global model without spatial autocorrelation was:

$$\text{Logit(fish pres)}_{ij} = \text{Depth}_{ij} + \text{Substratum}_{ij} + \text{Cover}_{ij} + \text{Height}_{ij} + \text{Density}_{ij} + \varepsilon_{ij}$$

Where the response was the logit transformed binomial of whether or not a fish was observed within a patch, depth was the mean depth of that patch measured along the transect, substratum was the continuous relative relief as calculated from the primary and secondary substratum types, Cover was the percent DSCS cover, Height was the mean height of all DSCS in each patch, Density was the densities of the DSCS taxa that could be associated with that fish taxa after the random forest screening, and e was the residual error. All fixed covariates were scaled (i.e., z-transformed) so that the model coefficient estimates were on a similar scale.

Our global model with spatial autocorrelation was nearly identical, but included an additional random effect (u) to account for spatial autocorrelation:

$$\text{Logit}\left(\text{fish pres}\right)_{ij} = \text{Depth}_{ij} + \text{Substratum}_{ij} + \text{Cover}_{ij} + \text{Height}_{ij} + \text{Density}_{ij} + u_{ij} + \varepsilon_{ij}$$

$$u_{ij} \sim GMRF(0, \Sigma)$$

The spatial autocorrelation term (u) is assumed to have a random intercept and come from a Gaussian Markov random field (GMRF) with mean 0 and covariance matrix S. The covariance matrix (σ) is calculated using the two parameters (κ and σ) estimated by the Matérn correlation function.

We used WAIC to conduct model selection and used area under the curve to assess model performance. We conducted our model selection in two stages to ensure that we accounted for depth- and substratum-related covariates. Our first model selection stage included only depth, substratum, DSCS height, and DSCS cover (i.e., we excluded DSCS densities) and, thus, compared a maximum of 16 models. The purpose of this stage was to identify the physical and biological covariates that accounted for as much variation in fish presence as possible prior to including individual DSCS taxa. We selected the model with the fewest covariates and a delta WAIC values less than 4. We used Bayesian model averaging (Hoeting et al., 1999) if more than one model was selected based on those criteria. In the second model selection stage, we included all potential DSCS taxa in addition to the physical and biological covariates selected in the first stage. Covariate were considered important in estimating whether or not a fish was present within a habitat patch based on whether the 90% credible interval (90% CrI) of that covariate included zero, indicating there was no effect of that covariate on the response. The 90% CrI is the interval in which there is a 90% probability that the true (unknown) parameter estimate exists, given the observed data. During the second model selection stage, we removed any covariates that had 90% CrI that overlapped with zero. We refit the model after removing covariates with 90% CrIs that included zero, and repeated this process until all covariates had 90% CrIs that did not include zero or until there were no significant covariates remaining. We used the area under the receiver operating characteristic curve (AUC) method to gauge the adequacy of the model relative to the observed data (Hosmer et al., 2013).

We chose a subset of nine rockfish taxa that were either of high commercial value or conservation concern to present our logistic regression results. We downloaded commercial landing data for 2000–2017 from the National Oceanic and Atmospheric Administration website: https://foss.nmfs.noaa.gov/apexfoss/. We calculated the average landings in California for each species over that time period and merged those data with our dataset. We then selected the top five most landed rockfish species by pounds. In addition to those species, we included bocaccio (*S. paucispinis*), canary rockfish (*S. pinniger*), cowcod (*S. levis*), and young-of-year rockfish, as these taxa were of conservation interest due to current (or recent) protection status and the importance of young-of-year growth to maturity in the definition of essential fish habitat. To visualize the effect of each covariate, we calculated the probability of fish presence over the observed range (1–99% quantiles) of an individual covariate based on the logistic regression coefficient estimates and plotted these values against the individual covariate. To isolate the effect of that covariate on individual taxa, we constrained the other covariates to median values. We refer to these figures as "response plots."

Gradient Forest Analysis

In addition to identifying the ecological covariates that are associated with an increased probability of fish presence, we also were interested in determining the covariates that increased fish biodiversity. We used gradient forest analysis (Ellis et al., 2012), which is a multivariate extension of the random forest method, to quantify multispecies responses to environmental gradients and to understand the drivers of differences in biodiversity (Pitcher et al., 2012). This method first uses a random forest to determine which covariates improve fit of the observations, and then uses a novel algorithm to determine the importance of each predicator for all species within a data set (Ellis et al., 2012; Pitcher et al., 2012). The gradient forest component collates the splits from each random forest along the gradient of each predictor (Ellis et al., 2012; Pitcher et al., 2012). See Ellis et al. (2012) for further statistical details regarding this approach. We ran the gradient forest for all observed fish taxa in the same patches as the logistic regression to determine which physical and biological variables had the largest influence fish biodiversity.

RESULTS

Data Summary

There were general trends of primary substratum type and biological community with depth. After removing 51 small

patches (<3 m^2) from the dataset, and 29 habitat patches that were repeated in multiple surveys, we were left with 5144 habitat patches. We observed a mean of 10.53 ± 0.77 habitat patches on each transect and the mean size of each patch was 72.00 ± 3.42 m^2. The ratio of hard to soft primary substratum types declined from 57% at 50 m to 33% at 200 m (**Figure 2**). Between 200 and 350 m, the ratio of hard to soft substrata was approximately 50%.

The number of observed DSCS taxa was dependent on both substratum type and depth (**Figure 3A**). The highest number of DSCS taxa occurred when both primary and secondary substrata were hard. In these patches, the number of DSCS taxa peaked between 250 and 300 m. Soft habitat patches had lower numbers of DSCS taxa and did not co-vary with depth. There was a similar relationship between percent DSCS cover and substratum type and depth, although percent cover declined at deeper depths, whereas the number of species stayed relatively constant (**Figure 3B**). Patches dominated by hard substrata had the lowest DSCS cover at shallow and deep depths, with the peak DSCS cover between 200 and 250 m (**Figure 3B**). In contrast, DSCS height was greatest in soft substratum patches and declined with depth in all substratum types (**Figure 3C**). This was primarily the result of sea pens (Pennatulacea), which were among the tallest DSCS taxa observed (**Table 1**) and were abundant in the soft substratum patches. The density of fishes was greatest at shallow depths in hard substrata. As with DSCS height, fish density declined with depth (**Figure 3D**).

Cluster Analysis

The indicator species and "NbClust" methods suggested that our data was best described by four to six clusters (**Figure 4**). Clusters were defined by the density of DSCS taxa (matrix columns, $n = 32$) within each habitat unit (matrix rows, $n = 213$). Using the indicator species method, the maximum number of species was observed with either four or five clusters (**Figure 4A**). Eight of the indices from "NbClust" indicated that two clusters best describe our data, which we suspect was driven by difference between shallow and deep DSCS communities (**Figure 4B**). However, based on our ecological knowledge, we believe there are more differences in DSCS communities than those based simply on depth. The second mode from "NbClust" was at six clusters (**Figure 4B**), which was similar to that from the indicator species approach. Upon further examination, there was no ecological difference between the clusters formed when there was either four, five, or six clusters. The difference between four and five clusters was that some of the habitat units were removed from Cluster 1 and put into a separate cluster. However, this new cluster had no associated indicator species for fishes or DSCS, and therefore does not change our interpretation ecologically. The addition of a sixth cluster split the soft bottom habitat units into two separate clusters: one with thinner sea pens (Pennatulacea) as the indicator species and the other with thicker sea pens (Pennatulidae) as the indicator species. We think these two types of habitats were ecologically similar and should be in the same cluster. Therefore, we selected four clusters as most representative of the SCB.

Habitat clusters were primarily differentiated by their depth and substratum, and they contained a wide range of indicator species in different taxa (**Table 2**). Clusters 1 and 2 primarily consisted of hard or mixed substratum types found within the 100–300 m depth range (**Table 2**). Cluster 3 comprised hard and mixed substratum types in the shallowest depths (50 m). Fishes in

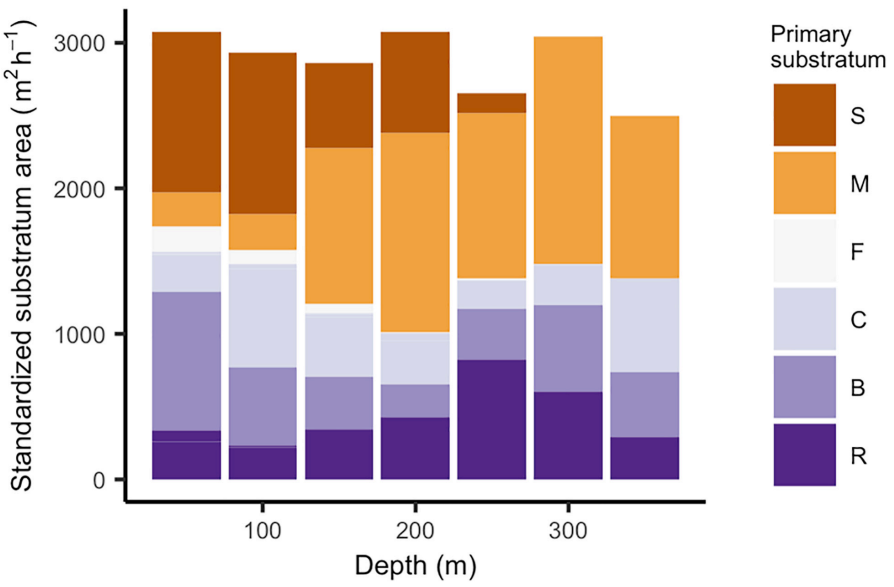

FIGURE 2 | Area of soft (orange hues) and hard (purple hues) primary substratum types within 50-m depth bins. We standardized substratum areas (m^2) by the number of survey hours (h) within each depth bin. Standardization was used because time at depth was variable. Substrata include sand (S), mud (M), cobble (C), boulder (B), rock (R). Three rarely observed substratum types were grouped with their closest substratum category based on relief. We categorized pinnacles with rock, and both pebbles and gravel with cobble.

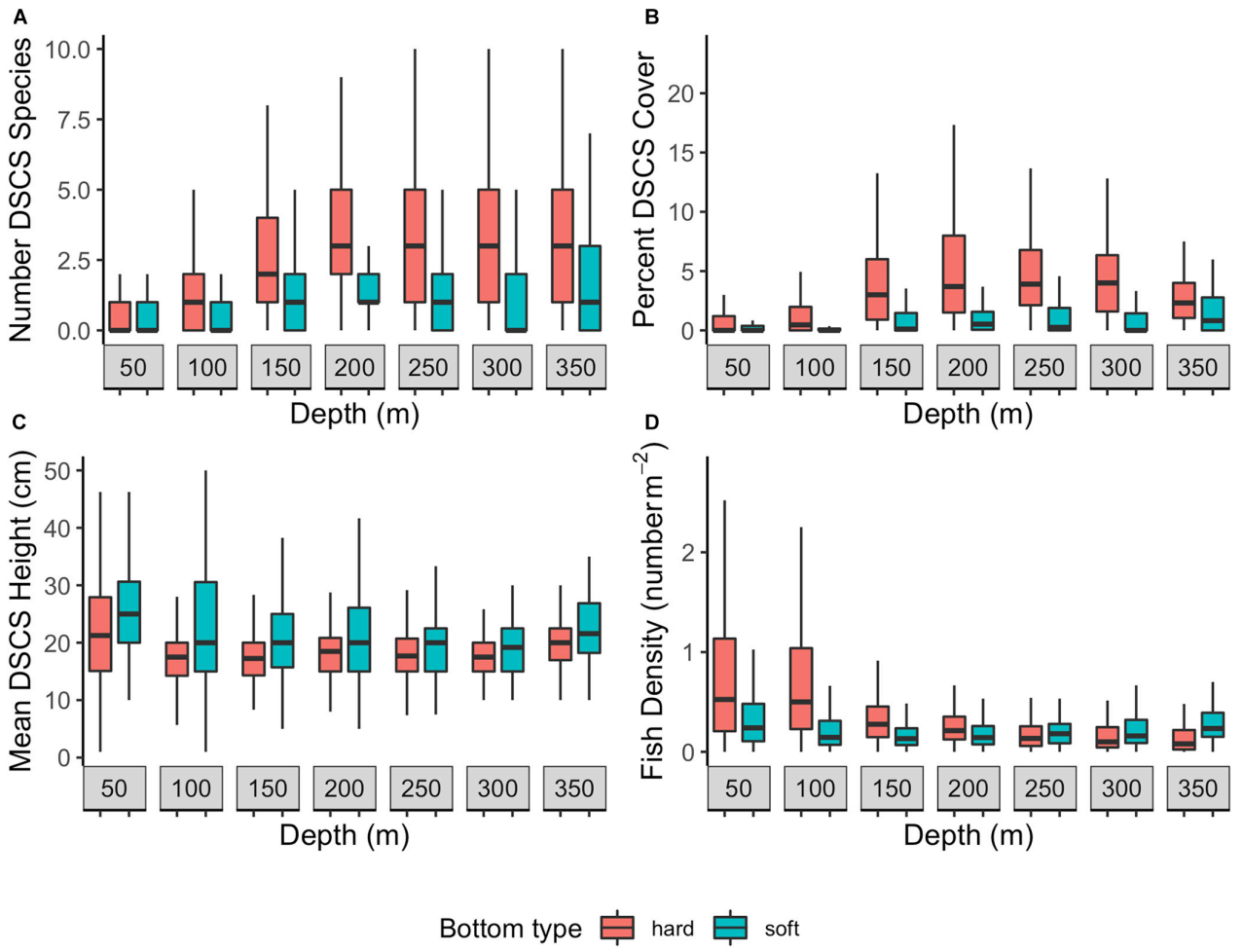

FIGURE 3 | The **(A)** number of Deep-sea coral and sponge (DSCS) species, **(B)** percent DSCS cover, **(C)** mean DSCS height, and **(D)** fish density by depth in habitats with hard (pinnacle rock, boulder, rugose rock, cobble, gravel, pebble, flat rock) and soft (mud and sand) primary substratum types.

the *Sebastes* genus dominated the significant indicator fish species in the first three clusters. The DSCS taxa in the first three clusters primarily were gorgonians from the order Alcyonacea. Notably, Cluster 2 also contained a single species of black coral and a single species of Scleractinian coral, *Antipathes dendrochristos* (Christmas tree coral) and *Lophelia pertusa* (white cup coral), respectively. Cluster 4 primarily comprised soft substrata at depths from 100 to 300 m. The fish indicator species in Cluster 4 were primarily flatfish, sculpins, combfish, eelpout, poachers, and pricklebacks. The two indicator DSCS species for this cluster were both corals commonly known as sea pens (order Pennatulacea).

The four clusters were defined both by the density of DSCS within the habitat units and the density of DSCS taxa within the cluster (**Figure 5**). Clusters 3 and 4 were clearly defined by their indicator DSCS taxa, whereas the indicator species in Clusters 1 and 2 were less dominant throughout the habitat units in those clusters. However, Cluster 2 had the most DSCS indicator species of any cluster. Cluster 3 included the shallowest habitat units and was well defined by five indicator DSCS taxa that were found in much higher abundance in the habitat units within this cluster than in any other habitat units: Plexauridae #2 (sea fan), *Placogorgia* sp. (primnoid), *Adelogorgia phyllosclera* (Orange gorgonian), *Eugorgia rubens* (Purple gorgonian), and *Leptogorgia chilensis* (Red gorgonian). Likewise, the soft bottom Cluster 4 was well defined by Pennatulidae and Pennatulacea, which are both sea pen taxa. In contrast, there were some sponges (Porifera #2, #3, and #5) that were found in both Clusters 1 and 2 (the deeper clusters with high substratum relief) in nearly equal abundances. **Table 1** provides descriptions of the observed size, depth, and abundances of these DSCS taxa.

Logistic Regression

From the logistic regression analysis, depth and coral height were the primary ecological covariates correlated with the increased probability of fish presence within habitat patches. The vast majority (43 of 45; 96%) of the fish taxa were better represented by models that included a spatial correlation term. Thus, fishes within these taxa were more likely to be present in patches closer to one another. Water depth was included as an important covariate in 76% (34 of 45) of the models

TABLE 1 | Summary of observations of DSCS and fish taxa mentioned throughout text.

Group	Taxon	Common name	Number observed (patches)	Density (num m^{-2})	Mean size (cm)	Mean Depth (m)
Coral	*Acanthogorgia* spp.	Gold coral	837 (82)	0.103	23.46	191
	Adelogorgia phyllosclera	Orange gorgonian	610 (42)	0.219	20.58	83
	Antipathes dendrochristos	Christmas tree coral	1,187 (464)	0.027	26.35	250
	Eugorgia rubens	Purple gorgonian	813 (99)	0.066	32.86	85
	Lophelia pertusa	White cup coral	554 (55)	0.085	8.19	197
	Pennatulacea	Thin sea pen	5,350 (338)	0.055	32.92	179
	Pennatulidae	Thick sea pen	41 (25)	0.003	23.05	133
	Placogorgia spp.	Primnoid	33 (26)	0.007	31.82	114
	Plexauridae #1	Sea fan (swiftia type)	876 (177)	0.050	22.52	203
	Plexauridae #2	Sea fan (swiftia type)	549 (53)	0.086	13.76	104
	Plumarella longispina	Primnoid	1,791 (383)	0.046	14.89	205
Sponge	*Farrea occa*	Lace foliose sponge	225 (115)	0.019	10.99	257
	Haliclona (gellius)	Trumpet sponge	625 (164)	0.031	15.30	149
	Porifera #1	Foliose sponge	4,353 (621)	0.072	12.87	188
	Porifera #2	Upright flat sponge	548 (264)	0.019	14.10	209
	Porifera #3	Barrel sponge	3,423 (935)	0.037	14.02	183
	Porifera #5	Vase sponge	895 (338)	0.025	14.10	220
	Rhabdocalyptus dawsoni	Brown barrel sponge	301 (151)	0.015	28.85	147
Fish	*Ophiodon elongatus*	Lingcod	255 (187)	0.011	46.29	111
	Sebastes chlorstictus	Greenspotted rockfish	815 (406)	0.018	20.56	117
	Sebastes constellatus	Starry rockfish	766 (422)	0.017	21.11	105
	Sebastes diploproa	Splitnose rockfish	1,349 (195)	0.063	19.41	345
	Sebastes ensifer	Swordspine rockfish	8,577 (1,014)	0.080	15.64	142
	Sebastes entomelas	Widow rockfish	570 (84)	0.052	24.75	113
	Sebastes hopkinsi	Squarespot rockfish	41,377 (898)	0.518	15.33	103
	Sebastes jordani	Shortbelly rockfish	3,841 (325)	0.087	18.29	253
	Sebastes levis	Cowcod	251 (201)	0.008	42.28	149
	Sebastes melanostomus	Blackgill rockfish	104 (65)	0.010	20.50	336
	Sebastes miniatus	Vermilion rockfish	408 (126)	0.026	34.8	97
	Sebastes ovalis	Speckled rockfish	701 (202)	0.032	26.45	109
	Sebastes paucispinis	Bocaccio	1,400 (390)	0.033	35.63	111
	Sebastes pinniger	Canary rockfish	9 (7)	0.008	36.67	101
	Sebastes rubrivinctus	Flag rockfish	156 (126)	0.008	20.91	122
	Sebastes rufus	Bank rockfish	2,682 (489)	0.056	25.09	230
	Sebastes semicinctus	Halfbanded rockfish	11,352 (671)	0.142	12.71	122
	Sebastes simulator	Pinkrose rockfish	1,680 (495)	0.029	18.13	219
	Sebastes umbrosus	Honeycomb rockfish	57 (39)	0.017	18.98	75
	Sebastes wilsoni	Pygmy rockfish	23,674 (802)	0.298	11.06	112
	Sebastes young-of-year	YOY rockfish	20,102 (624)	0.273	5.00	110
	Sebastolobus spp.	Thornyhead rockfish	215 (112)	0.016	20.05	356

The number observed is the total number of individuals observed, which is followed in parentheses by the number of patches in which at least one individual of that taxa was observed. The density is the number (num) per square meter within the observed habitats. The mean size is either the height (coral and sponges) or fork length (fish) of the measured taxa.

(**Supplementary Table 4**). As expected, the fish taxa could be categorized into fishes found in deeper water (positive depth odds ratio) or fishes in shallower water (negative depth odds ratio). DSCS height also was included in a large percentage (30 of 45; 67%) of fish taxa models. In contrast to water depth, all of the fishes had positive DSCS height odds ratios, indicating that fishes were more likely to be present in patches with taller-than-average DSCS. Interestingly, this correlation only became clear after we included the spatial correlation term in the models. The correlation between DSCS height and fish presence depended on which cluster the fishes were associated with. The vast majority of models (86%) for fishes in Cluster 4 included the DSCS height term, whereas only 36% of models for fishes in Cluster 2 included this term. Because Cluster 4 represents habitats with softer sediments without much relief except the exceptionally tall sea pens, it is intuitive that fishes in these habitats would more likely associate with taller corals. Although substratum was included in 58% of the logistic regression models, there was not an obvious pattern of how substratum was related to fish presence. Only 10 of the 45 fishes (22%) had models with a

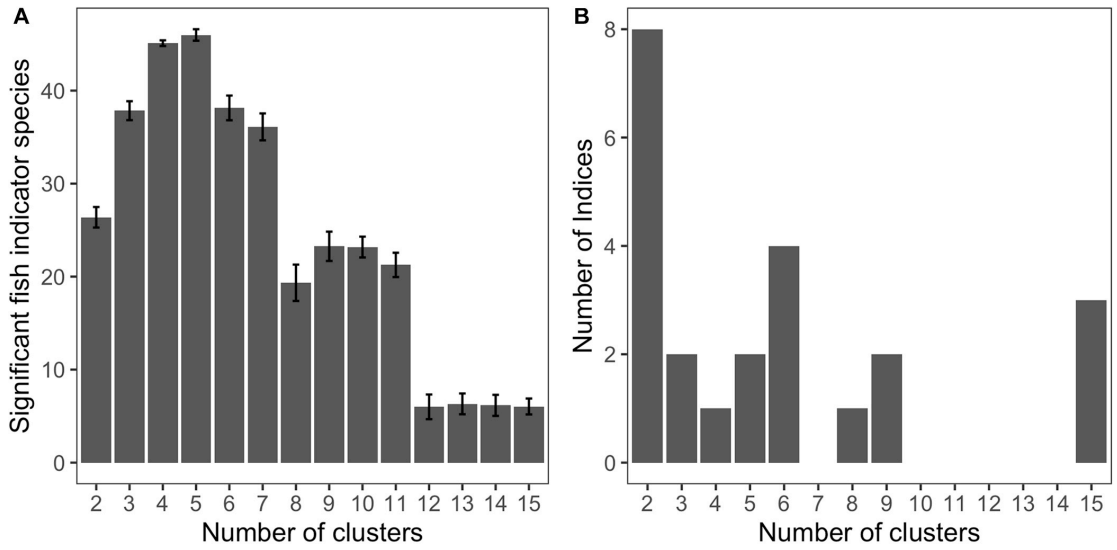

FIGURE 4 | (A) The number of significant indicator fish species as a function of the number of clusters. The error bars were calculated by a Monte Carlo resampling approach to determine which taxa are indicator species for each cluster. **(B)** A histogram of the number of indices from the NbClust library that selected various number of clusters as the most appropriate based on the habitat unit data.

TABLE 2 | Description of the fish and substratum types found within each of four DSCS clusters.

Cluster	Coral indicator species	Sponge indicator species	Rockfish indicator species	Other fish indicator species	Substratum type						Mean relief (m)	Mean depth (m)
					Rock	Boulder	Cobble	Flatrock	Sand	Mud		
1	1	0	1	0	10 (12)	16 (13)	30 (25)	15 (5)	5 (13)	10 (8)	0.18	139
2	7	16	9	3	24 (13)	20 (22)	16 (18)	1 (1)	4 (3)	14 (14)	0.26	217
3	5	0	13	5	4 (2)	6 (5)	4 (6)	3 (3)	2 (4)	7 (4)	0.12	84
4	2	0	2	12	0 (0)	0 (1)	0 (10)	0 (1)	16 (10)	16 (10)	0.00	159

Substratum type represents the number of habitat units with rock, boulder, cobble, flatrock, sand, and mud as the primary and secondary (in parentheses) substratum type. For the purposes of this summary table we categorized three rarely observed substratum types with their closest substratum category based on relief. We categorized Pinnacles with Rock, and both Pebbles and Gravel with Cobble.

positive odds ratio for substratum, implying that our measure of substratum relief was not a major driving factor in predicting fish presence for most taxa. There also were 16 taxa that had a negative substratum odds ratio, and 11 of those fishes (69%) were in Cluster 4. DSCS cover was only included in 22% of logistic regression models and had a negative odds ratio in all the models it was found in except one (**Supplementary Table 4**). As with substratum, the majority (67%) of fishes with logistic regression models that included DSCS cover were in Cluster 4.

A few key DSCS taxa were associated with fish presence, after accounting for depth, DSCS height, DSCS cover, and substratum (**Table 3**). The DSCS taxa that recurred in the most logistic regression models were sponges in the phylum Porifera. These sponges were important for multiple rockfish species (**Table 2**) as well as lingcod (*Ophiodon elongatus*) (**Supplementary Table 4**). Multiple coral taxa were positively associated with the increased probability of fish presence. In contrast to the sponge taxa, the coral taxa tended to be associated with only one or two fish taxa (**Table 2**). The results from the logistic regression models also suggest that the densities of DSCS taxa were less likely to affect the presence of rockfish taxa in deeper habitats. Most (11 of 14, 79%) of the rockfish taxa that had a negative depth odds ratio in their logistic regression models (i.e., they were more likely to be found in shallower depths) also had a positive association with the densities of at least one DSCS taxa. In contrast, none of the seven rockfish taxa that had positive depth odds ratios had a positive association with any DSCS taxa.

The top five rockfish taxa landed by pounds in commercial fisheries, and observed within at least 1% of patches in our dataset, were thornyhead (*Sebastolobus* spp.) (846,152 lbs), blackgill rockfish (*S. melanostomus*) (214,554 lbs), widow rockfish (*S. entomelas*) (152,604 lbs), bank rockfish (*S. rufus*) (137,770 lbs), and splitnose rockfish (*S. diploproa*) (98,722 lbs). Bank rockfish was the only one of these top commercially landed taxa that was positively associated with any DSCS taxa (**Table 3**). Both bank rockfish and widow rockfish were the only two of these five taxa found in shallower depths (i.e., negative depth odds ratios). As previously mentioned, the deeper rockfish taxa generally were not positively associated with DSCS taxa. The taxa of conservation interest we selected were young-of-year rockfish, cowcod, bocaccio, and canary

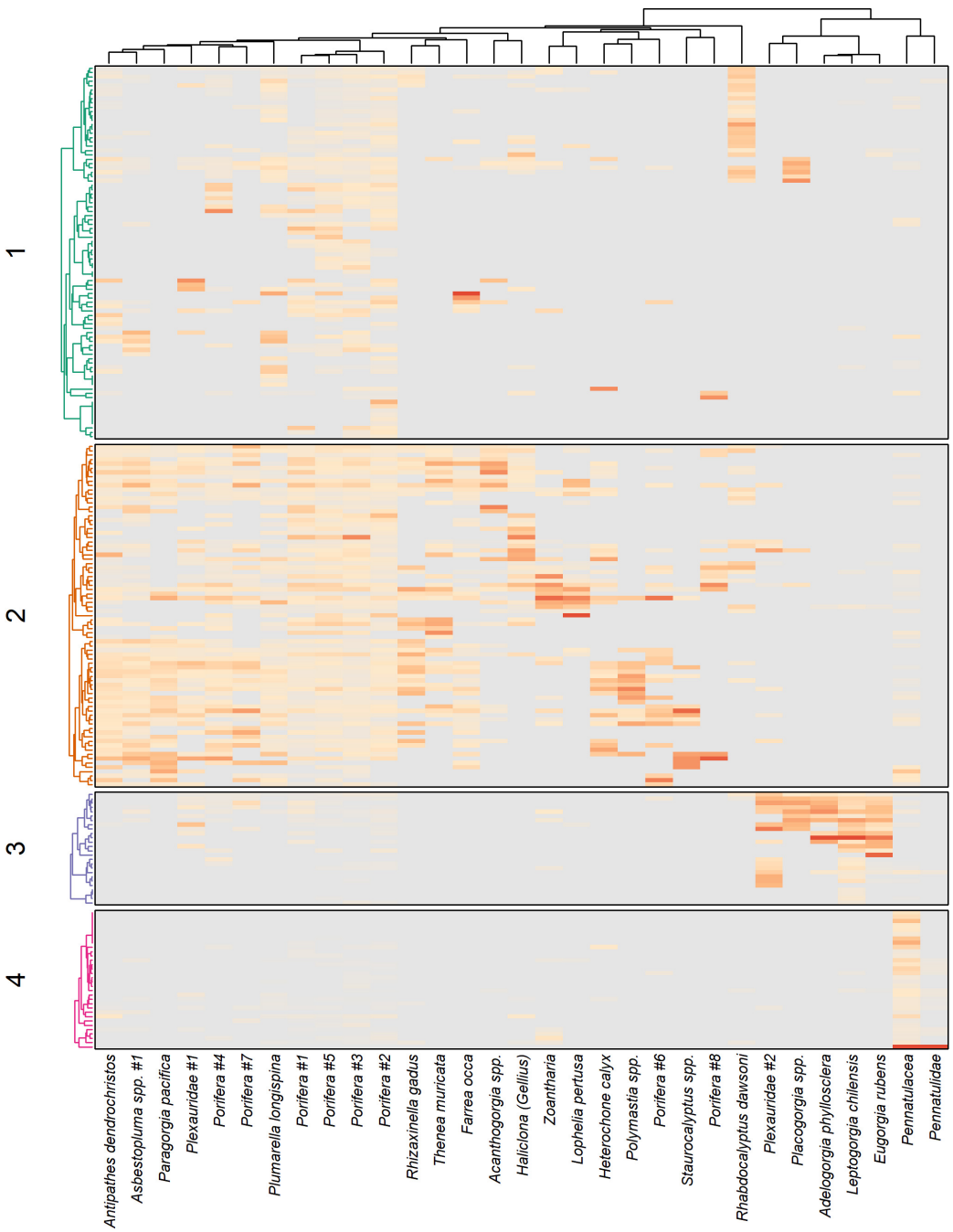

FIGURE 5 | Density (darker colors represent higher densities) of DSCS taxa within each of four clusters. Dendrograms on the left-hand side represent the classification of the habitat units (defined as the combination of primary and secondary substratum type together with 50 m depth bin). The large number of habitat units cannot be individually labeled, but the general classifications for each cluster are described in the text.

rockfish. We were unable to fit a reasonable model for canary rockfish using these ecological covariates. For the other three conservation taxa, DSCS height was positively correlated with the probability of fish presence, where substratum was only correlated with bocaccio (**Figure 6** and **Table 4**). Both young-of-year rockfish and cowcod were positively associated with multiple DSCS taxa (**Figure 6** and **Table 4**).

Gradient Forest

The gradient forest analysis generally supported the results from the logistic regression and indicated that depth and

TABLE 3 | Associated DSCS and *Sebastes* taxa and the percent increase in the probability of fish presence with a standard deviation increase in DSCS abundance based on the logistic regression models.

Type	DSCS species	Associated fish species	% increase
Coral	*Acanthogorgia* spp.	***Sebastes rufus***	15%
	Adelogorgia phyllosclera	*Sebastes miniatus*	12%
		Sebastes wilsoni	11%
		Sebastes umbrosus	10%
	Eugorgia rubens	*Sebastes chlorstictus*	11%
	Farrea occa	*Sebastes simulator*	14%
	Lophelia pertusa	***Sebastes* spp. *YOY***	7%
		Sebastes rufus	9%
	Plumarella longispina	***Sebastes* spp. *YOY***	9%
Sponge	*Haliclona (gellius)*	***Sebastes rufus***	12%
		Sebastes ensifer	12%
		Sebastes wilsoni	10%
	Plexauridae #1	*Sebastes semicinctus*	31%
	Porifera sp. *#1*	*Sebastes miniatus*	20%
		Sebastes ovalis	23%
		Sebastes wilsoni	29%
	Porifera sp. *#2*	***Sebastes* spp. *YOY***	15%
		Sebastes levis	16%
		Sebastes simulator	11%
		Sebastes constellatus	20%
		Sebastes hopkinsi	28%
		Sebastes wilsoni	35%
	Porifera sp. *#3*	***Sebastes* spp. *YOY***	9%
		Sebastes rufus	12%
		Sebastes levis	13%
		Sebastes miniatus	22%
		Sebastes rubrivinctus	14%
	Porifera sp. *#5*	*Sebastes jordani*	26%
	Rhabdocalyptus dawsoni	***Sebastes rufus***	24%
		Sebastes ensifer	7%

Sebastes taxa of management and conservation concern are identified in bold.

ecological covariates related to DSCS were the primary factors that influenced the biodiversity of demersal taxa throughout the SCB. Using the gradient forest method, the importance of each predictor can be evaluated based on their contributions to the accuracy importance (R^2) for each random forest and averaged across all species to provide an overall importance (see Ellis et al., 2012 for statistical details). Although we originally hypothesized that depth and substratum would be the strongest predictors of biodiversity, it was actually depth and percent DSCS cover that were the strongest predictors (**Figure 7**). Many of the same DSCS taxa that were associated with increased presence of fish taxa based on the logistic regression models also were associated with increased fish biodiversity based on the gradient forest. In fact, 6 of the top 10 DSCS taxa from the gradient forest (selected out of 39 potential DSCS taxa) were those also associated with increased probability of fish presence based on the logistic regression (**Table 3** and **Figure 7**).

Interesting patterns relative to depth and coral height were evident from a plot of the ratio of forest splits to observed data along the gradient of these variables (**Figure 8**). Locations on the gradient where splits density was greater than data density (**Figure 8**: blue line ratio > 1) indicate higher relative importance for species composition change (Pitcher et al., 2012). Note that because these values are standardized by the observed data, they represent the density of the random forest splits corrected for sampling bias. The depth results indicate that shallow depths (<100 m) and deeper depths (>250 m) have the greatest relative importance for species compositional change (**Figure 8A**). Likewise, DSCS taxa between 5 and 60 cm have the greatest relative importance for compositional change (**Figure 8C**). No clear patterns were apparent for percent DSCS cover or substratum.

The cumulative importance plots revealed varying levels of association between various rockfish taxa and each of the DSCS taxa (**Figure 9**). Some rockfish taxa were strongly associated with one DSCS taxon well beyond any of the other rockfish taxa. For example, bank rockfish and swordspine rockfish (*S. ensifer*) exhibited a strong association with *Acanthogorgia* spp. (gold coral). Although both fish taxa responded strongly to *Acanthogorgia* spp., it took larger densities of this DSCS taxa before the probability of bank rockfish presence increased compared with swordspine rockfish. This was indicated by a steep cumulative importance curve as the density of *Acanthogorgia* spp. increased, while most other fish taxa cumulative importance curves remained close to zero and relatively constant (**Figure 9A**). In contrast, the probability of fish presence increased for most fish taxa with increasing densities of Porifera #2 (**Figure 9B**). For most taxa, this increase occurred as Porifera #2 densities reached 0.05–0.15 individuals per m^2. Note there is an order of magnitude difference in the *y*-axis scale for the plots of these two DSCS taxa. This illustrates why Porifera #2 was ranked as the 4th most important DSCS taxa while *Acanthogorgia* spp. was the 19th most important DSCS taxa and does not even appear in **Figure 7**.

DISCUSSION

We described four communities of deep-sea coral, sponge, and fish assemblages in the SCB, and demonstrated that the density of DSCS taxa increased the probability of presence for multiple fish taxa and increased fish biodiversity. The results from two different analytical approaches indicated the same DSCS taxa were correlated with fish taxa in the SCB, which strongly suggests that these DSCS taxa play an important role in the ecosystem. From the logistic regression analysis, it was evident that increased densities of DSCS taxa increased the probability of at least three rockfish taxa of management and conservation interest, including young-of-year rockfish. These fish taxa occupy DSCS habitat potentially because DSCS provide benefits such as increased prey density (Quattrini et al., 2012), predation refuge (Krieger and Wing, 2002; Costello et al., 2005), and nursery habitat (Stone, 2006, 2014; Baillon et al., 2012).

Our finding that young-of-the-year rockfish are more likely to occur in habitat patches with taller DSCS supports the suggestion

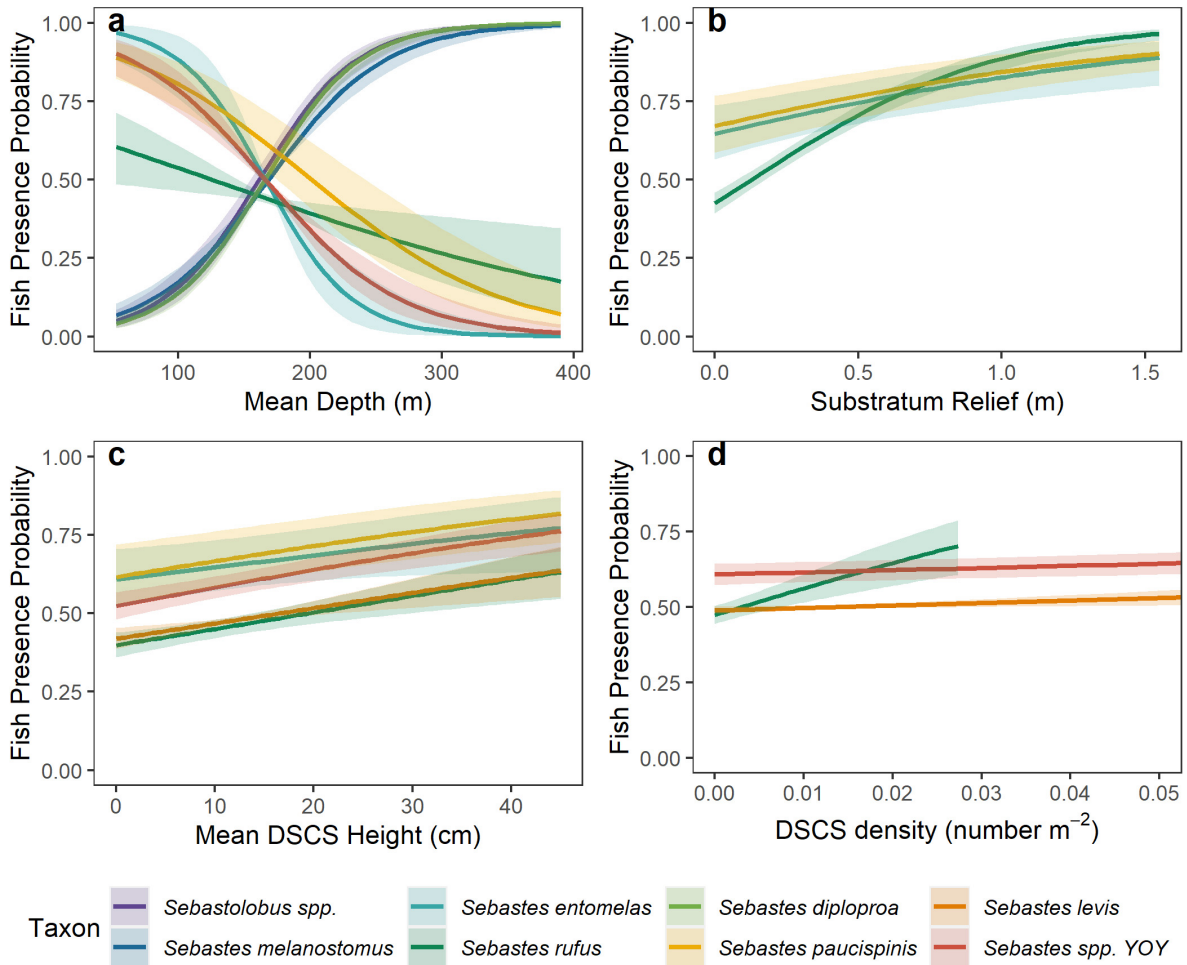

FIGURE 6 | Response plots of the probability of fish presence relative to **(A)** Mean Depth, **(B)** Substratum Relief, **(C)** Mean DSCS Height, and **(D)** DSCS density based on results from the covariates included in the logistic regression. No line is shown if a covariate was not included in the best model for that fish taxon. In **(D)** we only plot the relationship for DSCS taxa with the largest response greater than zero, because we were interested only in taxa that increased the probability of taxon presence.

that DSCS can provide important nursery habitats for these taxa (Edinger et al., 2007; Harter et al., 2009; Baillon et al., 2012). Specifically, our results suggest that young-of-the-year rockfish were more likely to be present in habitat patches with increased densities of *Lophelia pertusa* (7%), *Plumarella longispina* (9%), Porifera #2 (15%), and *Porifera* #3 (9%). In addition to the association with these specific taxa, the model also indicated that young-of-the-year rockfish were more likely to be present in patches with taller corals. Baillon et al. (2012) also observed Atlantic *Sebastes* larvae associated with deep-sea coral, implying that they use these habitats as nursery grounds. Multiple authors have noted the presence of gravid *Sebastes* near *Lophelia* reefs (Fosså et al., 2000 cited in Husebø et al., 2002; Costello et al., 2005), suggesting they may release their young near these reefs. However, we note that *Lophelia* has a much different reef-forming pattern in the Atlantic (where these previous studies were conducted) than in the Pacific so it is best not to assume that fishes are using these corals in the same way. The use of DSCS by gravid *Sebastes* may be due to the protection these corals provide from predators (Krieger and Wing, 2002; Costello et al., 2005) or the additional feeding opportunities, because researchers anecdotally have noted that zooplankton abundances are higher near DSCS (Costello et al., 2005). Although our study cannot establish why young-of-year rockfish are using these DSCS habitats, our results imply that DSCS are important to these fishes growth to maturity, which supports the classification of DSCS as essential fish habitat (Rosenberg et al., 2000). Similarly, a modeling study in the Northeast Atlantic (Foley et al., 2010) also concluded that *Sebastes* population dynamics were important to the intrinsic growth rate of the stock. Their results were consistent with corals serving as essential fish habitat and suggested that coral removal would result in the decline, and potential extirpation, of some *Sebastes* populations.

Our analytical approach also identified some specific DSCS taxa that were associated with demersal fishes, and we suggest species distribution maps should be developed for these taxa

TABLE 4 | Results of logistic regression showing the percent increase (black), or decrease (red), in the probability of fish presence with a standard deviation increase in each covariate.

Taxon (common name)	*n*	Depth	Substratum	DSCS cover	DSCS height	DSCS density	AUC
Sebastolobus spp. (Thornyhead)	145	911 (579, 1,471)		−36 (−56, −9)			0.96 (0.95, 0.97)
Sebastes melanostomus (Blackgill rockfish)	84	589(383, 913)					0.96(0.95, 0.97)
Sebastes entomelas (Widow rockfish)	92	−91 (−97, −75)	33(14, 53)		20(−3, 46)		0.89(0.86, 0.92)
Sebastes spp. YOY (young-of-year rockfish)	722	−81 (−88, −70)			33(20, 48)	*Plumarella*: 9 (−1, 19) *Lophelia*: 7 (−1, 15) Porifera #2: 15 (4, 27) Porifera #3: 9 (0, 18)	0.87(0.86, 0.88)
Sebastes rufus (Bank rockfish)	638	−39 (−57, −13)	99(78, 123)		29(14, 45)	*Acanthogorgia*: 15 (7, 25) *Lophelia*: 9 (1, 19) *Haliclona (gellius)*: 12 (1, 24) *Plumarella*: −11 (−21, −2) *Antipathes*: −8 (−15, 0) *Rhabdocalyptus*: 24 (13, 36) Porifera #3: 12 (4, 21)	0.90(0.89, 0.91)
Sebastes paucipinis (Bocaccio)	460	−69 (−78, −57)	32(22, 43)		31(16, 48)	Pennatulacea: −99 (−100, −78)	0.84(0.83, 0.86)
Sebastes diploproa (Splitnose rockfish)	294	962(681, 1,396)				Porifera #5: −24 (−46, 1)	0.97(0.97, 0.98)
Sebastes levis (Cowcod)	230				27(11, 44)	Porifera #2: 16 (3, 29) Porifera #3: 13 (3, 22)	0.79(0.76, 0.81)

These are calculated by subtracting 1 from the logistic regression odds ratio (exponentiated model coefficients) and multiplying by 100. The upper and lower 95% credible interval for each estimate is shown in parentheses under, or next to, each estimate. The value n is the number of habitat patches (out of 5,144 patches) where a fish taxon was observed. The area under the curve (AUC) is a measure of model fit (see text).

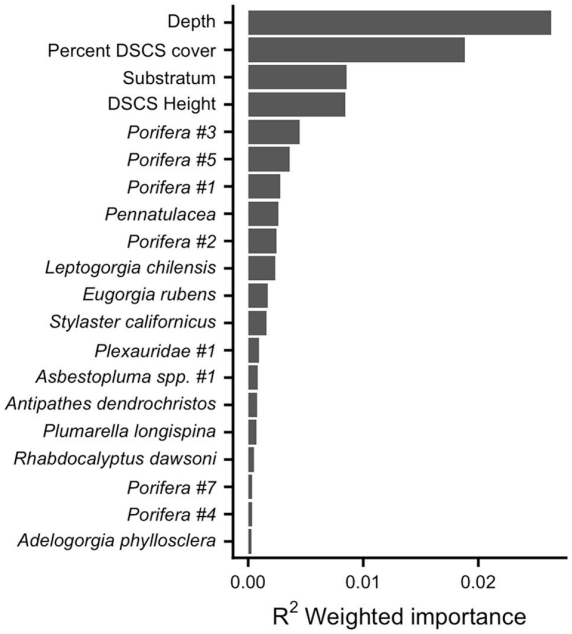

FIGURE 7 | Gradient forest estimated relative importance of physical and biological covariates for predicting biodiversity of fish taxa in the SCB. Note that for clarity we have displayed only the top 20 covariates out of a total of 43.

to ensure they are protected from future damaging human practices (e.g., benthic trawling). Deep-sea coral taxa are slow growing, so it can take a long time for them to recover once they have been damaged (Roberts et al., 2006; Althaus et al., 2009). Consequently, it is important to identify locations where these taxa may be found in the highest densities, validate their presence, and provide protection to these areas before any further damage is inflicted. Species distribution models are one method to identify the areas where taxa are expected to be found based on sample observations. Species distribution models have been developed for multiple DSCS taxa to examine the factors that influence habitat suitability at a variety of scales. Multiple authors have developed species distribution models to predict global habitat suitability for various DSCS taxa (Tittensor et al., 2009; Davies and Guinotte, 2011; Yesson et al., 2012). In general, these models have predicted that the majority of suitable coral habitat was on the continental shelves and slopes of the Atlantic, South Pacific, and Indian Oceans as well as seamounts along the northern Mid-Atlantic Ridge and in the South Pacific Ocean. Models developed at finer scales are perhaps more useful to identify areas that could be protected from anthropogenic practices that may potentially damage these fragile DSCS taxa (Rengstorf et al., 2013; Gullage et al., 2017; Sundahl et al., 2020). For example, Huff et al. (2013) found that the distribution of the Christmas tree coral in the SCB (*Antipathes dendrochristos*) was positively affected by a combination of persistently high surface productivity, water current velocity and direction near the seafloor, warmer temperature and shallow depth. Developing similar distribution maps for the other DSCS taxa associated with demersal fishes would be invaluable to fisheries managers seeking to protect these vulnerable fish habitats as they seek to rebuild overfished populations (Rosenberg et al., 2000).

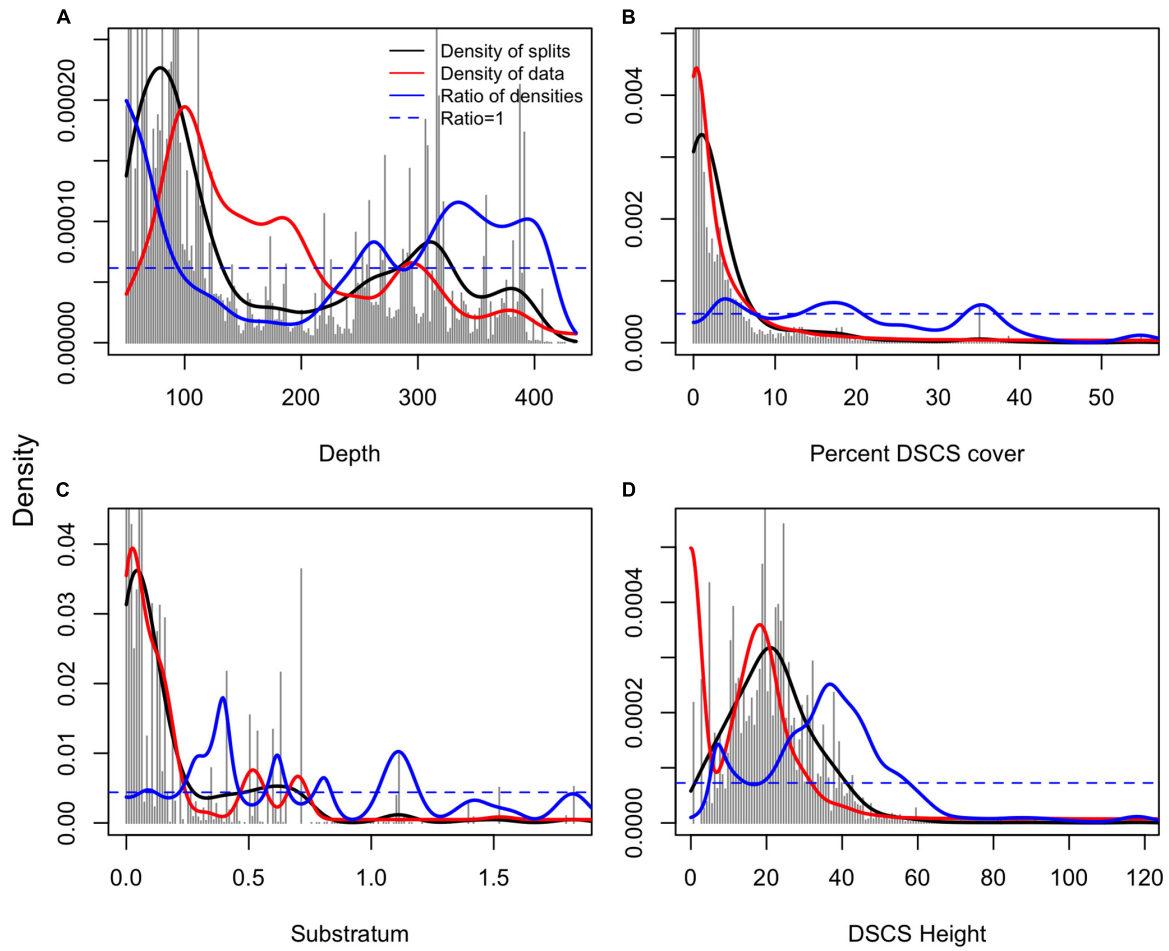

FIGURE 8 | Gradient forest output showing the locations of the random forest splits (gray bars), densities of splits (black line), densities of observations (red line), and the ratio of splits standardized by observation density (blue line). Ratio > 1 indicate locations of relative greater change in composition. Note that the panels are on different scales based on the number of bins the data is split between for each variable.

The logistic regression analysis indicated that the probability of fishes being associated with DSCS was primarily related to depth, and we did not find that this association was as highly related to substratum relief as we anticipated. All the fish taxa that had positive relationships with DSCS taxa were found in depths shallower than 230 m, while all the fish taxa at mean depths greater than 300 m generally had no, or negative, relationships with DSCS taxa. One explanation for this is that at deeper depths DSCS cover decreases (**Figure 3**). Therefore, we are less likely to see fishes and DSCS on the same transect, or in the same patch, unless we have a sufficiently high sample size. Unfortunately, the fish taxa with deeper mean depths had smaller sample sizes, thus we cannot say if there was an ecological reason why fishes at deeper depths were not associated with DSCS taxa (e.g., reduced need for predation refuge at depth) or if this was an artifact due to reduced sampling effort at deeper depths. We suspect that the probability of fish presence was correlated with relief for only a few taxa was likely due to the way we calculated relief from visual estimates of the primary and secondary substratum. Future surveys should consider simultaneously collecting bathymetry data (e.g., side-scan or multibeam sonar) to better correlate relief to fishes and invertebrate habitat.

Results from the gradient forest analysis revealed both expected and surprising relationships between fish biodiversity and various physical and biological covariates. Depth and DSCS cover had the largest influence on rockfish biodiversity in habitat patches. It was not surprising that depth was important, as we know that various taxa have specific depth preferences. Based on the results of the logistic regression, it was surprising that DSCS cover was considerably more important than either DSCS height or substratum type. This may have been due to the taxa that were included in this analysis (all 111 fish taxa were included in the gradient forest biodiversity analysis while just the significant indicator species were included in the logistic regression analysis), the fact that the gradient forest did not explicitly account for spatial autocorrelation as was done in the logistic regression, or simply that the two analyses were measuring different responses (univariate vs. multivariate). In any case, it is intuitive that fish diversity would increase as coral cover increases, as fishes generally associate with habitats having

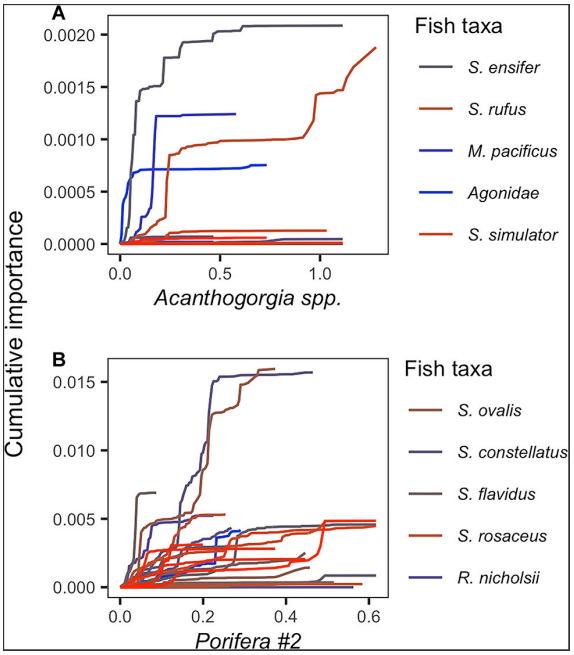

FIGURE 9 | Cumulative importance distributions of standardized random forest splits for each fish taxa (individual lines) along the observed density gradient for **(A)** *Acanthogorgia* spp. and **(B)** Porifera #2.

increased cover. It also is noteworthy that six of the top ten DSCS taxa selected from the gradient forest analysis (out of 39 possible taxa) also increased the probability of rockfish presence based on the results from the logistic regressions (**Table 3**). This suggests that some of the other DSCS taxa that were near the top of the gradient forest analysis (e.g., *Pennatulacea* and *Antipathes dendrochristos*) also may have important habitat roles for other taxa that were not included in our logistic regression models. Habitats with persistent localized upwelling resulting from complex seafloor topography are areas where *Antipathes dendrochristos* are denser (Huff et al., 2013), and also may comprise the most important benthic habitats for other DSCS and fishes in the SCB.

As with any survey method, there are weaknesses of using submersible observations to record species data that can potentially bias results. Behavioral reactions of fishes to submersibles have been documented, including avoidance, attraction, and no reaction (Stoner et al., 2008; Laidig et al., 2012; Sward et al., 2019). As Sward et al. (2019) state in their review of ROV surveys for visually assessing fish assemblages: "the type and severity of the reaction to the ROV can be influenced by a variety of factors, including the species, trophic position, and the body size and position of the individual relative to the seafloor as well as to different aspects of the ROV system (i.e., artificial lighting, thruster noise, speed)." Laidig et al. (2012) found that a smaller percentage of fishes (11%) reacted to the larger, manned submersibles (as we used in this study) than to ROVs (57%). Those fishes that did react to the manned submersibles tended to be smaller fishes, suggesting that it is more difficult to accurately count these smaller fishes (Laidig et al., 2012). Likewise, it is likely that the observers overlooked cryptic species that were able to hide among rocks, DSCS, and sediment. We would expect this bias to increase relative to the complexity of the habitat, although we are unaware of any studies that have conducted experiments to quantify this potential bias. Stoner et al. (2008) qualitatively noted that the reaction of most rockfish species was relatively low and concluded that bias was probably minimal. Finally, although the distribution of many species is influenced by time of day (Hart et al., 2010), our study only examined movement during the day. In their review, Sward et al. (2019) found that very few submersible studies (∼2%) were conducted at night. Telemetry studies have indicated that home ranges and behavior of Pacific rockfishes change both diurnally and seasonally (Tolimieri et al., 2009; Zhang et al., 2015). A quantitative comparison of diel habitat use using submersible video surveys would be an excellent future study.

Another area of future research is understanding the trophic dynamics of these DSCS habitats. Our results indicate that DSCS provide important habitat for multiple rockfish taxa, but we cannot identify what functional benefit these structure-forming invertebrates provide to the associated fishes. These DSCS may be found in areas where the hydrodynamics enhance the density of zooplankton and other potential prey items, some of which may be reliant on the DSCS (Husebø et al., 2002; George et al., 2007; Lessard-Pilon et al., 2010; Huff et al., 2013). Based on simplified trophic ecosystem models for deep-sea coral reef ecosystems, George et al. (2007) concluded that the degradation of corals and sponges would negatively impact populations of commercially important fish species. Thus, to understand the functional benefit of DSCS to fish populations, it would be valuable to compare diets of fishes in areas of high DSCS density and nearby habitats that have lower DSCS densities, such as those that have been disturbed by trawl fisheries.

Furthermore, future research could improve upon our results by incorporating a temporal component to the associations between fishes and DSCS taxa and by incorporating a measure of fishing impact as an additional covariate. Although DSCS generally are long-lived, and thus large changes in their distribution would not be expected over a short time scale, regional climatic variations (e.g., El Nino and ocean warming) can affect fish recruitment and distribution. These changes in the distribution and abundance of fishes could influence the interpretation of the observed associations between fishes and DSCS. Additionally, acute and rapid change in DSCS distribution can be caused by the impacts of benthic trawling (Yoklavich et al., 2018). For example, Clark and Rowden (2009) found differences in macro-invertebrate assemblages between fished and unfished seamounts in New Zealand. Future development of habitat suitability models could include amount of benthic trawling as a measure of habitat alteration. Continuing to collect long-term datasets of these deep-sea habitats and associated assemblages will help to understand the ecological importance of DSCS.

This study has provided evidence of the importance of DSCS as habitat for multiple taxa of fishes, including some with commercial importance, and re-enforces the importance of conserving these important structure-forming invertebrates. Previous research on structure-forming

deep-sea invertebrates primarily focused on larger species, and our results highlight the importance of sponges that are generally overlooked as habitat forming invertebrates. Sponges are often the largest structure-forming invertebrates in their associated habitats, and provide considerable biotic complexity, predator refuge, and enhanced food supply (Tissot et al., 2006; Buhl-Mortensen et al., 2010).

DSCS taxa throughout the world's seas are threatened by multiple factors. The impacts of fishing gear, primarily benthic trawling, have been documented on deep-sea reefs along the West Ireland continental shelf break (Hall-Spencer et al., 2002), Norway (Fosså et al., 2002), Tasmania (Koslow et al., 2001), and Alaska (Krieger and Wing, 2002; Heifetz et al., 2009). DSCS also are threatened due to climate change. A recent study estimated there would be no suitable habitat for deep-sea coral by 2099 assuming an upper temperature tolerance of 7°C (Thresher et al., 2015). Likewise, ocean acidification due to an increasing production of anthropogenic CO_2 has resulted in declining aragonite and calcite saturation states, which may impair the ability of DSCS taxa to build sufficiently robust skeletons (Guinotte et al., 2006). In the face of these potential threats, further conservation efforts are essential to protect these ecologically important DSCS.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because it was an observation only study.

AUTHOR CONTRIBUTIONS

MY contributed the data and essential background for the study. MH and DH conceived the analysis. MH conducted the statistical analysis and prepared result figures. All authors conceived the study, contributed to writing, editing, and approved the submitted manuscript.

FUNDING

The authors would like to acknowledge the NMFS Office of Habitat Conservation Deep Sea Coral Research and Technology Program for funding portions of this research.

ACKNOWLEDGMENTS

We thank Tom Laidig, Linda Snook, and Mary Nishimoto for analyzing the visual surveys, and Tom Laidig and Diana Watters for maintaining the databases used in this study. We also thank Tom Hourigan, Chris Rooper, and four reviewers for reviewing early drafts of the manuscript and providing comments that greatly improved the final product. We appreciate the assistance of many experts in the identification of deep-sea corals and sponges. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2020.593844/full#supplementary-material

REFERENCES

Addamo, A. M., Vertino, A., Stolarski, J., García-Jiménez, R., Taviani, M., and Machordom, A. (2016). Merging scleractinian genera: the overwhelming genetic similarity between solitary Desmophyllum and colonial Lophelia. *BMC Evol. Biol.* 16:108. doi: 10.1186/s12862-016-0654-8

Althaus, F., William, A., Schlacher, T. A., Kloser, R. J., Green, M. A., Barker, B. A., et al. (2009). Impacts of bottom trawling on deep-coral ecosystems of seamounts are long-lasting. *Mar. Ecol. Prog. Ser.* 397, 279–294. doi: 10.3354/meps08248

Auster, P. J. (2005). "Are deep-water corals important habitats for fishes?," in *Cold-water corals and ecosystems*, eds A. Freiwald and J. M. Roberts (Berlin: Springer), 747–760. doi: 10.1007/3-540-27673-4_39

Baillon, S., Hamel, J.-F., Wareham, V. E., and Mercier, A. (2012). Deep cold-water corals as nurseries for fish larvae. *Front. Ecol. Environ.* 10, 351–356. doi: 10.1890/120022

Belbin, L., Faith, D. P., and Milligan, G. W. (1992). A comparison of two approaches to beta-flexible clustering. *Multivar. Behav. Res.* 27, 417–433. doi: 10.1207/s15327906mbr2703_6

Bowden, D. A., Rowden, A. A., Leduc, D., Beaumont, J., and Clark, M. R. (2016). Deep-sea seabed habitats: Do they support distinct mega-epifaunal communitites that have different vulnerabilities to anthropogenic disturbance? *Deep Sea Res. I* 107, 31–47. doi: 10.1016/j.dsr.2015.10.011

Breiman, L. (2001). Random forests. *Machine Learning* 45, 5–32.

Buhl-Mortensen, L., Vareusel, A., Gooday, A. J., Levin, L. A., Priede, I. G., Buhl-Mortensen, P., et al. (2010). Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. *Mar. Ecol.* 31, 21–50. doi: 10.1111/j.1439-0485.2010.00359.x

Charrad, M., Ghazzali, N., Bioteau, V., and Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *J. Statist. Soft.* 61, 1–36.

Clark, M. R., and Rowden, A. A. (2009). Effect of deepwater trawling on the macro-invertebrate assemblages of seamounts on the Chatham Rise. *N Z. Deep Sea Res. I* 56, 1540–1554. doi: 10.1016/j.dsr.2009.04.015

Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna: R Core Team.

Costello, M. J., McCrea, M., Freiwald, A., Lundalv, T., Jonsson, L., Brett, B. J., et al. (2005). "Role of cold-water Lophelia pertusa coral reefs as fish habitat in the NE Atlantic," in *Cold-water corals and ecosystems*, eds A. Freiwald and J. M. Roberts (Berlin: Springer), 771–805. doi: 10.1007/3-540-27673-4_41

D'Onghia, G., Maiorana, P., Sion, L., Giove, A., Capezzuto, F., Carlucci, R., et al. (2010). Effects of deep-water coral banks on the abundance and size structure of the megafauna in the Mediterranean Sea. *Deep Sea Res. II* 57, 397–411. doi: 10.1016/j.dsr2.2009.08.022

D'Onghia, G., Maiorano, P., Carlucci, R., Capezzuto, F., Carluccio, A., Tursi, A., et al. (2012). Comparing deep-sea fish fauna between coral and non-coral 'megahabitats' in the Santa Maria di Leuca cold-water coral province (Mediterranean Sea). *PLoS One* 7:e44509. doi: 10.1371/journal.pone.0044509

Dailey, M. D., Reish, D. J., and Anderson, J. W. (1993). *Ecology of the Southern California Bight: a synthesis and interpretation*. Berkeley, CA: University of California Press.

Davies, A. J., and Guinotte, J. M. (2011). Global habitat suitability for framework-forming cold-water corals. *PLoS One* 6:e18483. doi: 10.1371/journal.pone.0018483

De Caceres, M., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

Du Preez, C., and Tunnicliffe, V. (2011). Shortspine thornyhead and rockfish (Scorpaenidae) distribution in response to substratum, biogenic structures and trawling. *Mar. Ecol. Prog. Ser.* 425, 217–231. doi: 10.3354/meps09005

Dufrene, M., and Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* 67, 345–366. doi: 10.2307/2963459

Edinger, E. N., Wareham, V. E., and Haedrich, R. L. (2007). Patterns of groundfish diversity and abundance in relation to deep-sea coral distributions in Newfoundland and Labrador waters. *Bull. Mar. Sci.* 81, 101–122.

Ellis, N., Smith, S. J., and Pitcher, R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93, 156–168. doi: 10.1890/11-0252.1

Foley, N. S., Kahui, V., Armstrong, C. W., and van Rensburg, T. M. (2010). Estimating linkages between redfish and cold water coral on the Norwegian coast. *Mar. Resour. Econom.* 25, 105–120. doi: 10.5950/0738-1360-25.1.105

Fosså, J. H., Mortensen, P. B., and Furevik, D. M. (2000). Lophelia-korallrev langs norskekysten. *Forekomst og tilstand. Fisken og Havet* 2:94.

Fosså, J. H., Mortenson, P. B., and Furevik, D. M. (2002). The deep-water coral Lophelia pertusa in Norwegian waters: distribution and fishery impacts. *Hydrobiologia* 471, 1–12.

Freese, J. L., and Wing, B. (2003). Juvenile red rockfish. Sebastes sp., associations with sponges in the Gulf of Alaska. *Mar. Fish. Rev.* 65, 38–42.

Georganos, S., Grippa, T., Gadiaga, A. N., Linard, C., Lennert, M., Vanhuysse, S., et al. (2019). Geographical random forest: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* 2019, 1–16. doi: 10.1080/10106049.2019.1595177

George, R. Y., Okey, T. A., Reed, J. K., and Stone, R. P. (2007). "Ecosystem-based fisheries management of seamount and deep-sea coral reefs in U.S. waters: conceptual models for proactive decisions," in *Conservation and adaptive management of seamount and deep-sea coral ecosystems*, eds R. Y. George and S. D. Cairns (Florida, FL: University of Miami).

Green, K. M., Greenley, A. P., and Starr, R. M. (2014). Movements of blue rockfish (Sebastes mystinus) off Central California with comparisons to similar species. *PLoS One* 9:e98976. doi: 10.1371/journal.pone.0098976

Greene, G., Yoklavich, M. M., Starr, R. M., O'Connell, V. M., Wakefield, W. W., Sullivan, D. E., et al. (1999). A classification scheme for deep seafloor habitats. *Oceanol. Acta* 22, 663–678. doi: 10.1016/s0399-1784(00)88957-4

Guinotte, J., Orr, J., Cairns, S., Freiwald, A., Morgan, L., and George, R. (2006). Will human-induced changes in seawater chemistry alter the distribution of deep-sea scleractinian corals? *Front. Ecol. Environ.* 4, 141–146. doi: 10.1890/1540-92952006004[0141:WHCISC]2.0.CO;2

Gullage, L., Devillers, R., and Edinger, E. (2017). Predictive distribution modelling of cold-water corals in the Newfoundland and Labrador region. *Mar. Ecol. Prog. Ser.* 582, 57–77. doi: 10.3354/meps12307

Hall-Spencer, J. M., Allain, V., and Fosså, J. (2002). Trawling damage to Northeast Atlantic ancient coral reefs. *Proc. R. Soc. B Biol. Sci.* 269, 507–511. doi: 10.1098/rspb.2001.1910

Hart, T. D., Clemons, J. E. R., Wakefield, W. W., and Heppell, S. S. (2010). Day and night abundance, distribution, and activity patterns of demersal fishes on Heceta Bank. *Oregon. Fish. Bull.* 108, 466–477.

Harter, S. L., Ribera, M. M., Shepard, A. N., and Reed, J. K. (2009). Assessment of fish populations and habitat on Oculina Bank, a deep-sea coral marine protected area off eastern Florida. *Fish. Bull.* 107, 195–206.

Heifetz, J., Stone, R., and Shotwell, S. K. (2009). Damage and disturbance to coral and sponge habitat of the Aleutian Archipelago. *Mar. Ecol. Prog. Ser.* 397, 295–303. doi: 10.3354/meps08304

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statist. Sci.* 14, 382–401.

Hosmer, D. W. Jr., Lemshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*, 3rd Edn, New Jersey. John Wiley & Sons, Inc.

Huff, D. D., Yoklavich, M. M., Love, M. S., Watters, D. L., Chai, F., and Lindley, S. T. (2013). Environmental factors that influence the distribution, size, and biotic relationships of the Christmas tree coral Antipathes dendrochristos in the Southern California Bight. *Mar. Ecol. Prog. Ser.* 494, 159–177. doi: 10.3354/meps10591

Husebø, A., Nøttestad, L., Fosså, J. H., Furevik, D. M., and Jørgensen, S. B. (2002). Distribution and abundance of fish in deep-sea coral habitats. *Hydrobiologia* 471, 91–99.

Jorgensen, S. J., Kaplan, D. M., Klimley, A. P., Morgan, S. G., O'Farrell, M. R., and Botsford, L. W. (2006). Limited movement in blue rockfish Sebastes mystinus internal structure of home range. *Mar. Ecol. Prog. Ser.* 327, 157–170. doi: 10.3354/meps327157

Knudby, A., Brenning, A., and LeDrew, E. (2010). New approaches to modelling fish-habitat relationships. *Ecol. Modell.* 221, 503–511. doi: 10.1016/j.ecolmodel.2009.11.008

Koslow, J. A., Gowlett-Holmes, K., Lowry, J. K., O'Hara, T., Poore, G. C. B., and Williams, A. (2001). Seamount benthic macrofauna off southern Tasmania: community structure and impacts of trawling. *Mar. Ecol. Prog. Ser.* 213, 111–125. doi: 10.3354/meps213111

Krieger, K. J., and Wing, B. L. (2002). Megafauna associations with deepwater corals (Primnoa spp.) in the Gulf of Alaska. 1st International Deep-Sea Coral Symposium. *Hydrobiologica* 471, 83–90.

Laidig, T. E., Krigsman, L. M., and Yoklavich, M. M. (2012). Reactions of fishes to two underwater survey tools, a manned submersible and a remotely operated vehicle. *Fish. Bull.* 111:67.

Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716

Lessard-Pilon, S. A., Podowski, E. L., Cordes, E. E., and Fisher, C. R. (2010). Megafauna community composition associated with Lophelia pertusa colonies in the Gulf of Mexico. *Deep Sea Res. II* 57, 1882–1890. doi: 10.1016/j.dsr2.2010.05.013

Lindgren, F., and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *J. Statist. Soft.* 63, 1–25. doi: 10.1002/9781118950203.ch1

Love, M. S. (2006). "Subsistence, Commercial, and Recreational Fisheries. 567-594," in *The Ecology of Marine Fishes: California and Adjacent Waters*, eds Allen, Horn, and Pondella (California: University of California Press).

Love, M. S., Yoklavich, M., and Schroeder, D. M. (2009). Demersal fish assemblages in the Southern California Bight based on visual surveys in deep water. *Environ. Biol. Fishes* 84, 55–68. doi: 10.1007/s10641-008-9389-8

Love, M. S., Yoklavich, M., and Thorsteinson, L. (2002). *The Rockfishes of the Northeast Pacific*. California, CA: University of California Press, 405.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2017). *cluster: Cluster analysis basics and extensions. R package version 2.0.6*.

McCune, B., Grace, J. B., and Urban, D. L. (2002). *Analysis of ecological communities*. Gleneden Beach, OR: MJM Software Design.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017). *vegan: Community Ecology Package. R package version 2.4-5*. Available Online at: https://CRAN.R-project.org/package=vegan

Pitcher, C. R., Lawton, P., Ellis, N., Smith, S. J., Incze, L. S., Wei, C.-L., et al. (2012). Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. *J. Appl. Ecol.* 49, 670–679. doi: 10.1111/j.1365-2664.2012.02148.x

Prasad, A. M., Iverson, L. R., and Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199. doi: 10.1007/s10021-005-0054-1

Quattrini, A. M., Ross, S. W., Carlson, M. C. T., and Nizinski, M. S. (2012). Megafaunal-habitat associations at a deep-sea coral mound off North Carolina. *U S A. Mar. Biol.* 159, 1079–1094. doi: 10.1007/s00227-012-1888-7

Reese, E. (1989). Orientation behavior of behavior of butterflyfishes (family Chaetodontidae) on coral reefs: spatial learning of route specific landmarks and cognitive maps. *Environ. Biol. Fish.* 25, 79–86. doi: 10.1007/bf00002202

Rengstorf, A. M., Yesson, C., Brown, C., and Grehan, A. J. (2013). High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. *J. Biogeogr.* 40, 1702–1714. doi: 10.1111/jbi.12123

Roberts, J. M., Wheeler, A. J., and Freiwald, A. (2006). Reefs of the deep: the biology and geology of cold-water coral ecosystems. *Science* 312, 543–547. doi: 10.1126/science.1119861

Roberts, J. M., Wheeler, A., Freiwald, A., and Cairns, S. (2009). *Cold Water Corals: The Biology and Geology of Deep-Sea Coral Habitats*. Cambridge University Press, Cambridge. doi: 10.1017/CBO9780511581588

Rosenberg, A., Bigford, T. E., Leathery, S., Hill, R. L., and Bickers, K. (2000). Ecosystem approaches to fishery management through essential fish habitat. *Bull. Mar. Sci.* 66, 535–542.

Ross, S. W., and Quattrini, A. M. (2009). Deep-sea reef fish assemblage patterns on the Blake Plateau (Western North Atlantic Ocean). *Mar. Ecol.* 30, 74–92. doi: 10.1111/j.1439-0485.2008.00260.x

Sharma, S., Legendre, P., Boisclair, D., and Gauthier, S. (2012). Effects of spatial scale and choice of statistical model (linear versus tree-based) on determining species-habitat relationships. *Can. J. Fish. Aqua. Sci.* 69, 2095–2111. doi: 10.1139/cjfas-2011-0505

Stone, R. P. (2006). Coral habitat in the Aleutian Islands of Alaska: depth distribution, fine-scale species associations, and fisheries interactions. *Coral Reefs* 25, 229–238. doi: 10.1007/s00338-006-0091-z

Stone, R. P. (2014). *The ecology of deep-sea coral and sponge habitats of the central Aleutian Islands of Alaska*. Washington, DC: U.S. Department of Commerce, 1–52.

Stoner, A. W., Clifford, H. R., Parker, S. J., Auster, P. J., and Wakefield, W. W. (2008). Evaluating the role of fish behavior in surveys conducted with underwater vehicles. *Can. J. Fish. Aqua. Sci.* 65, 1230–1243. doi: 10.1139/f08-032

Sundahl, H., Buhl-Mortensen, P., and Buhl-Mortensen, L. (2020). Distribution and suitable habitat of the cold-water corals Lophelia pertusa. Paragorgia arborea, and Primnoa resedaeformis on the Norwegian continental shelf. *Front. Mar. Sci.* 7:213. doi: 10.3389/fmars.2020.00213

Sward, D., Monk, J., and Barrett, N. (2019). A systematic review of remotely operated vehicle surveys for visually assessing fish assemblages. *Front. Mar. Sci.* 6:134. doi: 10.3389/fmars.2019.00134

Thresher, R. E., Guinotte, J. M., Matear, R. J., and Hobday, A. J. (2015). Options for managing impacts of climate change on a deep-sea community. *Nat. Clim. Change* 5, 635–639. doi: 10.1038/nclimate2611

Tissot, B. N., Yoklavich, M. M., Love, M. S., York, K., and Amend, M. (2006). Benthic invertebrates that form habitat on deep banks off southern California, with special reference to deep-sea coral. *Fish. Bull.* 104, 167–181.

Tittensor, D. P., Baco, A. R., Brewin, P. E., Clark, M. R., Consalvey, M., Hall-Spencer, J., et al. (2009). Predicting global habitat suitability for stony corals on seamounts. *J. Biogeogr.* 36, 1111–1128. doi: 10.1111/j.1365-2699.2008.02062.x

Tolimieri, N., Andrews, K., Williams, G., Katz, S., and Levin, P. S. (2009). Home range size and patterns of space use by lingcod, copper rockfish and quillback rockfish in relation to diel and tidal cycles. *Mar. Ecol. Prog. Ser.* 380, 229–243. doi: 10.3354/meps07930

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *J. Machine Learning Res.* 14, 867–897.

Wolda, H. (1981). Similarity indices, sample size and diversity. *Oecologia* 50, 296–302. doi: 10.1007/bf00344966

Yesson, C., Taylor, M. L., Tittensor, D. P., Davies, A. J., Guinotte, J., Baco, A., et al. (2012). Global habitat suitability of cold-water octocorals. *J. Biogeogr.* 39, 1278–1292. doi: 10.1111/j.1365-2699.2011.02681.x

Yoklavich, M. M., Laidig, T. E., Graiff, K., Clarke, M. E., and Whitmire, C. E. (2018). Incidence of disturbance and damage to deep-sea corals and sponges in areas of high trawl bycatch near the California and Oregon border. *Deep Sea Res. Part II* 150, 156–163. doi: 10.1016/j.dsr2.2017.08.005

Zhang, Y., Xu, Q., Alós, J., Liu, H., Xu, Q., and Yang, H. (2015). Short-term fidelity, habitat use and vertical movement behavior of the black rockfish Sebastes schlegelii as determined by acoustic telemetry. *PLoS One* 10:e0134381. doi: 10.1371/journal pone.0134381

Zuur, A., Ieno, E. N., and Saveliev, A. A. (2017). *Beginner's guide to spatial, temporal, and spatial-temporal ecological data analysis with R-INLA. Volume I: Using GLM and GLMM*. Newburgh: Highland Statistics Ltd.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 23 February 2021
doi: 10.3389/fmars.2021.631657

Northern Shortfin Squid (*Illex illecebrosus*) Fishery Footprint on the Northeast US Continental Shelf

Brooke A. Lowman[1,2], Andrew W. Jones[3], Jeffrey P. Pessutti[3], Anna M. Mercer[3], John P. Manderson[4] and Benjamin Galuardi[2,5]*

[1] ERT, Inc. under Contract to Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Narragansett, RI, United States, [2] School for Marine Science and Technology, University of Massachusetts Dartmouth, New Bedford, MA, United States, [3] Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Narragansett, RI, United States, [4] OpenOcean Research, Philadelphia, PA, United States, [5] Greater Atlantic Regional Fisheries Office, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Gloucester, MA, United States

Northern shortfin squid (*Illex illecebrosus*) have presented a challenge for US fishery management because of their life history traits and broad population distribution. They are characterized by a short semelparous lifespan and high interannual variability in recruitment. Much of the stock resides outside of the boundaries of existing US fisheries surveys and US fishing effort. Based on the annual migration pattern and broad geographic distribution of shortfin squid, it is believed that the US squid fishery in the Mid-Atlantic has not had a substantial impact on the stock; however, recent catches are viewed as tightly constrained by quotas. To better estimate the potential impact of fishing on the resource, we worked with industry representatives, scientists, and managers to estimate the availability of the northern shortfin squid stock on the US continental shelf to the US fishery. Taking a novel analytical approach, we combine a model-based estimate of the area occupied by northern shortfin squid with the empirical US commercial shortfin squid fishery footprint to produce estimates of the area of overlap. Because our method overestimates the fishery footprint and underestimates the full distribution of the stock, we suggest that our estimates of the overlap between the area occupied by the squid and the fishery footprint is a way to develop a conservative estimate of the potential fishery impact on the stock. Our findings suggest a limited degree of overlap between the US fishery and the modeled area occupied by the squid on the US continental shelf, with a range of 1.4–36.3%. The work demonstrates the value of using high-resolution, spatially explicit catch and effort data in a species distribution model to inform management of short-lived and broadly distributed species, such as the northern shortfin squid.

Keywords: *Illex illecebrosus*, fishery footprint, northern shortfin squid, species distribution model, spatiotemporal model

OPEN ACCESS

Edited by:
Mark J. Henderson, U.S. Geological Survey, United States

Reviewed by:
Francois Bastardie, Technical University of Denmark, Denmark
Stephanie Brodie, University of California, Santa Cruz, United States

****Correspondence:***
Brooke A. Lowman
brooke.lowman@noaa.gov

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: *20 November 2020*
Accepted: *21 January 2021*
Published: *23 February 2021*

Citation:
Lowman BA, Jones AW, Pessutti JP, Mercer AM, Manderson JP and Galuardi B (2021) Northern Shortfin Squid (Illex illecebrosus) Fishery Footprint on the Northeast US Continental Shelf. Front. Mar. Sci. 8:631657. doi: 10.3389/fmars.2021.631657

INTRODUCTION

There are many uncertainties inherent in fisheries science and management. For example, natural mortality, catchability, and recruitment dynamics are often unknown. These uncertainties are exacerbated when surveys are not designed for the species of interest, its lifespan is very short, and recruitment is highly variable. Fishery footprints (i.e., the geographical area exposed to

fishing effort) have been used as a means of quantifying the potential impact of fishing on a population (Swartz et al., 2010; Jennings et al., 2012; Amoroso et al., 2018; Kroodsma et al., 2018). Species distribution modeling allows for the identification and estimation of areas critical to species' populations and is often used for management applications such as designing spatial or spatiotemporal fishery closures (Tserpes et al., 2008; Jalali et al., 2015; Rooper et al., 2019). Together, these two quantities can provide insight into the relative severity of fishing pressure. We propose to use the proportion of the occupied area on the US continental shelf as estimated by a species distribution model overlapped by the US fishery footprint to calculate a conservative estimate of stock availability to the fishery as an approximation of the potential impact of the fishery.

Northern shortfin squid (hereafter shortfin squid), *Illex illecebrosus*, live <1 year, die soon after spawning, and have highly variable recruitment that is believed to be environmentally controlled (Dawe and Beck, 1997; Hendrickson, 2004). Since 1996, assessments of this squid stock have recommended in-season assessment and fishery management to ensure sufficient spawner escapement from the US fishery to provide adequate recruitment levels in the subsequent year (Hendrickson et al., 1996). Subsequent stock assessments applied depletion-based models (using a weekly time step) using tow-based shortfin squid fishery catch per unit effort (CPUE) data, reported electronically by shortfin squid harvesters in real time (Hendrickson et al., 2003), to demonstrate the utility of this type of management regime (Northeast Fisheries Science Center [NEFSC], 1999, 2003, 2006). However, the depletion-based methods have not been effective due, in part, to the continuous immigration of cohorts into the relatively small US fishery in some years. Given the limited information available, US fisheries management has set the acceptable biological catch based on biomass and catch history because an overfishing limit cannot be determined by the stock assessment (Federal Register, 2012, 2018). Maximum fishery catches have been limited by quotas during the late 1990s and early 2000s and during recent years. Methods for estimating possible levels of fishing mortality and spawner escapement for the squid would be valuable for informing specifications of acceptable biological catch.

The shortfin squid stock ranges from Florida (approximately 25° N) to southern Labrador (approximately 52° N) in the Northwest Atlantic Ocean and occupies continental shelf to slope sea habitats, which they use as spawning, nursery, and feeding grounds (Dawe and Hendrickson, 1998; Roper et al., 2010; O'Dor and Dawe, 2013). In the spring, some proportion of the shortfin squid stock migrates inshore from the shelf edge to occupy summer and fall feeding and spawning habitats (Hendrickson, 2004) on the US and Canadian continental shelf and in waters managed by the Northwest Atlantic Fisheries Organization (NAFO), while the remaining proportion of adults and juveniles remain in the shelf slope sea (e.g., Rathjen, 1981; Roper et al., 2010; Shea et al., 2017). In the fall, the inshore portion of the stock migrates off-shelf (Hendrickson and Holmes, 2004). Analyses of spatial patterns of sexual maturity using US and Canadian shelf-wide surveys and fisheries-dependent biosampling collections indicate that shortfin squid migrate onto and off of the continental shelf at approximately the same maturity stages and sizes in US and Canadian waters at approximately the same time (Northeast Fisheries Science Center [NEFSC], 1999). The US and Canadian fisheries operate exclusively on the continental shelf (Hendrickson and Showell, 2019; **Figure 1**).

Under the assumption that shortfin squid move onto and off of the shelf over a broad area of the US and Canadian continental shelf as noted above (Northeast Fisheries Science Center [NEFSC], 1999), the vulnerability of shortfin squid to the fishery can be roughly approximated in two dimensions by the ratio of the area fished, A_f, to the area occupied by the stock, A_o. This spatial overlap can be considered an index of availability $\rho = A_f/A_o$ of the stock to the fishery. The complement of ρ (i.e., $1 - \rho$) is the proportion of the area occupied by the stock that is not fished. This statistic can be viewed as an index of proportional area of escapement from the fishery.

Shortfin squid occupy an area much larger than the Northeast US continental shelf, including Labrador, the Flemish cap, Baffin Island, and Southern Greenland, and shelf slope sea (Roper et al., 2010). However, the current analysis focuses on the southern component of the stock that constitutes the US management unit. We adopted a conservative approach to develop estimates of the availability of shortfin squid to the fishery (ρ) and proportional escapement ($1 - \rho$) by confining analysis to fishery-dependent and fishery-independent survey data collected in US continental shelf waters. The shelf slope sea has not routinely been surveyed, and although shelf-wide bottom trawl surveys are conducted in northern waters, including the Scotian Shelf, Bay of Fundy, and Flemish Cap where shortfin squid are abundant, effort data are unavailable for the small Canadian inshore jig fishery (Hendrickson and Showell, 2019). Therefore, we did not include the shelf slope sea and northern shelf waters in our analysis of the area occupied. Our estimates of fishery overlap (ρ) are therefore overestimated, while estimates of proportional escapement ($1 - \rho$) are underestimated.

In this paper, we quantify the area occupied by the southern component of the shortfin squid stock on the US shelf available to the fishery and describe how the area of squid occupancy and overlap with the fishery has changed through time. We illustrate how the proportional availability of the stock varies with differing thresholds of probability of occurrence used to characterize occupancy. We then discuss the value of our approach and findings to precautionary fisheries management.

MATERIALS AND METHODS

Data Sets

We used shortfin squid catch data from bottom trawl surveys conducted in the fall in offshore waters by the Northeast Fishery Science Center (NEFSC) and inshore waters by the Northeast Area Monitoring and Assessment Program (NEAMAP) and state agencies of Maine and New Hampshire (MENH) as well as commercial fishery data from a cooperative study fleet in this analysis (**Figure 2** and **Table 1**).

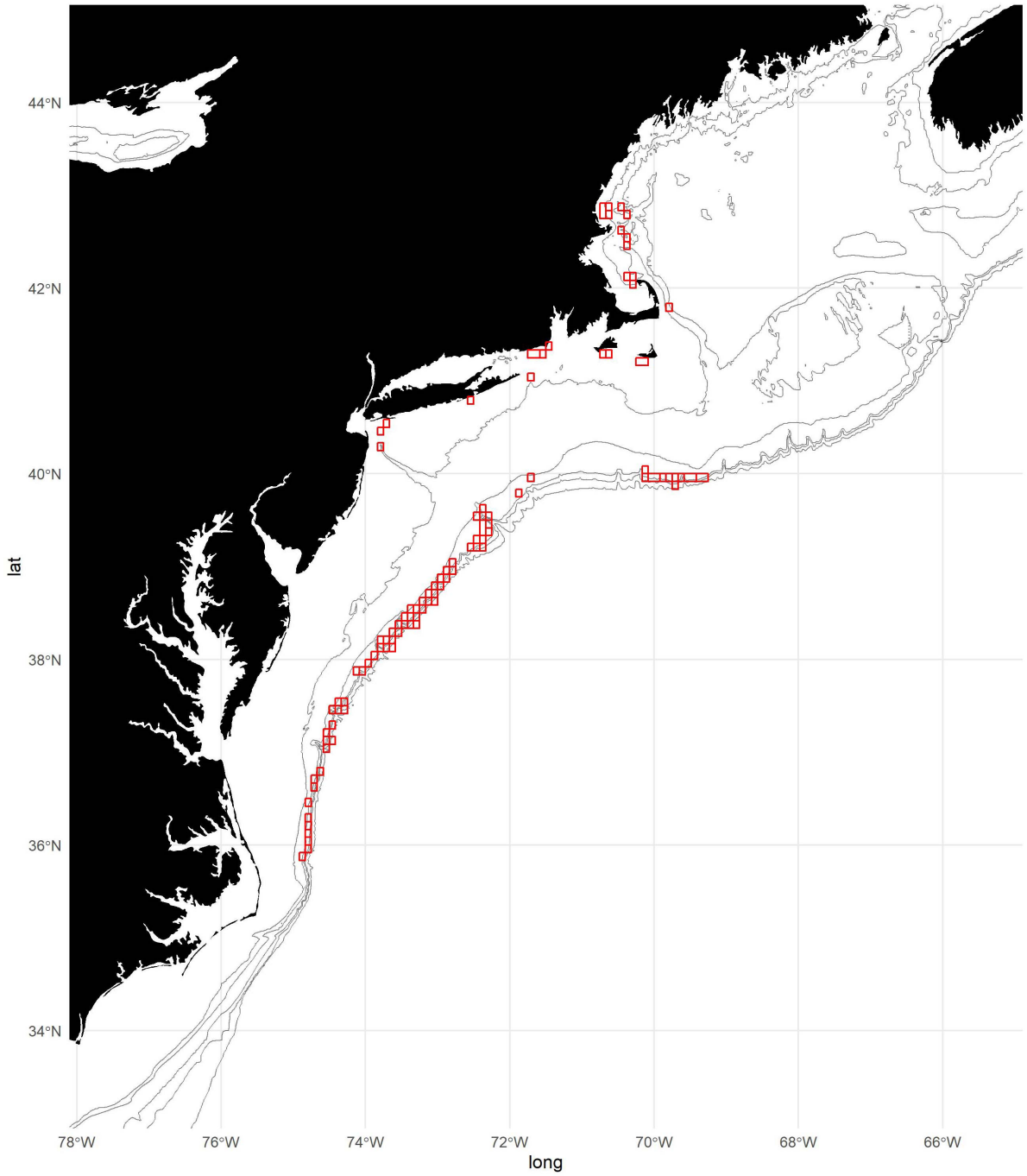

FIGURE 1 | Map of study area with 50, 100, 300, 500, and 1,000 m isobaths. Locations of fishing effort aggregated to 5-min squares are shown in red. Note that some areas of fishing effort have been excluded from the figure to maintain confidentiality.

The NEFSC fall bottom trawl survey is conducted in September–November, and tows are made during both day and night. The survey follows a stratified random design and used a standardized Yankee 36 trawl prior to 2009 and a three-bridle, four-seam trawl thereafter. The vessel used for conducting the survey transitioned from the Albatross to the Bigelow in 2009, following a calibration study in 2008 (Miller et al., 2010). The NEFSC bottom trawl survey gear and protocols are described in Politis et al. (2014). The NEAMAP fall bottom trawl survey is conducted from September to October, and tows are made during the day. The NEAMAP survey follows a stratified random design and uses a trawl with the same design as used in the surveys conducted by the Bigelow but with a 3-in cookie sweep instead of a rockhopper sweep. Full details of the survey protocols

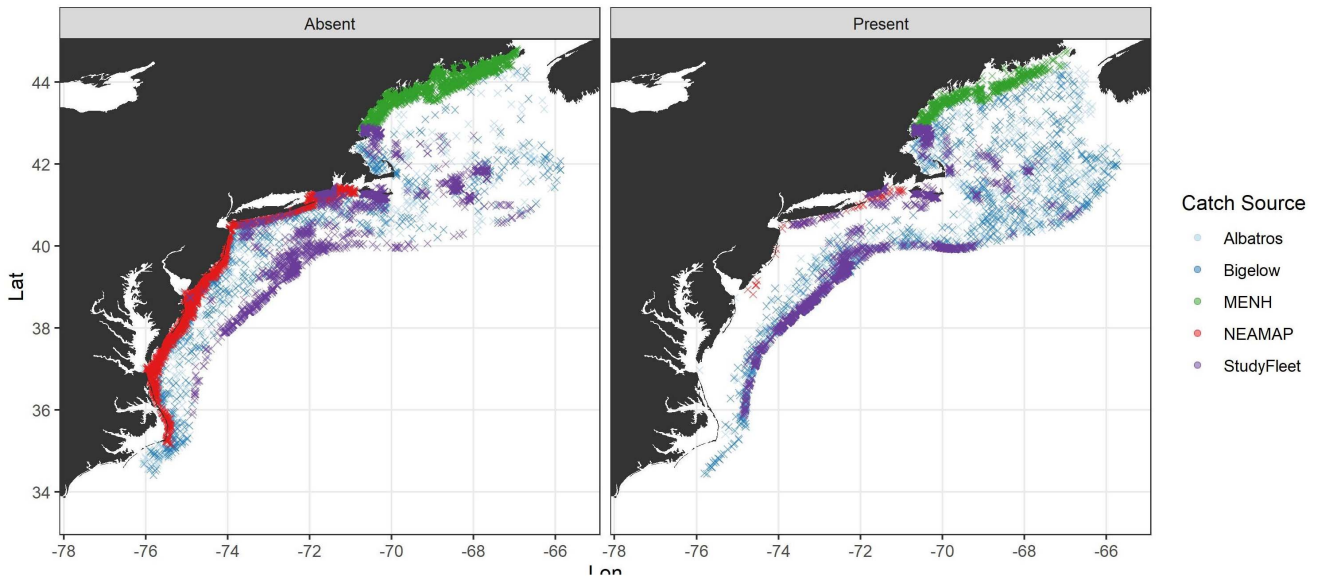

FIGURE 2 | Map of fishery-independent surveys [bottom trawls conducted by the Northeast Fishery Science Center (NEFSC), Northeast Area Monitoring and Assessment Program (NEAMAP), and the Maine Department of Marine Resources and New Hampshire Fish and Game Department (MENH)] and Study Fleet coverage 2000–2018. Data from the two research vessels (Albatros and Bigelow) that have been used in the NEFSC bottom trawl are shown separately.

are described in Bonzek et al. (2017). The MENH fall bottom trawl survey is conducted from October to November, and tows are made during the day. The survey follows a stratified random design and uses a modified shrimp net design. For a complete description of the MENH survey sampling protocols, see Sherman et al. (2005). We analyzed data collected on the Fall MENH survey from 2000 to 2018 (**Table 1**). We used Fall NEFSC bottom trawl survey data from 2000 to 2018. We used NEAMAP survey data from 2007 (the first year of the survey) to 2018.

The surveys used for this analysis have limited spatial overlap, which could be problematic for teasing apart differences in spatial effects versus vessel effects on catch abundance. For this application, however, we considered only the presence and absence of shortfin squid. We used this approach to minimize the impacts of variable shortfin squid detectability in the survey resulting from differences in survey vessel characteristics and net efficiencies.

All bottom trawl survey data were filtered to account for variations in shortfin squid detectability, as described below. Each of the surveys are designed for multispecies sampling and, thus, use different gear than the shortfin squid fishery. Since shortfin squid probably have a low detectability in gears used in the fishery-independent surveys, catch information was reclassified as presence/absence data for this analysis. Furthermore, shortfin squid exhibit diel vertical migration and are typically associated with bottom water during the day. Thus, we only used bottom trawl survey data from "daytime" tows in this analysis. Following the method of Jacobson et al. (2015), we used the astrocalc function in the fishmethods package (Nelson, 2019) to derive the solar zenith angle at the time and geographic position of each tow. The solar zenith angle is 90° when the sun is at the horizon (i.e., local sunrise and sunset), so tows that correspond to solar zenith angles of <90° occurred during the day. The final factor considered when filtering data was seasonality. Fall bottom trawl surveys are conducted near the end of the shortfin squid fishing season, while spring bottom trawl surveys occur during the period of inshore migration of shortfin squid from the slope sea. Thus, the proportion of positive shortfin squid tows and relative abundance indices for the spring survey are much lower than for the fall survey (Northeast Fisheries Science Center [NEFSC], 2006). As a result, only fall bottom trawl survey data are used in assessments to estimate shortfin squid stock size (Northeast Fisheries Science Center [NEFSC], 1999). For consistency with assessments, we also only used data from fall surveys that are more closely aligned with the fishing season.

After filtering the survey data using these criteria, we were left with a data set of 2,836 tows from the NEFSC surveys, 2,227 NEAMAP tows, and 3,284 tows from the MENH survey (**Table 1**). Along with the fishery-independent survey data, we included 5,170 tows from the NEFSC Study Fleet.

The NEFSC Study Fleet began in 2006 and is comprised of approximately 50 commercial fishing vessels (Bell et al., 2017; Blackburn, 2017). The captains join the program voluntarily and are paid for their participation. Captains report tow-specific effort, location, gear characteristics, catch, bycatch, and environmental conditions (e.g., bottom temperature) during their normal fishing operations, and the data are collected through an electronic logbook system. Our current analysis uses data from 2013 to 2018 because there were few shortfin squid trips reported in earlier years.

We include the Study Fleet data in the habitat model along with fishery-independent data to provide temporal and spatial representation of shortfin squid across the shelf during the fishing season. The shortfin squid fishery operates primarily in the

TABLE 1 | Fishery-independent surveys used for building shortfin squid habitat map.

Survey	Years	Months	Maximum depth (m)	Number of hauls	Number of shortfin squid positive hauls
NEFSC fall bottom trawl	2000–2018	Sept–Nov	>183	2,836	1,461
NEAMAP	2007–2018	Sept–Oct	>36.6	2,227	73
MENH	2000–2018	Oct–Nov	>102	3,284	747

Only fall surveys were used, and tows were filtered for daylight hours based on solar zenith angle <90°.

summer months, while the surveys are conducted in the fall. Depending on the productivity of the fishery, there may be little to no temporal overlap with the surveys (e.g., in some highly productive years, the fishery has closed in August). Given the goal of determining fishery footprint overlap with habitat, it is important to incorporate data from the fishing season, and model results based on the surveys alone become difficult to interpret across years when there is temporal overlap between the fishery and surveys in some years and no temporal overlap in others. The authors recognize that using commercial fishing data to determine habitat presents the problem of nonrandom sampling (i.e., the fishery will complete more tows in an area where they expect to catch squid than in other areas). Moreover, using a subset of fishery data (i.e., the Study Fleet) in both the numerator and denominator of our fishery overlap metric is somewhat circular (i.e., some amount of fishery/habitat overlap is guaranteed by the mere fact that a portion of the same data are used to define both the footprint and the habitat). However, the logistical considerations described above justify the use of the Study Fleet data in the habitat model. Finally, the potential direction of bias introduced by the inclusion of Study Fleet data in the habitat model would err on the side of conservatism by tending to increase the degree of fishery/habitat overlap. For these reasons, it is the authors' belief that the benefits gained by including the Study Fleet data in the habitat model outweigh the complications created by using it.

The Fishery and the US Fishery Footprint

The US fishery targets shortfin squid during the warm summer months primarily at depths of 109–365 m on the outer edge of the southern New England and Mid-Atlantic Bight continental shelf. Shortfin squid are a highly perishable seafood product that sells for relatively low prices. To be profitable, the vessels must catch large volumes of squid and process them quickly at shoreside plants and at sea (Manderson, 2019). Since 1999, the US fishery has accounted for approximately 97% of the annual catch of shortfin squid in the Northwest Atlantic (Hendrickson and Showell, 2019). The fishery uses large mesh bottom trawls towed primarily during daylight hours when the squid, which migrate diurnally, are usually concentrated near the seabed. Fishermen report that squid abundance on the shelf break varies at length scales of 10–20 km along the shelf, 0.09–0.5 km cross-shelf, and at time scales of 1–2 days (Manderson, 2019). These space–time scales are similar to those characterizing the dynamics of the shelf slope front (Chen and He, 2010; Todd et al., 2012; Gawarkiewicz et al., 2018). The fishing area is largely determined by technical and regulatory constraints. The fleet is currently prevented from fishing in water deeper than 400–600 m in the mid-Atlantic and New England by coral protection areas, including the Frank R. Lautenberg Deep-Sea Coral Protection Area. Furthermore, the current fleet of vessels is generally not capable of fishing in waters deeper than about 700 m, and capital investments required for deep water trawling in the slope sea are not justified by the current market economics of the fishery.

To develop the fishery footprint, we used Vessel Trip Report (VTR) data provided by the US Greater Atlantic Regional Fisheries Office (GARFO). Records of fishing locations for trips that reported any shortfin squid landings were aggregated to 5-min squares (∼9.25 × 6.90 km = 63.8 km^2 at 42°N) for each year from 2000 to 2019 (**Figure 1**). This approach is at a finer scale than the 10-min square regularly used to characterize the spatial dynamics of the shortfin squid fishery (Hendrickson, 2019). Each 5-min square was attributed as presence/absence of fishing. Vessel Monitoring System (VMS) data were considered for this analysis but ultimately were not used because complete years were only available from 2017 to 2019.

Species Distribution Model

The area shortfin squid occupied within the surveyed portion of the shelf was estimated with a Vector Autoregressive Spatiotemporal (VAST version 3.3.0) model (Thorson J.T., 2019) in R version 3.6.2 (R Core Team, 2019). VAST is a spatiotemporal generalized linear mixed model (GLMM), which by default is a delta style model to model probability of occurrence and a conditional positive catch component. Following the methods of Grüss et al. (2017, 2018), we used only the probability of occurrence model component by turning off all parameters used in the conditional abundance equation (Thorson J., 2019). The probability of occurrence model uses a binomial distribution and logit link. We used 500 spatial knots to fit the model, and we built an extrapolation grid based on the NEFSC survey strata with prediction points placed on a 3 × 3 nautical mile (5.56 × 5.56 km) grid. Area swept is accounted for directly, and we allowed for overdispersion by turning on random effects of vessels on the catchability. The probability of occurrence (p_i) for each sample i was estimated by the binomial GLMM as:

$$p_i = \left[(logit)\right]^{(-1)} (\beta(t_i) + \omega(s_i) + \varepsilon(s_i, t_i) + \eta(v_i))$$

where $\beta(t_i)$ is an intercept for year t_i, $\omega(s_i)$ is a random spatial effect at location s_i, $\varepsilon(s_i,t_i)$ is a random spatiotemporal effect at location s_i in year t_i, and $\eta(v_i)$ is a random effect of vessel v_i.

The NEAMAP, MENH, and NEFSC fall surveys and NEFSC Study Fleet catch data were used in the model to determine the area of the US shelf waters occupied by the shortfin squid southern stock component based on the probability of occurrence. We did not include environmental covariates, such as

bottom temperature in the VAST model because measurements were unavailable for a large number of stations in each of the surveys. We considered filling data gaps with model estimates from the Regional Oceanographic Modeling System (ROMS; Wilkin et al., 2005) but found model-based estimates to be inaccurate compared with *in situ* measurement.

The output prediction points were converted to Voronoi polygons, then joined to form polygons based on the probability of occurrence (i.e., <20%, 20–39.9%, 40–59.9%, 60–79.9%, and >80% probability of occurrence) using the sf package (Pebesma, 2018). These areas of binned probability of occurrence from the VAST model analysis were annual estimates of US continental shelf area occupied by the southern stock component of shortfin squid (A_o), which served as the denominator in computations of the US shortfin squid fishery overlap.

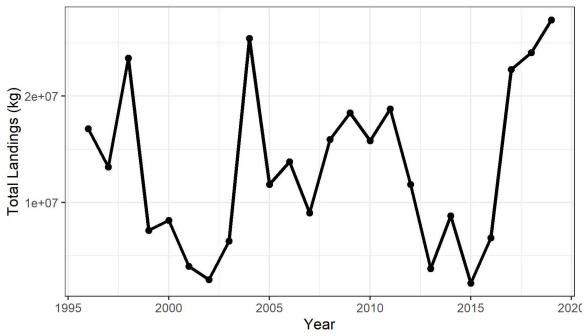

FIGURE 3 | Shortfin squid landings in the US fishery from 1996 to 2018.

Overlap of Fishery Footprint and SDM

Raster files of shortfin squid fishing effort were converted to polygons, then intersected with the habitat areas using the sf (Pebesma, 2018) and raster (Hijmans, 2020) packages. The habitat area overlapping with fishing effort divided by the total habitat area (at each threshold) is the metric of availability of the shortfin squid stock to the fishery such that the spatial estimate of the overlap of the fishery with the stock is given as $\rho = A_f/A_o$, where A_f is the area fished and A_o is the area occupied by the stock.

RESULTS

Species Distribution Model

Model diagnostics showed no evidence that the model did not converge: parameter estimates did not approach upper or lower bounds, the final gradient for all parameters was close to zero (maximum gradient = 9.3×10^{-9}), and the Hessian matrix was positive definite (Appendix A). Observed encounter frequencies were within the 95% confidence interval for nearly all predicted probabilities <0.8, and the observed encounter frequencies at high predicted probabilities tended to be greater than the predicted value and slightly outside the 95% CI (Appendix A). Pearson residual values did not suggest spatial or temporal trends in errors for probability of occurrence (Appendix A).

Differences in spatial patterns of occurrence did not vary systematically between years of high (e.g., 2004, 2017, 2018) and low landings (e.g., 2001, 2002, 2013, 2015) (**Figures 3**, **4**). Model-based estimates of areas occupied by shortfin squid were broadly similar across time in the Mid-Atlantic region but were more variable in the Gulf of Maine and Georges Bank (**Figure 4**). Much of the Gulf of Maine is characterized by relatively low probabilities of occurrence (mostly <40%) in 2000 and 2002, with areas of intermediate probability of occurrence (40–79%) increasing through 2007. Concurrently, the probability of occurrence remained slightly higher on much of Georges Bank (except for the center of the bank, which remains an area of low probability, likely due to lack of sampling). From 2007 to 2019, the probability of occurrence is high in most of the Gulf of Maine and Georges Bank area, except in 2010 and 2015–2016. The highest probabilities of occurrence over the largest area occurred in 2007 and 2018 (**Figure 4**).

The shortfin squid habitat area on the Northeast US continental shelf ranged from 4,262 to 22,656 km^2 using the 80% probability threshold of habitat area and from 51,311 to 151,382 km^2 using the 40% threshold (**Figure 5**). The wide range of habitat area reflects the highly variable nature of shortfin squid catch. The area occupied by squid based on the 40 and 60% thresholds increased from 2000 to 2007. This was followed by a decrease and then a period of lower variability from 2010 to 2016. The area occupied increased at the end of the time series (**Figure 5**). The habitat area based on the 80% threshold is relatively constant throughout the series.

Fishing Footprint

The spatial distribution of shortfin squid fishing effort is consistent at the shelf break in the Mid-Atlantic where the commercial fishery has traditionally been located (**Figure 4**). Fishing effort was more widespread, covering more inshore areas in early years (2000–2004). From 2005 to 2019, fishing effort is mostly confined to a narrow band along the shelf break. In a few years, fishing effort was evident inshore in the Gulf of Maine and on Georges Bank (e.g., 2012 and 2014); however, these areas are not typical for directed shortfin squid trips and appear to be indicative of incidental catch since the squid are targeted in the Mid-Atlantic.

Overlap of Fishery Footprint and SDM

The proportion of shortfin squid habitat on the US continental shelf that is accessed by the fishery (i.e., proportion of fished area overlapping with habitat area) varied each year probability threshold chosen to define habitat area (the largest difference is approximately 30 percentage points, and the average difference is approximately 11 percentage points) (**Figure 5C**). Across years, the minimum estimate for the percent of fishery/habitat overlap was 1.4% (2007 based on 40% threshold), and the maximum estimate for the percent of fishery/habitat overlap was 36.3% (2016 based on 80% threshold). The estimates of proportional area of shortfin squid escapement ranged from a maximum of 98.6% to a minimum of 63.7%.

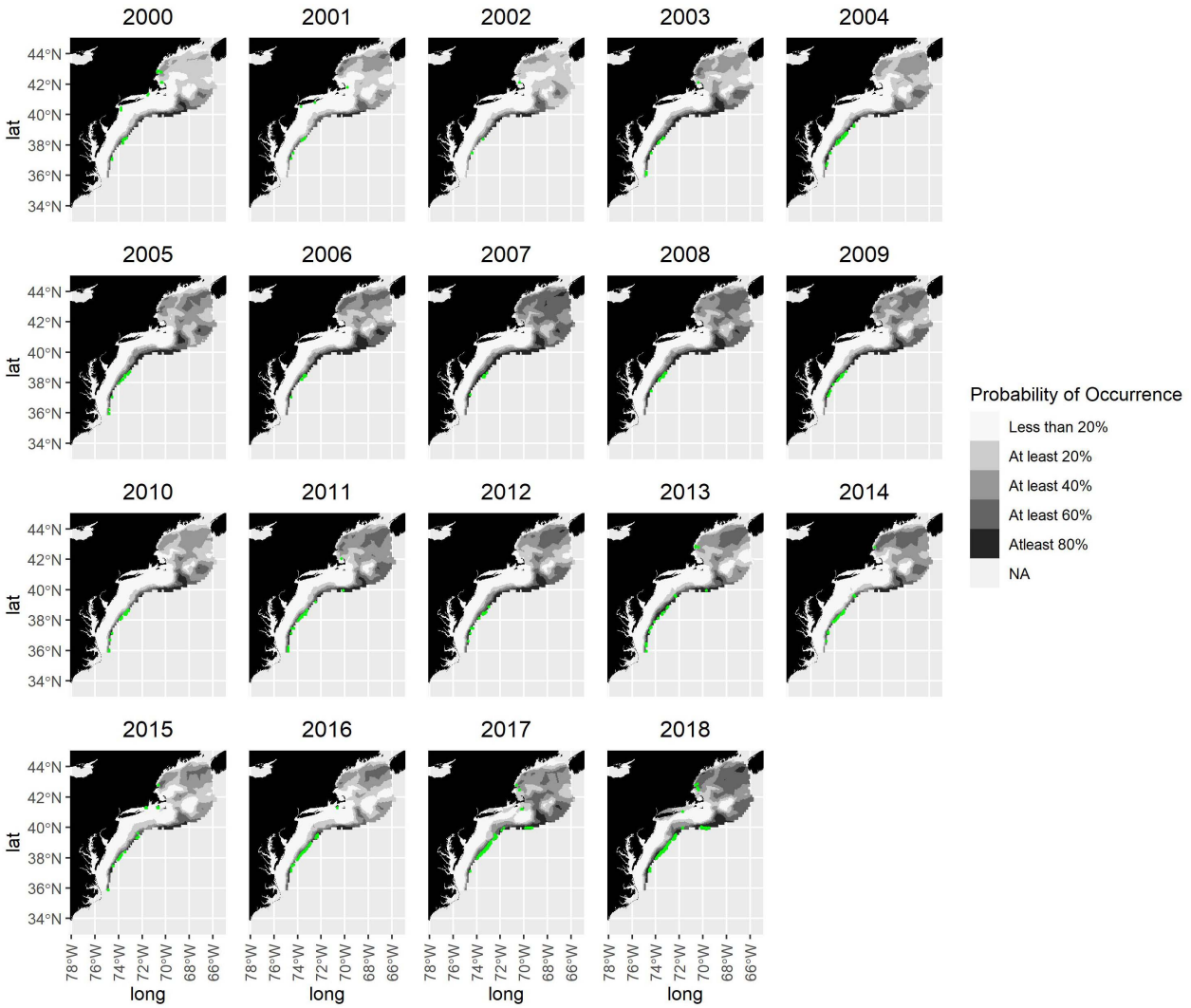

FIGURE 4 | Shortfin squid probability of occurrence map. Quintiles of probability of occurrence are shown in shading. Locations of fishing effort aggregated to 5-min squares are shown in green. Note that some areas of fishing effort have been excluded from the figure to maintain confidentiality.

DISCUSSION

We developed estimates of availability (ρ) and proportional escapement (1 – ρ) for the southern stock component of shortfin squid by confining our analysis to US continental shelf waters where the fishery is well monitored and routine fishery-independent bottom trawl surveys are conducted. Results suggest that even when considering only the US shelf as habitat, a relatively small proportion of the resource (1.4–36.3%) interacts with the fishery, regardless of the threshold chosen to indicate habitat. Given that (1) shortfin squid are known to occupy waters in the shelf slope sea much deeper than those sampled in the available surveys (Rathjen, 1981; Vecchione and Pohle, 2002; Roper et al., 2010; Harrop et al., 2014: Shea et al., 2017) as well as areas beyond the northern and southern extent of surveys considered here (Dawe and Beck, 1985; Hendrickson and Hart, 2006; Roper et al., 2010) and (2) fishing effort is aggregated to a coarse scale representing a much larger area than the actual tow path, the results of this research provide a conservative estimate of habitat with an overestimate of fishing footprint.

The shortfin squid habitat estimated by this research is conservative for several reasons. First, the geographic range of shortfin squid extends far beyond the spatial domain of this research, from South of Cape Hatteras, North Carolina in the Florida Straits (Dawe and Beck, 1985; Hendrickson and Hart, 2006; Roper et al., 2010), Northeast to Labrador, the Flemish cap, Baffin Island, and Southern Greenland. There are confirmed reports of shortfin squid farther east in Iceland, the Azores, and in the Bristol Channel, England (O'Dor and Lipinski, 1998; Roper et al., 1998; O'Dor and Dawe, 2013). In addition, the squid occupy shelf slope sea habitats as adults as well as in the juvenile and larval phases. Bottom and midwater trawl and submersible surveys of the shelf slope sea have documented high concentrations of shortfin squid to bottom depths up to 2,000 m

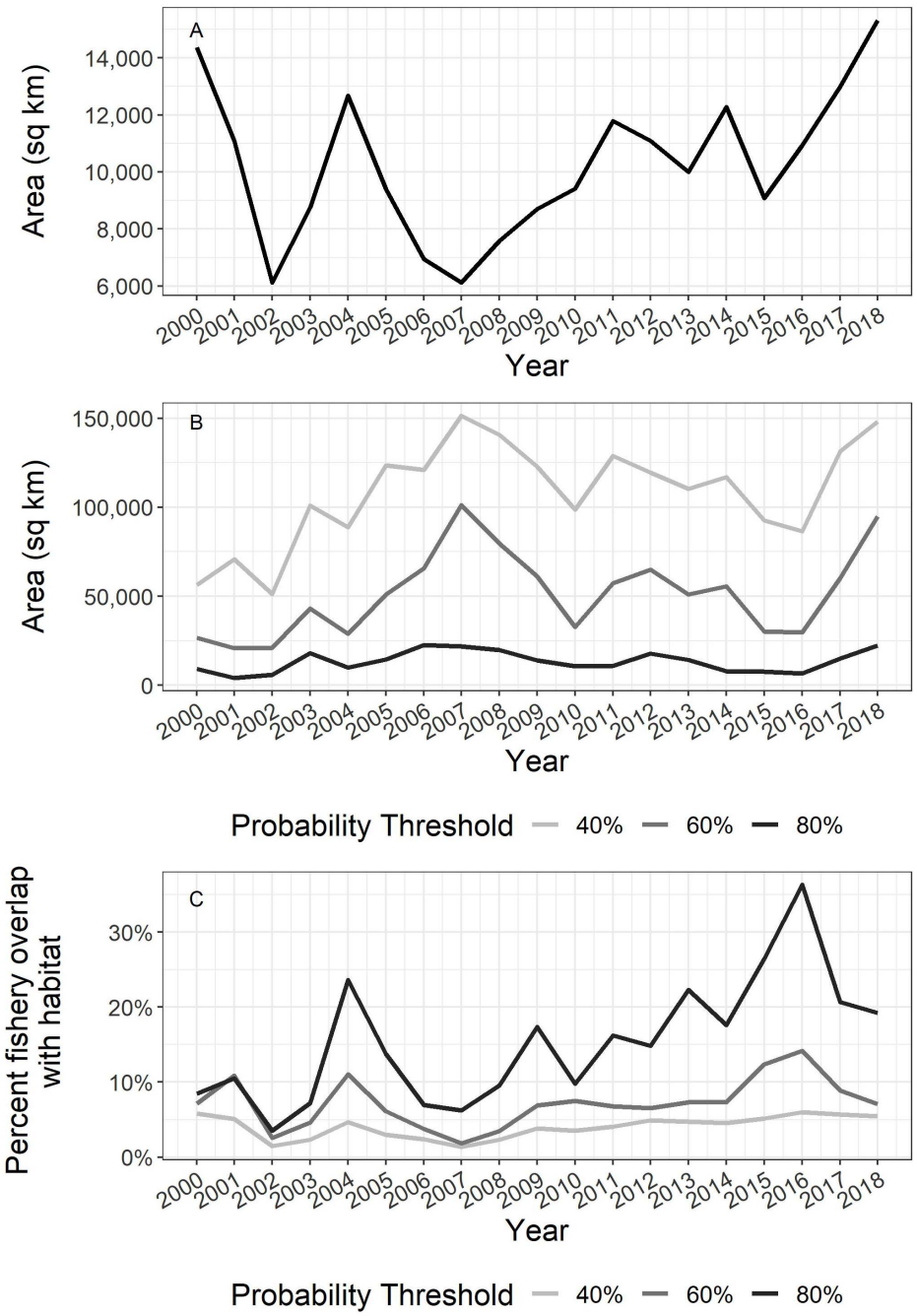

FIGURE 5 | (A) Approximate area fished based on Vessel Trip Reports (VTR) aggregated to 5 minute squares. **(B)** Shortfin squid habitat area on the US continental shelf and **(C)** US fishery overlap based on 40%, 60%, and 80% probability of occurrence threshold for defining habitat.

(Rathjen, 1981; Vecchione and Pohle, 2002; Roper et al., 2010; Harrop et al., 2014; Shea et al., 2017), far beyond the domains of fishery-independent bottom trawl surveys of the US continental shelf (max depth = 542 m; Politis et al., 2014). Stomach content analysis of large pelagic fishes caught in the central Atlantic showed that ommastrephid squids appeared to be the primary food source for these fishes in this region of the ocean (Logan et al., 2013). Thus, the geographic area occupied by shortfin squid is far larger than the area included in this analysis, leading to a conservative estimate of habitat.

Shortfin squid also spend significant amounts of time in pelagic habitats on the continental shelf and the shelf slope sea. Submersible as well as mid-water trawl surveys of the slope sea have observed large concentrations of adult shortfin squid in the water column (Vecchione and Pohle, 2002; Harrop et al., 2014; Shea et al., 2017). There is also evidence that shortfin

squid occupy the pelagic environment on the US and Canadian continental shelf (Froerman, 1981; Brodziak and Hendrickson, 1998; Roper et al., 2010). The pelagic lifestyle of shortfin squid makes its space use volumetric rather than areal. However, this analysis used daytime data when shortfin squid are more closely associated with the seafloor. A volumetric calculation of the availability (ρ) of the squid to these fisheries may be different than the value we calculated using surface areas occupied by shortfin squid and the fishery.

The shortfin squid fishing footprint estimated by this research is likely overestimated, as the scale of fishing effort data is far coarser than actual tow paths. Quantifying fishery footprints has become an increasingly common exercise in recent years on regional (e.g., Jennings et al., 2012; Amoroso et al., 2018) as well as global scales (Kroodsma et al., 2018) as a means to approximate the impacts of fishing. Amoroso et al. (2018) quantified fishery footprints on two dozen shelf/slope areas using VMS and logbook data and found that trawling footprints tended

TABLE 2 | Indices ofshortfin squid abundance (mean kilogram per tow, mean number per tow) from fishery-independent bottom trawls surveys: Fall North East Fisheries Science Center (US NEFSC), Fall southern Gulf of St. Lawrence (Div4t StLau), July Scotian Shelf and Bay of Fundy (DFO SS), Fall Grand Banks (3LNO GB), and July Flemish Cap.

Year	Fall US NEFSC	Fall Div4t StLau	July DFO SS	Fall 3LNO GB	July Flemish Cap
2010	0.05, 8.70	0.18, 0.88	1.08, 19.60	0.00, 0.00	43, NA
2011	0.50, 10.00	0.10, 0.86	1.90, 23.00	0.00, 0.00	89, NA
2012	0.05, 6.30	0.12, 0.88	1.50, 16.90	0.03, 0.22	38, NA
2013	0.40, 8.00	0.01, 0.11	0.10, 1.4	0.00, 0.01	0, NA
2014	0.60, 8.30	0.06, 0.28	1.10, 10.10		3, NA
2015	0.50, 9.50	0.00, 0.00	0.20, 2.40	0.01, 0.09	0.001, NA
2016	0.66, 7.60	0.03, 0.39	0.40, 10.90	0.0185, 0.117	3, NA
2017		0.28, 1.35	16.10, 119.90	0.162, 0.907	2,359, NA
2018	1.30, 15.80	0.89, 5.07		0.2794, 1.648	49, NA
2019			32.10, 196.00		363, NA
Median: 2000–present	0.60, 8.7	0.10, 0.495	1.5, 16.15	0.03, 0.117	79, NA

Data from Hendrickson and Showell (2019).

TABLE 3 | Shortfin squid landings (in metric tons, MT) and percent of total landings in US (NAFO 5&6), and Canadian waters (NAFO 3&4) since 1999 when Canadians ceased licensing foreign fishing on the Nova Scotia Shelf.

	Total landings	US waters Scotian NAFO 5&6		Gulf St Lawrence/ Shelf NAFO 4		Newfoundland–Flemish Cap NAFO 3		Total allowable catch MT	
Year	MT	MT	% Total	MT	%	MT	%	CAN (NAFO 3 + 4)	US (NAFO 5–6)
1999	7,693	7,388	96	286	4	19	0	75,000	19,000
2000	9,377	9,011	96	38	0	328	3	34,000	24,000
2001	4,066	4,009	99	34	1	23	1	34,000	24,000
2002	3,010	2,750	91	30	1	230	8	34,000	24,000
2003	7,524	6,391	85	46	1	1,087	14	34,000	24,000
2004	28,671	26,097	91	34	0	2,540	9	34,000	24,000
2005	12,591	12,013	95	30	0	548	4	34,000	24,000
2006	20,924	13,943	67	24	0	6,957	33	34,000	24,000
2007	9,268	9,022	97	16	0	230	2	34,000	24,000
2008	16,434	15,900	97	11	0	523	3	34,000	24,000
2009	19,136	18,418	96	42	0	676	4	34,000	24,000
2010	15,945	15,825	99	18	0	102	1	34,000	24,000
2011	18,935	18,797	99	50	0	88	0	34,000	23,328
2012	11,756	11,709	100	29	0	18	0	34,000	22,915
2013	3,819	3,792	99	27	1	0	0	34,000	22,915
2014	8,788	8,767	100	21	0	0	0	34,000	22,915
2015	2,437	2,422	99	14	1	0	0	34,000	22,915
2016	6,836	6,682	98	18	0	134	2	34,000	22,915
2017	22,881	22,516	98	52	0	313	1	34,000	22,915
2018	25,663	24,117	94	70	0	1,476	6	34,000	22,915
2019		26,922						34,000	24,825

to be smaller in areas where fishery management reference points were being met. Jennings et al. (2012) estimated fishing area and landings by VMS records assigned to a 3 × 3-km grid. They found that the total fishing footprint (i.e., the area that accounted for 100% of fishing effort) ranged from about 43 to 51% of the total study area and included "core areas" where much of the effort was concentrated as well as large "margins" that contained areas with much less of the effort. Similarly, our finding that the proportion of fishing effort overlap is considerably greater based on the 80% threshold relative to the 60 and 40% thresholds (**Figure 5**) suggests that captains are able to identify and target prime shortfin squid habitat within the study area. Even with this targeting behavior, the impact of the fishing fleet on the population is limited to approximately one-third of the "best" habitat within our study area on the US continental shelf.

This research suggests that shortfin squid have ample opportunity for escapement from the fishery on the northeast continental shelf. Additional opportunities for escapement may be provided in the northern stock area and areas closed to fishing, as explained below. We limited our analysis to US waters, despite the availability of Canadian fishery-independent survey data (**Table 2**) because the Canadian commercial fishery and recreational fishery are not as well monitored as the US fishery (Hendrickson and Showell, 2019). However, examination of available fishery statistics indicates that the capacity of the Canadian commercial fishery is currently quite small when compared with the US fishery (**Table 3**). Since the prohibition of foreign vessels in the Canadian Fishery in 1999, the US summer bottom trawl fishery has accounted for approximately 97% of the total landings of shortfin squid in the Northwest Atlantic (**Table 3**). Fisheries operating in the Gulf of Saint Laurence, Scotian Shelf, and Newfoundland have been responsible for approximately 3% of the landings. The Canadian fishery has achieved only about 1% of the Total Allowable Catch for NAFO areas 3 and 4 (range, 0–21%) since 2000. Thus, the northern shortfin squid stock area, which is not included in our analysis, represents an additional portion of the species range that provides for escapement of potential spawners because the Canadian fishery has remained small in capacity (Hendrickson and Showell, 2019; **Table 3**).

US fishery regulations that prevent fishing in areas on the outer continental shelf and slope sea also provide shortfin squid with permanent regions of escapement from the fishery. These areas include the *Frank R. Lautenberg Deep-Sea Coral Protection Area*, the tilefish and lobster gear-restricted areas, and other regulated mesh areas in Gulf of Maine and Georges Bank that prohibit the use of fine mesh trawls used by the squid fishery. The Coral Protection Area occurs along the shelf break at depths >450 m, covers 38,000 square nautical miles including a large area of the slope sea, and 15 canyon areas where the fishery cannot operate. Large concentrations of shortfin squid have also been observed in the slope sea near seamounts within the 4,913 square mile Monument Area that is now closed to mobile fishing gear (Shea et al., 2017). Given these areas of squid occupancy outside of our current study area, our results may overestimate the availability of shortfin squid to the fishery despite our fishery footprint area being based on presence/absence and unscaled to catch or effort.

A clear next step for this research is to incorporate environmental factors in the shortfin squid habitat model. A study of 67 cephalopod time series indicated population increases from the 1950s through the 2010s across various taxa and life histories, suggesting that common large-scale processes drive the increase, aided by biological aspects of cephalopods (Doubleday et al., 2016). Similarly, an examination of the relationship between oceanographic characteristics and spatial distribution of cephalopods in the Yellow Sea suggested that shifts in spatial distribution of cephalopods over the study period was consistent with environmental drivers rather than fishing pressure (Jin et al., 2020).

In conclusion, our findings are consistent with advice to regional management that the shortfin squid stock is unlikely to be negatively impacted by a modest increase in catch (Didden, 2018) because the US fishery overlaps a small portion of the area occupied by the southern stock component. The overlap of fishing area with areas where shortfin squid are likely to occur does not account for the variations in density. Fishing is concentrated on the shelf break because this is where squid are concentrated prior to their subsequent use of shoreward habitats. In some years, these areas may have had sufficient densities or detectability by the fishing fleets to support commercial harvest. In view of the limited understanding of recruitment dynamics of shortfin squid, the potential impacts of harvests on spawning stock escapement are not known. By the same measure, there is no direct evidence of recruitment overfishing for shortfin squid. However, several lines of evidence suggest low potential effects of fishing activity. The near absence of fishing activity in the known historical range of shortfin squid in the US and Canada and the occurrence of shortfin squid at depths and distances well offshore suggest a large region of unfished resource. A high fishing mortality on the entire resource would be possible only if a large fraction of the resource passed through the actual fishing areas of the US. Thus, it is unlikely that the US fishery has had a substantial impact on the southern stock component of shortfin squid.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Maine-New Hampshire survey data are available through the Maine Marine Resources data portal (https://mainedmr.shinyapps.io/MaineDMR_Trawl_Survey_Portal/). The NEAMAP survey data are available by request through the NEAMAP data access page (http://www.neamap.net/dataAccess.html). The NEFSC survey data are available by request, and more information can be found in the NOAA data catalog (https://catalog.data.gov/dataset/fall-bottom-trawl-survey). Study Fleet and Vessel Trip Report data are confidential but can be made available in aggregated form upon request.

AUTHOR CONTRIBUTIONS

BL, AJ, JP, AM, and JM conceived of the study. AJ, BG, and JM curated the data. BL performed the data analysis. BL and JM wrote the manuscript with contributions from all authors.

FUNDING

This research was provided by the National Oceanic and Atmospheric Administration, Northeast Fisheries Science Center.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance from Sara Murray, James Gartland, Rebecca Peters, Kim McKown, Matt Camisa, and Linda Barry for providing data from the state inshore surveys.

We are grateful to Tyler Pavlowich and Rich Bell for providing assistance with coding. Meghan Lapp offered useful suggestions on an earlier version of the text. We thank Paul Rago, Jason Didden, and Charles Adams for providing helpful feedback on early versions of the manuscript. We thank the captains and crew of vessels who participate in the Study Fleet. We also thank the shortfin squid harvesters and processors who provided useful insight. Members of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee and the Illex Working Group provided useful comments and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.631657/full#supplementary-material

REFERENCES

Amoroso, R. O., Pitcher, C. R., Rijnsdorp, A. D., McConnaughey, R. A., Parma, A. M., Suuronen, P., et al. (2018). Bottom trawl fishing footprints on the world's continental shelves. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10275–E10282. doi: 10.1073/pnas.1802379115

Bell, R. J., Gervelis, B., Chamberlain, G., and Hoey, J. (2017). Discard estimates from self-reported catch data: an example from the US Northeast Shelf. *N. Am. J. Fish.* 37, 1130–1144. doi: 10.1080/02755947.2017.1350219

Blackburn, J. (2017). *Study Fleet Inventory of Data Uses: A Tool for Stakeholder Engagement*. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-06. (Woods Hole, MA: Northeast Fisheries Science Center), 18. doi: 10.7289/V5/RD-NEFSC-17-06

Bonzek, C. F., Gartland, J., Gauthier, D. J., and Latour, R. J. (2017). *Northeast Area Monitoring and Assessment Program (NEAMAP) Data Collection and Analysis in Support of Single and Multispecies Stock Assessments in the Mid—Atlantic: Northeast Area Monitoring and Assessment Program Near Shore Trawl Survey*. (Gloucester Point, VA: Virginia Institute of Marine Science, College of William and Mary). doi: 10.25773/ 7206-KM61

Brodziak, J., and Hendrickson, L. (1998). An analysis of environmental effects on survey catches of squids *Loligo pealei* and *Illex illecebrosus* in the northwest Atlantic. *Fish. Bull.* 97, 9–24.

Chen, K., and He, R. (2010). Numerical investigation of the middle Atlantic bight shelfbreak frontal circulation using a high-resolution ocean hindcast model. *J. Phys. Oceanogr.* 40, 949–964. doi: 10.1175/2009jpo4262.1

Dawe, E. G., and Beck, P. C. (1985). Distribution and size of short-finned squid (*Illex illecebrosus*) larvae in the Northwest Atlantic from winter surveys in 1969, 1981 and 1982. *J. Northwest Atl. Fish. Sci.* 6, 43–55. doi: 10.2960/j.v6.a5

Dawe, E. G., and Beck, P. C. (1997). Population structure, growth, and sexual maturation of short-finned squid (*Illex illecebrosus*) at Newfoundland. *Can. J. Fish. Aquat. Sci.* 54 137–146. doi: 10.1139/f96-263

Dawe, E. G., and Hendrickson, L. C. (1998). A review of the biology, population dynamics, and exploitation of short-finned squid in the northwest Atlantic Ocean, in relation to assessment and management of the resource. *NAFO SCR* Doc. 98/59, Ser. No. N3051, 33.

Didden, J. (2018). *Ilex ABC [Memorandum]. Mid-Atlantic Fishery Management Council*. Available online at: https://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/5b8990764ae237ecb08cb785/1535742070744/IllexABC.pdf (accessed August 31, 2018).

Doubleday, Z. A., Prowse, T. A. A., Arkhipkin, A., Pierce, G. J., Semmens, J., Steer, M., et al. (2016). Global proliferation of cephalopods. *Curr. Biol.* 26, R406–R407. doi: 10.1016/j.cub.2016.04.002

Federal Register (2012). *Fisheries of the Northeastern United States; Atlantic Mackerel, Squid, and Butterfish Fisheries;* Framework Adjustment 6, 77 Federal Register 51857 (August 27, 2012) (codified at 50 C.F.R. 648). Federal Register. Washington, D.C.

Federal Register (2018). *Fisheries of the Northeastern United States; Mid-Atlantic Fishery Management Council; Omnibus Acceptable Biological Catch* Framework Adjustment, 83 Federal Register 15512 (April 11, 2018) (codified at 50 C.F.R. 648). Federal Register. Washington, D.C.

Froerman, Y. M. (1981). *Approach to Stock Assessment of Nerito-Oceanic Squids of the Family Ommastrephidae in the Atlantic Ocean on the Example of North-Western Atlantic Shortfin Squid Illex illecebrosus (Lesueur, 1821). Status of Resources and Basis of Rational Fishery in the Atlantic Ocean*. (Kaliningrad: AtlantNIRO publishing), 60–69.

Gawarkiewicz, G., Todd, R. E., Zhang, W., Partida, J., Gangopadhyay, A., Monim, M.-U.-H., et al. (2018). The changing nature of shelf-break exchange revealed by the OOI Pioneer Array. *Oceanography* 31, 60–70. doi: 10.5670/oceanog.2018.110

Grüss, A., Thorson, J. T., Sagarese, S. R., Babcock, E. A., Karnauskas, M., Walter, J. F. I. I. I., et al. (2017). Ontogenetic spatial distributions of red grouper (*Epinephelus morio*) and gag grouper (*Mycteroperca microlepis*) in the US Gulf of Mexico. *Fish. Res.* 193, 129–142. doi: 10.1016/j.fishres.2017.04.006

Grüss, A., Thorson, J. T., Babcock, E. A., and Tarnecki, J. H. (2018). Producing distribution maps for informing ecosystem-based fisheries management using a comprehensive survey database and spatio-temporal models. *ICES J. Mar. Sci.* 75, 158–177. doi: 10.1093/icesjms/fsx120

Harrop, J., Vecchione, M., and Felley, J. D. (2014). In situ observations on behaviour of the ommastrephid squid genus *shortfin squid* (Cephalopoda: Ommastrephidae) in the northwestern Atlantic. *J. Nat. Hist.* 48, 2501–2516. doi: 10.1080/00222933.2014.937367

Hendrickson, L. C. (2004). Population biology of northern shortfin squid (Illex illecebrosus) in the Northwest Atlantic Ocean and initial documentation of a spawning area. *ICES J. Mar. Sci.* 61, 252–266. doi: 10.1016/j.icesjms.2003.10.010

Hendrickson, L. C. (2019). *2019 Fishery and Survey Data Update Report for the Southern Stock Component of Northern Shortfin Squid (Illex illecebrosus). Report to Mid-Atlantic Fishery Management Council Scientific and Statistical Meeting*. Available online at: http://www.mafmc.org/ssc-meetings/2019/may-7-8 (accessed March 13, 2020).

Hendrickson, L. C., and Hart, D. R. (2006). An age-based cohort model for estimating the spawning mortality of semelparous cephalopods with an application to per-recruit calculations for the northern shortfin squid, *Illex illecebrosus*. *Fish. Res.* 78, 4–13. doi: 10.1016/j.fishres.2005.12.005

Hendrickson, L. C., and Holmes, E. M. (2004). *Essential Fish Habitat Source Document: Northern Shortfin Squid, Illex illecebrosus, Life History and Habitat Characteristics*: NOAA Tech Memo NMFS-NE-191, 2nd Edn. (Woods Hole, MA: Northeast Fisheries Science Center), 46.

Hendrickson, L. C., and Showell, M. A. (2019). *2019 Assessment of Northern Shortfin Squid (Illex illecebrosus) in Subareas 3+4*. (Dartmouth, NS: Northwest Atlantic Fisheries Organization). Serial No. N6973

Hendrickson, L. C., Brodziak, I., Basson, M., and Rago, P. (1996). *Stock Assessment of Northern Shortfin Squid in the Northwest Atlantic During 1993*. Northeast

Fish. Sci. Cent. Ref. Doc. 96-05g. (Woods Hole, MA: Northeast Fisheries Science Center), 63.

Hendrickson, L. C., Hiltz, D. A., McBride, H. M., North, B. M., and Palmer, J. E. (2003). *Implementation of Electronic Logbook Reporting in a Squid Bottom Trawl Study Fleet During 2002.* Northeast Fish. Sci. Cent. Ref. Doc. 03-07. (Woods Hole, MA: National Marine Fisheries Service), 30.

Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling. R Package Version 3.0-12*. Available online at: https://CRAN.R-project.org/package=raster (accessed February 27, 2020).

Jacobson, L. D., Hendrickson, L. C., and Tang, J. (2015). Solar zenith angles for biological research and an expected catch model for diel vertical migration patterns that affect stock size estimates for longfin inshore squid (*Doryteuthis pealeii*). *Can. J. Fish. Aquat. Sci.* 72, 1329–1338. doi: 10.1139/cjfas-2014-0436

Jalali, M. A., Ierodiaconou, D., Monk, J., Gorfine, H., and Rattray, A. (2015). Predictive mapping of abalone fishing grounds using remotely-sensed LiDAR and commercial catch data. *Fish. Res.* 169, 26–36. doi: 10.1016/j.fishres.2015.04.009

Jennings, S., Lee, J., and Hiddink, J. G. (2012). Assessing fisher footprints and the trade-offs between lanings value, habitat sensitivity, and fishing impacts to inform marine spatial plannin and an ecosystem approach. *ICES J. Mar. Sci.* 69, 1053–1063. doi: 10.1093/icesjms/fss050

Jin, Y., Jin, X., Gorfine, H., Wu, Q., and Shan, X. (2020). Modeling the oceanographic impacts on the spatial distribution of common cephalopods during autumn in the Yellow Sea. *Front. Mar. Sci.* 7:432. doi: 10.3389/fmars.2020.00432

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., et al. (2018). Tracking the global footprint of fisheries. *Science* 359, 904–908. doi: 10.1126/science.aao5646

Logan, J. M., Toppin, R., Smit, S., Galuardi, B., Porter, J., and Lutcavage, M. (2013). Contribution of cephalopod prey to the diet of large pelagic fish predators in the central North Atlantic Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 95, 74–82. doi: 10.1016/j.dsr2.2012.06.003

Manderson, J. P. (2019). *Summary Report of the Northern Shortfin Squid (Illex Illecebrosus) Population Ecology & the Fishery Summit, November 25-26, 2019 Wakefield, Rhode Island*. Available online at: https://www.dropbox.com/s/ph2t7l1hjpzzeae/DraftSummary_NorthernShortfinSquidPopulationEcologyAndTheFisherySummit_03242020.pdf?dl=0 (accessed March 24, 2020).

Miller, T. J., Das, C., Politis, P. J., Miller, A. S., Lucey, S. M., Legault, C. M., et al. (2010). *Estimation of Albatross IV to Henry B. Bigelow Calibration Factors.* Northeast Fish. Sci. Cent. Ref. Doc. 10-05. (Woods Hole, MA: National Marine Fisheries Service), 233.

Nelson, G. A. (2019). *fishmethods: Fishery Science Methods and Models. R Package Version 1.11-1*. Available online at: https://CRAN.R-project.org/package=fishmethods (accessed June 1, 2020).

Northeast Fisheries Science Center [NEFSC] (1999). *Report of the 29th Northeast Regional Stock Assessment Workshop (29th SAW): Stock Assessment Review Committee SARC) Consensus Summary of Assessments:* Northeast Fish. Sci. Cent. Ref. Doc. 99–14. (Woods Hole, MA: Northeast Fisheries Science Center [NEFSC]).

Northeast Fisheries Science Center [NEFSC] (2003). *Report of the 37th Northeast Regional Stock Assessment Workshop (37th SAW): Stock Assessment Review Committee (SARC) Consensus Summary of Assessments:* Northeast Fish. Sci. Cent. Ref. Doc. 03–16. (Woods Hole, MA: Northeast Fisheries Science Center [NEFSC]).

Northeast Fisheries Science Center [NEFSC] (2006). *42nd Northeast Regional Stock Assessment Workshop (42nd SAW) Stock Assessment Report Part A: Silver Hake, Mackerel, & Northern Shortfin Squid:* Northeast Fish. Sci. Cent. Ref. Doc. 06–09a. (Woods Hole, MA: Northeast Fisheries Science Center).

O'Dor, R. K., and Dawe, E. G. (2013). "*Illex illecebrosus*, northern short-finned squid," in *Advances in Squid Biology, Ecology and Fisheries*, eds R. Rosa, G. Pierce, and R. O'Dor (Hauppauge, NY: Nova Science Publishers).

O'Dor, R. K., and Lipinski, M. R. (1998). "The genus *Shortfin Squid* (Cephalopoda, Ommastrephidae): characteristics, distribution and fisheries," in *Squid Recruitment Dynamics. The Genus Illex as a Model, the Commercial Illex Species and Influences on Variability*, eds P. G. Rodhouse, E. G. Dawe, and R. K. O'Dor (Rome: FAO fisheries technical paper). doi: 10.1016/s0165-7836(02)00006-1

Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *R J.* 10, 439–446. doi: 10.32614/RJ-2018-009

Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. (2014). *Northeast Fisheries Science Center Bottom Trawl Survey Protocols for the NOAA Ship Henry B. Bigelow US Dept Commer 14-06*. (Woods Hole, MA: Northeast Fisheries Science Center), 138.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rathjen, W. F. (1981). Exploratory squid catches along the continental slope of the Eastern United States. *J. Shellfish Res.* 1, 153–159.

Rooper, C. N., Hoff, G. R., Stevenson, D. E., Orr, J. W., and Spies, I. B. (2019). Skate egg nursery habitat in the eastern Bering Sea: a predictive model Mar. *Ecol. Prog. Ser.* 609, 163–178. doi: 10.3354/meps12809

Roper, C. F. E., Lu, C. C., and Vecchione, M. (1998). "A revision of the systematics and distribution of *Shortfin squid* species (Cephalopoda: Ommastrephidae)," in *Systematics and Biogeography of Cephalopods*, eds N. A. Voss, M. Vecchione, R. B. Toll, and M. J. Sweeney (Tokyo: Tokai University Press), 405–423.

Roper, C. F. E., Nigmatullin, C., and Jereb, P. (2010). "Family ommastrephidae," in *Cephalopods of the World. An Annotated and Illustrated Catalogue of Species Known to Date. Volume 2. Myopsid and Oegopsid Squids. FAO Species Catalogue for Fishery Purposes*, Vol. 2, eds P. Jereb, and C. F. E. Roper (Rome: FAO), 269–347.

Shea, E. K., Judkins, H., Staudinger, M. D., Dimkovikj, V. H., Lindgren, A., and Vecchione, M. (2017). Cephalopod biodiversity in the vicinity of Bear Seamount, western North Atlantic based on exploratory trawling from 2000 to 2014. *Mar. Biodivers.* 47, 699–722. doi: 10.1007/s12526-017-0633-3

Sherman, S. A., Stepanek, K., and Sowles, J. (2005). *Maine-New Hampshire Inshore Groundfish Trawl Survey Procedures and Protocols.* ME DMF Research Reference Document 05/01, 42. Augusta, ME. Maine Department of Marine Resources.

Swartz, W., Sala, E., Tracey, S., Watson, R., and Pauly, D. (2010). The spatial expansion and ecological footprint of fisheries (1950 to present). *PLoS One* 5:e15143. doi: 10.1371/journal.pone.0015143

Thorson, J. (2019). *VAST Model Structure and User Interface [PDF File]*. Available online at: https://github.com/James-Thorson-NOAA/VAST/blob/master/manual/VAST_model_structure.pdf (accessed December 10, 2019).

Thorson, J. T. (2019). Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Todd, R. E., Gawarkiewicz, G. G., and Owens, W. B. (2012). Horizontal scales of variability over the middle Atlantic bight shelf break and continental rise from finescale observations. *J. Phys. Oceanogr.* 43, 222–230. doi: 10.1175/jpo-d-12-099.1

Tserpes, G., Politou, C.-Y., Peristeraki, P., Kallianiotis, A., and Papconstantinou, C. (2008). Identification of hake distribution pattern and nursery grounds in the Hellenic seas by means of generalized additive models. *Hydrobiologia* 612, 125–133. doi: 10.1007/s10750-008-9486-x

Vecchione, M., and Pohle, G. (2002). Midwater cephalopods in the western North Atlantic Ocean off Nova Scotia. *Bull. Mar. Sci.* 71, 883–892.

Wilkin, J. L., Arango, H. G., Haidvogel, D. B., Lichtenwalner, C. S., Glenn, S. M., and Hedström, K. S. (2005). A regional ocean modeling system for the long-term ecosystem observatory. *J. Geophys. Res.* 110:C06S91. doi: 10.1029/2003JC002218

Conflict of Interest: BL was employed by ERT, Inc. JM was employed by OpenOcean Research.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 01 April 2021
doi: 10.3389/fmars.2021.632389

Assessing Habitat Suitability Models for the Deep Sea: Is Our Ability to Predict the Distributions of Seafloor Fauna Improving?

David A. Bowden[1], Owen F. Anderson[1], Ashley A. Rowden[1,2], Fabrice Stephenson[3] and Malcolm R. Clark[1]*

[1] National Institute of Water and Atmospheric Research, Wellington, New Zealand, [2] School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand, [3] National Institute of Water and Atmospheric Research, Hamilton, New Zealand

OPEN ACCESS

Edited by:
Mark J. Henderson, United States Geological Survey, United States

Reviewed by:
Chris Rooper, National Marine Fisheries Service (NOAA), United States
Chris Yesson, Zoological Society of London, United Kingdom
Derek Tittensor, Dalhousie University, Canada

***Correspondence:**
David A. Bowden
David.Bowden@niwa.co.nz

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: *23 November 2020*
Accepted: *23 February 2021*
Published: *01 April 2021*

Citation:
Bowden DA, Anderson OF, Rowden AA, Stephenson F and Clark MR (2021) Assessing Habitat Suitability Models for the Deep Sea: Is Our Ability to Predict the Distributions of Seafloor Fauna Improving? Front. Mar. Sci. 8:632389. doi: 10.3389/fmars.2021.632389

Methods that predict the distributions of species and habitats by developing statistical relationships between observed occurrences and environmental gradients have become common tools in environmental research, resource management, and conservation. The uptake of model predictions in practical applications remains limited, however, because validation against independent sample data is rarely practical, especially at larger spatial scales and in poorly sampled environments. Here, we use a quantitative dataset of benthic invertebrate faunal distributions from seabed photographic surveys of an important fisheries area in New Zealand as independent data against which to assess the usefulness of 47 habitat suitability models from eight published studies in the region. When assessed against the independent data, model performance was lower than in published cross-validation values, a trend of increasing performance over time seen in published metrics was not supported, and while 74% of the models were potentially useful for predicting presence or absence, correlations with prevalence and density were weak. We investigate the reasons underlying these results, using recently proposed standards to identify areas in which improvements can best be made. We conclude that commonly used cross-validation methods can yield inflated values of prediction success even when spatial structure in the input data is allowed for, and that the main impediments to prediction success are likely to include unquantified uncertainty in available predictor variables, lack of some ecologically important variables, lack of confirmed absence data for most taxa, and modeling at coarse taxonomic resolution.

Keywords: habitat suitability, species-environment models, distributions, deep sea, benthos, epifauna, predictive models, AUC

INTRODUCTION

Understanding and managing ecosystem effects of human activities, such as bottom-contact fishing and mineral extraction in the deep sea (depths greater than ca. 200 m), requires quantitative information on the distributions of benthic habitats and fauna (Kaiser et al., 2016; Pitcher et al., 2017). Because such information is generally sparse in waters beyond coastal areas, management

decision-making relies increasingly on outputs from habitat suitability models (also known as species distribution models), which develop correlations between point-sampled faunal occurrence records and spatially continuous environmental variables to predict probabilities of suitable habitat or taxon occurrence across unsampled environmental space (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009; Guisan et al., 2013; Vierod et al., 2014; Reiss et al., 2015). Methods commonly in use include Boosted Regression Trees (BRT, Friedman et al., 2000; Leathwick et al., 2006; De'ath, 2007), Generalized Additive Models (GAM), Maximum Entropy (MaxEnt, Phillips et al., 2006), and Random Forests (RF, Breiman, 2001). These and other types of habitat suitability models are in constant development (e.g., Warton et al., 2015) and are used increasingly in a broad range of applications (Guisan et al., 2013; Robinson et al., 2017; Araujo et al., 2019). The fundamental requirements of all methods, however, are the same: accurate and sufficient point-sample data about where a taxon has been recorded and, ideally, where it is has been confirmed to be absent, and accurate and ecologically relevant environmental data as continuous layers spanning the region of interest.

The relative paucity, patchiness, and taxonomic selectivity of available faunal sample data in the deep sea, a lack of spatial resolution and local validation of environmental layers, and limited understanding of biotic interactions and historical factors that might influence present distributions, in combination, can result in high levels of uncertainty being associated with the outputs from habitat suitability models (Fielding and Bell, 1997; Araujo and Guisan, 2006; Vierod et al., 2014; Reiss et al., 2015; Anderson et al., 2016a). This uncertainty is exacerbated by the cross-validation methods commonly used to evaluate model performance, in which subsets of the input taxon occurrence data are withheld from the model and used as test sites to assess predictions. While this approach is practical, it can generate overly optimistic values of model performance (Bahn and Mcgill, 2013; Ploton et al., 2020) that may not be supported by field validation (Anderson et al., 2016a) because the data used in cross-validation methods are not independent from those used to build the model itself and are likely to be spatially biased (e.g., Lobo et al., 2008; Ploton et al., 2020).

Evaluation against data collected independently of those used in the modeling process is the most convincing approach to model assessment because it is directly relevant to the way in which model predictions are used in practice: if we are to have confidence in the model outputs, we need to know how reliable they are in relation to independent observations of the target taxa (Verbyla and Litvaitis, 1989; Fielding and Bell, 1997; Pearce and Ferrier, 2000; Araujo et al., 2019). An important point here is that such independent observations should be made using methods that detect the taxa of interest reliably. In most studies of benthic invertebrate taxa in the deep sea, taxon occurrence data are compiled from sampling methods, typically demersal fish trawls, that are not efficient at catching benthos, leading to unquantified uncertainty in relation to detection and selectivity. Evaluation against independent data is rare, however, for the same reasons that habitat suitability modeling itself is a useful tool. That is, sample data about species' occurrences are usually sparse because such data are time-consuming, logistically challenging, and expensive to collect and habitat suitability modeling approaches have been developed as a more pragmatic, rapid, and affordable way to map distributions. However, because independent validation of models is rarely undertaken, confidence in their predictions can be low, limiting their credibility for use in environmental management (Anderson et al., 2016a; Winship et al., 2020). Model uncertainty can be reduced by development of more sophisticated modeling methods or by increasing the quality and quantity of data inputs but without evaluation of performance against independent data, we cannot be sure that such developments translate into practical gains. Therefore, in places where successive models have been developed, with progressive updates to input data and modeling techniques, it is important to understand whether more recent models represent improvements in terms of increased prediction accuracy and thereby build confidence in their use for fisheries and other management purposes.

In New Zealand and the wider southwest Pacific region, growing concern about the ecosystem effects of fisheries and potential seabed mineral extraction operations has stimulated interest in improving knowledge about the distributions of seafloor fauna (Rowden et al., 2019). Habitat suitability modeling has been used in several studies of seafloor faunal distributions, mostly for sessile invertebrate taxa such as corals and sponges that are recognized as being particularly sensitive or vulnerable to anthropogenic disturbances (e.g., Tracey et al., 2011; Anderson et al., 2014) but also for demersal fishes (Leathwick et al., 2006) and mobile benthic fauna (Compton et al., 2013; Bowden et al., 2019a). The potential of habitat suitability modeling methods to predict distributions across unsampled space is of particular appeal in the region because, despite being rich in biological and mineral resources, relatively little of its seafloor has been surveyed in detail, other than in areas of particular interest for fisheries research. Many of the broad-scale habitat suitability modeling initiatives in this region arose as direct or indirect consequences of concerns about the seabed impacts of commercial bottom-contact fisheries. Bottom trawl fisheries target hoki (*Macruronus novaezelandiae*) and other demersal species on smooth substrata over large areas of New Zealand's Exclusive Economic Zone (EEZ) in depths of 300–1,400 m and deep-sea species including orange roughy (*Hoplostethus atlanticus*), oreo (Oreosomatidae), and alfonsino (primarily *Beryx splendens*) on seamounts and other underwater topographic features throughout the region (Fisheries New Zealand, 2020). Much of what is known about the distributions of non-target seafloor taxa comes from bycatch records from these fisheries and the research trawl surveys that inform catch advice for them (O'Driscoll et al., 2011), and most habitat suitability models in the region have been based on occurrence data from these records in combination with records from museum and other specimen collection databases.

The only evaluation of deep-sea habitat suitability model predictions using data collected independently and by methods designed to detect the target taxa to date in the region is the study of Anderson et al. (2016a), in which the authors of the present paper first developed models for cold-water coral taxa, then designed and ran a seabed photographic

survey specifically to test their predictions on the Louisville Seamount Chain to the east of New Zealand. We found that the models performed poorly in practice and attributed this to a number of potential factors, including a lack of reliably supported taxon absence records, low precision of available environmental variables, particularly bathymetry, and lack of ecologically relevant variables, such as substrate type, which are key determinants of benthic taxon distributions. In light of these results and a general lack of confidence in modeled distributions for use in environmental management, Fisheries New Zealand and the National Institute of Water and Atmospheric Research initiated a project to independently assess the predictive performance of existing models and improve confidence in future predictions. Focusing on a major fisheries area in New Zealand, Chatham Rise, the first stage of this project was to generate a fully independent, quantitative, dataset of seabed faunal distributions derived from photographic surveys that would enable objective assessment of existing habitat suitability models for the region (Bowden et al., 2019b).

Here, we use this independent dataset to assess the usefulness of outputs from published habitat suitability models for the region. We first use best-practice model building standards proposed by Araujo et al. (2019) to rank the existing models in order of their expected predictive performance, then by comparing model predictions against the independent data, we generate five metrics describing performance: area under the receiver operating characteristic curve (AUC); true skill statistic (TSS); results from *t*-tests comparing the mean of published model probability values for all locations at which a taxon was present in the independent test data against the mean value for locations at which it was absent; and correlation strength between predicted probability of suitable habitat for a given taxon and both its prevalence in the test data (R^2_{prev}) and its standardized population density in the test data (R^2_{dens}). We use these metrics to assess how well each model performs in absolute terms, and to rank them in order of realized performance, which we hypothesize should match the expected ranking based on the Araujo et al. (2019) criteria. We then refer to the Araujo et al. (2019) criteria to discuss which aspects of model development have the greatest influence on realized model performance.

MATERIALS AND METHODS

Study Area

The study focuses on Chatham Rise because this part of New Zealand's EEZ has the highest density of seafloor photographic survey data available for use in model evaluation, is physically central to many of the habitat suitability models available for evaluation in the region, and is the source of much of the specimen data that informed development of these models (**Figure 1**). Chatham Rise is a continental rise that extends eastward from the South Island of New Zealand for approximately 1,000 km, with the Chatham Islands toward the eastern end. The Sub-Tropical Front coincides with and is partially constrained by the rise, making it the most biologically productive fisheries region in the EEZ (McClatchie et al., 1997; Clark et al., 2000; Marchal et al., 2009; Nodder et al., 2012). Recent summaries of bottom-contact trawl history across Chatham Rise (Black et al., 2013; Black and Tilney, 2015; Baird and Mules, 2019) show high trawling intensity, primarily from the hoki fishery, at a 450–700-m depth west of Mernoo Bank and on the southern and northern central flanks of Chatham Rise, with locally very high intensities of trawling for orange roughy, oreo, and alfonsino on seamount and knoll features on the northern, eastern, and southern flanks. At present, initiatives to protect benthic habitats and fauna are limited to closures to fishing on some seamounts in the "Graveyard" and "Andes" regions (since 2001) on the northwest and southeast flanks of the rise, respectively (Brodie and Clark, 2003; Clark and Dunn, 2012), and establishment in 2007 of two benthic protection areas (BPAs): the Mid Chatham Rise BPA and the East Chatham Rise BPA (Helson et al., 2010).

Existing Models

Predictive models of habitat suitability for benthic epifaunal invertebrate taxa that encompass Chatham Rise have been published, primarily, in eight separate studies by our research team since 2011 (Tracey et al., 2011; Baird et al., 2013; Compton et al., 2013; Anderson et al., 2014, 2015, 2016a,b; Georgian et al., 2019). Most of these studies have focused on protected corals (Tracey et al., 2011; Baird et al., 2013; Anderson et al., 2014, 2015) and vulnerable marine ecosystem (VME, sensu FAO, 2009) indicator taxa (Anderson et al., 2016a,b; Georgian et al., 2019), with only one study producing models for individual species across a wide range of taxonomic groups (Compton et al., 2013). We selected models for taxa that were well represented in our independent dataset (see below), with presence records at nine or more sites within the spatial domain of the model. Across the eight studies, 47 individual models spanning 31 taxa (**Table 1**) were suitable for assessment against the independent dataset.

All of the studies were undertaken with the principal aim of predicting occurrence across unsampled geographic space within the geographic range of their input faunal data (Prediction, sensu Araujo and Guisan, 2006; Araujo et al., 2019). Three modeling techniques were used across the studies: BRT, RF, and MaxEnt, with BRT being the most commonly used. The treatment of input data varied across studies, particularly in the approach to defining absence records. Five studies worked with presence–*background* data, using either randomly selected or spatially structured background points from the study area as assumed "pseudo-absences" (Tracey et al., 2011; Baird et al., 2013; Anderson et al., 2014, 2016a; Georgian et al., 2019), while the others used presence–*absence* data, deriving absence sites either from a combination of research trawl bycatch and museum databases or, in the case of Compton et al. (2013), from the source photographic survey data. Presence–*absence* models give an indication of the probability of a taxon being present, whereas models using pseudo-absences provide only a measure of the probability of suitable habitat being present.

The spatial extents of the studies range from the entire South Pacific Regional Fisheries Management Organisation (SPRFMO) Convention area (Anderson et al., 2016a) down to a section of the central part of the New Zealand EEZ (Compton et al., 2013), but all are centered on Chatham Rise (**Figure 1**).

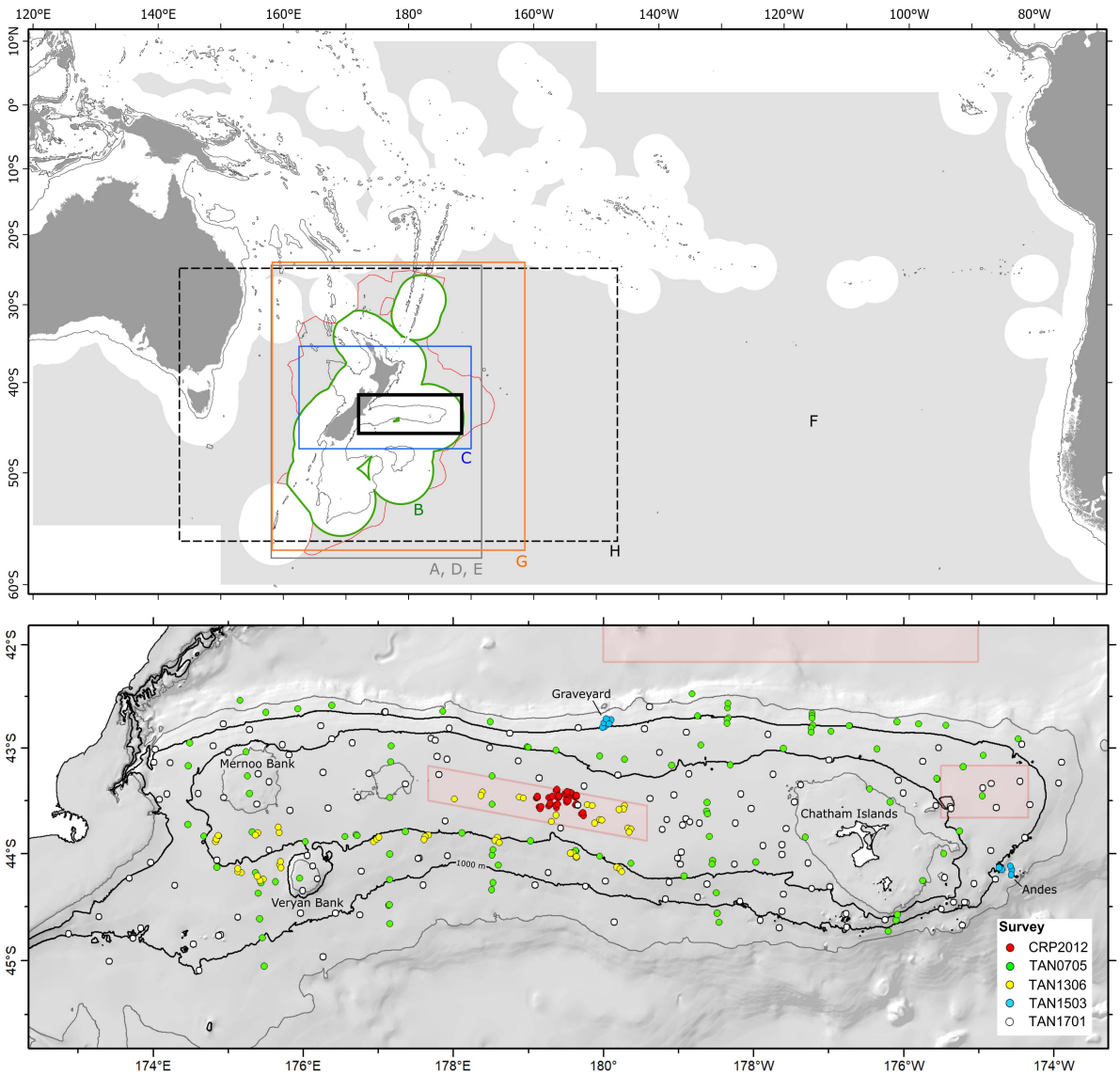

FIGURE 1 | (Top) areas of eight published studies from which predictive models were assessed: A, Tracey et al. (2011); B, Baird et al. (2013); C, Compton et al. (2013); D, Anderson et al. (2014); E, Anderson et al. (2015); F, Anderson et al. (2016a), G, Anderson et al. (2016b); and H, Georgian et al. (2019). Boundary B is the New Zealand Exclusive Economic Zone (EEZ). Also showing New Zealand's Extended Continental Shelf boundary (red polygon), the 1,000-m isobath, and Chatham Rise (bold black rectangle). **(Bottom)** Chatham Rise showing the location of DTIS photographic transect stations (color-coded points, see legend) for the five surveys from which the independent test dataset was developed. The Graveyard and Andes seamount complexes are indicated, isobaths show 250-, 500-, 1, 000-, and 1,500-m depths, and red polygons show Benthic Protection Areas (BPAs), which are protected from seabed trawl fishing.

Spatial resolution for all studies was constrained to 1 km^2 by the resolution of available environmental predictor data. This resolution approximates to that of most of the methods used to collect the underlying sample data—primarily towed sampling gear including trawls, dredges, and epibenthic sleds—and matches closely the length of photographic transects from which the independent data were compiled.

All but one of the published studies were trained on benthic invertebrate sample data from physical specimens from research trawl surveys, fisheries bycatch, and museum collections, with most occurrence records coming from within the New Zealand EEZ, and many of these from Chatham Rise itself. Compton et al. (2013), by contrast, used observation data from photographic transects and epibenthic sled samples collected during two dedicated surveys of benthic biodiversity, one of which was TAN0705 (see below for relevance). All studies used k-fold cross-validation to evaluate model performance, a technique in which portions of the available sample data are iteratively withheld from the model training phase and then used to generate performance metrics based on how well the model predicts their values. The detail of how this validation was performed varied across studies, from random withholding of sites across

TABLE 1 | Summary details of the eight existing SDM studies and the individual models suitable for assessment against the independent dataset.

Name	Spatial extent	Model type	Assessed taxa	Model resolution
Tracey et al., 2011	New Zealand ECS	BRT	*Goniocorella dumosa*	1 km
Baird et al., 2013	New Zealand EEZ	BRT	Coral-reef, solitary small, Gorgonacea, Stylasteridae, Antipatharia, Scleractinia	1 km
Compton et al., 2013	Central New Zealand region	BRT	*Anthomastus robustus*, Corallimorpharia, *Flabellum* spp., Galatheidae, *Hyalinoecia tubicola*, *Metanephrops challengeri*, *Radicipes* spp., Scaphopoda, Serolidae, Spatangidae, *Taiaroa tauhou*, Volutidae, Zoantharia, Paguridae, *Gracilechinus multidentatus*, Echinothuroida	1 km
Anderson et al., 2014	New Zealand EEZ	BRT	Coral-reef, *Goniocorella dumosa*, Antipatharia	1 km
Anderson et al., 2015	New Zealand region	BRT	Coral-reef, *Goniocorella dumosa*, Antipatharia	1 km
Anderson et al., 2016a	SPRFMO convention area and EEZ	BRT and MaxEnt	Coral-reef (REEF; BRT and MaxEnt), *Goniocorella dumosa* (MaxEnt)	30 arc-seconds (ca. 1 km N-S)
Anderson et al., 2016b	New Zealand region	Ensemble (BRT + MaxEnt)	*Goniocorella dumosa*, Brisingidae, Antipatharia, Stylasteridae, Crinoidea, Demospongiae, Hexactinellida, Pennatulacea	1 km
Georgian et al., 2019	South West Pacific Ocean	Ensemble: (BRT + RF + MaxEnt)	*Goniocorella dumosa*, Antipatharia, Stylasteridae, Demospongiae, Hexactinellida, Pennatulacea, Alcyonacea	1 km

*ECS, **e**xtended **c**ontinental **s**helf; EEZ, **e**xclusive **e**conomic **z**one; Coral-reef, scleractinian branching corals Goniocorella dumosa, Enallopsammia rostrata, Solenosmilia variabilis, and Madrepora oculata combined; solitary small = hydrocorals and cup corals.*

the entire model domain (e.g., Tracey et al., 2011) to selection within longitudinal bands to compensate for spatial structuring of sample data (Anderson et al., 2016b). Spatial autocorrelation in the input sample data was addressed explicitly only in the most recent study (Georgian et al., 2019), by inclusion of a residual auto-covariate (RAC) predictor variable in their models (Crase et al., 2012). All of the studies used the area under the receiver-operating characteristic curve (AUC, Hanley and Mcneil, 1982; Swets, 1988; Bradley, 1997) as their metric of model predictive success.

Expected Performance

Before evaluating the published models against the independent dataset, we ranked the eight studies in order of their expected performance by applying the standards for best practice in habitat suitability modeling proposed by Araujo et al. (2019). The aim of this ranking was to place our results in the context of an objective framework that might subsequently help identify which aspects of the models contributed most to their predictive performance when assessed against independent data. The standards span the four broad components of model design: response variables, predictor variables, model building, and model evaluation, nested within which there are 15 "issues" (**Table 2**), each with guidelines allowing a given model to be ranked as either "Gold" (best practice), "Silver," "Bronze," or "Deficient." Each author in the present paper scored each study for each of the 15 issues independently. Scores were then discussed, adjusted by consensus, and the studies ranked in order of overall score, with the expectation that models from higher-ranking studies should perform better against the independent data than those from lower-ranking ones.

Independent Dataset

Source Data

A dataset of benthic mega-epifauna density records from Chatham Rise was assembled from quantitative analyses of seabed video and still-image transects from five research surveys conducted between 2007 and 2017 (**Figure 1**). Voyages TAN0705 (Bowden, 2011; Compton et al., 2013), TAN1701 (Bowden et al., 2017), and TAN1306 (Bowden and Leduc, 2017) were broad-scale surveys of benthic biodiversity following stratified random designs, while voyage TAN1503 was focused on seamounts, with multiple summit-to-base camera transects on features in the Graveyard and Andes seamount complexes (Clark et al., 2019). Voyage CRP2012 (Rowden et al., 2014) focused on areas of phosphorite-rich sediments on the central crest area of Chatham Rise, using a design with replicate transects within multiple survey blocks. Data derived from these surveys are independent from those used to train the published models in that they were collected without reference to the original source data or the surveys from which they were compiled. They are, however, from a region of the published model domains that has the highest density of sample data and, thus, are spatially interspersed with the original training data.

Quantitative data on the occurrence of benthic invertebrate fauna were extracted from imagery from each survey under separate research projects over a period of 10 years (see survey references above), but the data extraction methods used were similar throughout, being run by the same team of researchers (DB, AR, and MC). These methods, and the auditing procedures that were used to create a combined dataset of faunal occurrences, are described in detail by Bowden et al. (2019b). In brief, seafloor photographic transects of approximately 1 km distance were

TABLE 2 | Predictive model assessment criteria (1–4) and issues (A–E) proposed by Araujo et al. (2019).

	1 – Response variables	2 – Predictor variables	3 – Model building	4 – Model evaluation
A	Sampling	Selection	Complexity	Evaluation of model assumptions
B	Identification of taxa	Spatial resolution	Treatment of bias and noise in response variables	Evaluation of model outputs
C	Spatial accuracy	Uncertainty	Treatment of collinearity	Measures of model performance
D	Environmental extent		Dealing with modeling and parameter uncertainty	
E	Geographic extent			

run at each survey site, recording either high-definition digital color video (HD1080 format), digital still images (at 8-, 10-, 12-, or 24-megapixel resolution, depending on survey), or, for most surveys, both formats simultaneously. Four of the surveys used the same towed camera system (NIWA's Deep Towed Imaging System, DTIS, Hill, 2009; Bowden and Jones, 2016), which records continuous HD video with intermittent high-resolution still images captured simultaneously at 15-s intervals and was deployed using the same operating procedures and methods for logging navigational and observational data on all surveys. The CRP2012 survey was conducted by remotely operated vehicle (ROV) on the central Chatham Rise crest. It was designed by the same research group (AR and DB) specifically to generate data compatible with standard DTIS surveys but the ROV used lower-resolution video and still-image cameras.

For surveys TAN0705, TAN1306, and TAN1701, the full length of each video transect was reviewed by analysts using Ocean Floor Observation Protocol (OFOP, Huetten and Greinert, 2008) software to record the occurrence and taxonomic identities of all fauna visible (larger than ca. 5 cm) on the seabed and referring to the high-resolution still images to confirm taxonomic identifications in consultation with taxonomic experts. In this method, each occurrence is referenced by spatial coordinates and time, enabling direct retrospective audit of individual records by examination of the original imagery. For surveys CRP2012 and TAN1503, still images were analyzed, rather than video; for the former because video quality was too low, and for the latter, to be comparable with data from earlier surveys (Clark et al., 2019). Merging data from the five surveys involved three stages: (1) checking and aligning taxon identities to ensure consistency of identifications and nomenclature; (2) comparing taxon presence and counts in areas of survey overlap to check for systematic survey or analyst bias; and (3) aggregating taxa into higher groupings where necessary to match those used in the nine published models under evaluation. For example, several of the modeling studies produced models for all reef-forming stony coral species combined; therefore, observations of *Goniocorella dumosa*, *Enallopsammia rostrata*, *Solenosmilia variabilis*, and *Madrepora oculata* in the independent data were combined under a single taxon label "coral-reef" or "REEF" for comparison with these models. Similarly, records of comatulid and stalked crinoids were combined to match model predictions which did not differentiate between these forms.

The independent test dataset spanned the full extent of Chatham Rise from 172° 50′ E to 173° 53′ W and 42° 29′ S to 45° 5′ S and from 40- to 1,850-m depth. It consisted of 125,658 observations of individual benthic organisms from analyses of 358 seabed photographic transects, with 109,161 records from analyses of video, and 15,795 from still images. In the full dataset, there were 354 taxa across 13 phyla, with taxonomic level ranging from phylum to species, and the initial taxon aggregation process yielded a set of 79 "aggregated" taxa, ranging in taxonomic level from species level for distinctive taxa (e.g., the decapod crustacean *Metanephrops challengeri*), to family (e.g., Primnoidae and Brisingidae), order (e.g., Ceriantharia), class (e.g., Asteroidea and Holothuroidea), and phylum (e.g., Brachiopoda and Bryozoa). Full details of the data are given in Bowden et al. (2019b).

Density and Prevalence Measures

Standardized population density estimates (as individuals 1,000 m^{-2} of seafloor) for each taxon recorded in the photographic surveys were derived from the observation data, using seafloor swept areas calculated as the product of transect length and average image frame width for video (see Bowden et al., 2019b for details), and summed areas of all individual images for still photographs. While density estimates are ideally suited for assessment of predictions from abundance-based models, they are not strictly comparable with the probability values generated by models based on presence–absence data. Because none of the existing models available for evaluation were based on abundance data, we also calculated *prevalence* (i.e., occurrence rate, see Anderson et al., 2016a) for each taxon at each site, which more closely approximates to measures of probability of occurrence or suitable habitat. Prevalence was calculated in two ways, depending on the type of imagery. For the video-based analyses (surveys TAN0705, TAN1306, and TAN1701), each transect was divided into 1-, 5-, and 10-min time segments (three alternative values chosen to allow for a segment-length effect). Time, rather than distance, was used here for simplicity of calculation, but as tow speeds during individual transects are relatively constant, differences in resulting distance at the seabed are minor. The number of segments in which the taxon was recorded at least once was then divided by the total number of segments in the transect to calculate its prevalence at the site (**Supplementary Figure 1**). For the still-image-based analyses (CRP2012 and TAN1503), prevalence in each transect was estimated simply by calculating the proportion of the total number of images analyzed in which the taxon of interest was identified.

Habitat suitability values associated with the midpoint location of each photographic transect were extracted from the model grids of each of the published models for each taxon, using functions in the *raster* and *rgdal* packages in R (R Core Team 2017). Because transects were approximately 1 km long and the

environmental predictors were gridded at 1 km, it is likely that a proportion of the transects cross boundaries between grid cells. However, because the spatial domains of the models were large in relation to the grid size, and because the 1-km grid of the predictor variables is a convenient minimum scaling that does not necessarily reflect the native resolution of the data that inform them, fine-scale adjustments to allow for boundary crossing are unlikely to affect our results or to yield reliable insights at the scale of the study.

Model Assessment

The level of agreement between model predictions and the independent data was assessed using five metrics, three based on ability to predict presence–absence correctly and two on ability to predict abundance correctly:

(1) AUC_{ind}—area under the receiver operating characteristic curve, using predicted probabilities of occurrence from the existing models against presence in the independent dataset. This is a presence–absence comparison, with AUC yielding a single metric of discrimination across all possible thresholds for predicted presence (Fielding and Bell, 1997; Lobo et al., 2008). AUC is a standard measure of predictive model performance and in this context can be defined as the area under a plot of the proportion of true positives versus the proportion of false positives; a value of 0.5 indicates a model with no discriminatory power, and a value of 1 indicates a model that correctly identifies all records. There are no formally agreed thresholds for interpreting AUC values but there is some consensus that models with values greater than 0.7 can be considered useful and those with values greater than 0.85 reliable (Swets, 1988; Fielding and Bell, 1997; Wiley et al., 2003; Glover and Vaughn, 2010).

(2) TSS—true skill statistic (Allouche et al., 2006). This is a presence–absence comparison, calculated as sensitivity (i.e., the probability of predicting presences correctly) plus specificity (the probability of predicting absences correctly) minus one. It is proposed as a prevalence-independent measure of model success. TSS takes into account both omission and commission errors, and scales from −1 to 1. A value of 1 indicates perfect prediction success, while values of 0 or less indicate a performance no better than random or a systematically incorrect prediction. Models with TSS values 0.6 or more are considered to be useful (Allouche et al., 2006).

(3) *t*-test—results from one-tailed independent sample *t*-tests comparing the mean of published model probability values for all locations at which a taxon was present in the independent test data ($\bar{x}P$) against the mean value for all locations at which it was absent ($\bar{x}A$). Prior to testing, distributions of the model probabilities for each taxon were examined and log transformations applied in some cases to reduce skewness in the data and better approximate the normal distribution. This test is also a presence–absence comparison, based on the simple expectation that modeled probabilities should, on average, be greater at sampled sites where a taxon is present than at sites where it is absent (i.e., $\bar{x}P > \bar{x}A$) in the independent dataset. The resulting *p*-values are presented as three categories: not supported ($p \geq 0.05$, "NS"); true ($0.05 > p > 0.01$, "T"); or significant ($p < 0.01$, "TS").

(4) R^2_{prev}—correlation strength from a linear model fitting the published model probabilities to prevalence values from the independent dataset. Separate fits were assessed for taxon prevalence calculated from the 1-, 5-, and 10-min time segments, and results presented as the mean and standard deviation of these. This is a test of prediction success against a measure that is intermediate between presence–absence and density.

(5) R^2_{dens}—correlation strength from a linear model fitting the published model probabilities to measured taxon density values from the independent dataset. This is a test of prediction success against the full quantitative detail of the independent data.

The challenge associated with correct prediction increases with the level of information demanded of the prediction, with prediction of presence or absence being a simpler task than prediction of relative or absolute abundance (Bahn and Mcgill, 2013). Therefore, we expected better performance against the three presence–absence metrics (AUC, TSS, and *t*-tests) than against the prevalence and abundance metrics (R^2_{prev} and R^2_{dens}) but still with the expectation that more recent models and models ranked higher in our initial qualitative assessment would perform better than earlier and lower ranked models.

AUC is the most widely used metric of prediction performance and was used in all the existing published studies as the primary metric. Therefore, calculation of AUC using the independent data here enabled direct comparison against the published AUC values calculated by k-fold cross-validation (AUC_{kcv}). Both AUC and TSS are considered to be largely independent of differences in prevalence (the proportion of sites at which a target taxon is present) and might be expected to yield comparable results because the underlying logic of their calculation is similar (Allouche et al., 2006; Somodi et al., 2017). The t-test comparison also reduces the required predictions to presence–absence and thus was expected to yield results comparable to those from AUC and TSS. The two quantitative metrics here, R^2_{prev} and R^2_{dens}, are intended to evaluate predictive performance in terms of how these habitat suitability model outputs are likely to be used in practice: to answer questions about both where taxa are likely to be encountered and at what relative densities (Bahn and Mcgill, 2007). Strictly interpreted, presence-only models predict only the probability of suitable habitat for a taxon being present and thus should not be expected to predict the occurrence of a taxon or its population density. We include a density comparison here, however, because the outputs from presence-only models are often intuitively interpreted as predictions of distribution, particularly in environmental management scenarios, and the availability of fully quantitative independent evaluation data here affords a rare opportunity to demonstrate in practice what the consequences of inferring population density from predictions of habitat suitability might be.

For most models, the performance metrics were calculated by comparing against the full independent dataset (i.e., including data from all five of the Chatham Rise photographic surveys), with modeled taxa being considered for assessment only if they could be matched reliably with taxon names in the independent dataset and were present at 10 or more sites. However, because most of the models developed by Compton et al. (2013) were constructed using taxon occurrence data from TAN0705, data from this voyage were excluded from the test set for assessment of models from this study, with the exception of those for *Taiaroa tauhou*, *Hyalinoecia tubicola*, and Serolidae, which were based solely on physical specimen data.

We generated two graphs to compare AUC_{ind} against AUC_{kcv}. First, we plotted all AUC values in chronological order of the studies, together with mean values for AUC_{kcv} and AUC_{ind} per study. Second, we plotted AUC_{ind} against AUC_{kcv}, with results viewed in the context of how similar the two values were (proximity to a 1:1 regression line) and how they placed in relation to a threshold value of 0.7. For Anderson et al. (2016b), we plotted both of the AUC values reported for each of the 8 taxa they modeled: one calculated using random cross-validation sites, the other using spatially discrete (i.e., in longitudinal bands) sets of sites. We also visualized trends in model performance in relation to taxonomic resolution by plotting AUC_{kvc} and AUC_{ind} values against taxon level (Class, Order, Family, Genus, and Species), and by plotting model sensitivity (true positive rate) and specificity (true negative rate) in relation to the independent data against taxon level. In this analysis, the combined reef-forming coral grouping (REEF) was assigned to Family, and Scaphopoda was assigned to Genus, rather than Class because all recent specimen records from Chatham Rise are of *Fissidentatum* spp. (NIWA Invertebrate Collection, unpublished data).

To compare model performance against the *expected* rank performance generated using the Araujo et al. (2019) criteria, and to assess potential trends of improving model performance (prediction success) with time, we generated a "realized" ranking of the models by comparing the mean values of each model's rank scores across the five assessment metrics listed above. The expectation, again, was that models with higher expected performance should also rank higher in terms of realized performance.

RESULTS

Expected Performance

Against the standards of Araujo et al. (2019), the two most recent studies, Anderson et al. (2016b) and Georgian et al. (2019), ranked highest. Below these, however, there was no clear temporal trend in expected performance (**Supplementary Table 1**). All studies were assessed as being "deficient" against standards for dealing with uncertainty in predictor variables (Issue 2C), while later models were assessed to be improvements in terms of treatment of bias and noise in response variables (Issue 3B), treatment of collinearity (Issue 3C), dealing with modeling and parameter uncertainty (Issue 3D), and measures of model performance (Issue 4C). It was also noted that the two most recent studies made allowance for spatial autocorrelation, which is not listed explicitly in the Araujo et al. (2019) criteria (**Supplementary Table 1**). The final consensus rank order of the studies from highest to lowest was Georgian et al. (2019) > Anderson et al. (2016b) > Anderson et al. (2016a) = Compton et al. (2013) > Tracey et al. (2011) > Anderson et al. (2015) = Anderson et al. (2014) = Baird et al. (2013).

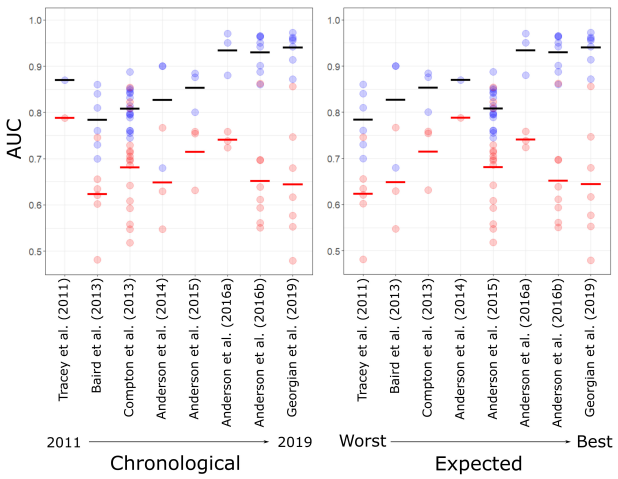

FIGURE 2 | AUC values generated using internal cross-validation (blue) and independent test data (red) for individual habitat suitability models in eight published studies (S1–S8, see **Table 1** for details) in order of time of publication from 2011 to 2019 **(Left)**, and *a priori* ranking of expected model performance based on the criteria of Araujo et al. (2019) **(Right)**. Cross bars show mean values for each study.

Model Assessment

AUC and TSS

AUC values from cross validation in the published studies (AUC_{kcv}) increased both with time and when ordered by expected rank performance, with all but one model ("REEF" in Anderson et al., 2014) scoring at least 0.7 and all models in the two most recent studies scoring higher than 0.85 (**Figure 2**). AUC values based on the independent test data (AUC_{ind}), by contrast, did not show matching increases over time and were lower than corresponding AUC_{kcv} values for all but two of the models, and values of 0.7 or higher were recorded for only 18 of the 47 models (38.3%), nine of these coming from a single study (Compton et al., 2013) (**Figure 3**).

The two models that scored more highly against the independent test data ($AUC_{ind} > AUC_{kcv}$) were those for the molluscan class Scaphopoda in Compton et al. (2013) and the stony coral *G. dumosa* in Anderson et al. (2016b). For the latter model, however, this was only the case when AUC_{kcv} was calculated using spatial banding in cross validation. Of the remaining 29 models, 22 had AUC_{ind} values of less than 0.65, despite all but one of these (REEF in Anderson et al., 2014) scoring higher than 0.7 by internal cross-validation. The exceptions here, again, were the models in Anderson et al.

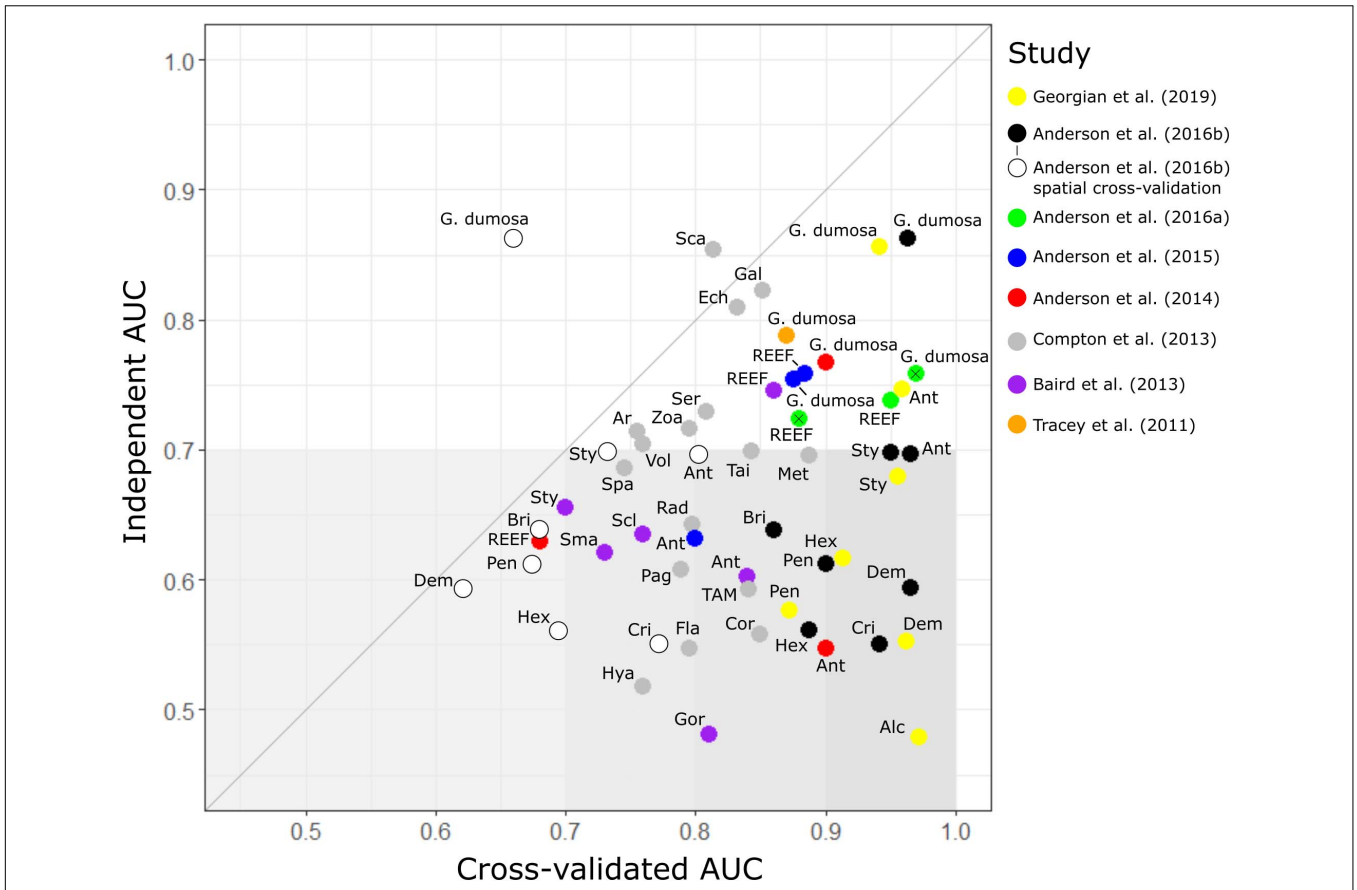

FIGURE 3 | Model assessment. Comparison of benthic invertebrate species-distribution model performance (as area under the receiver operator curve, AUC) when assessed against sample data withheld from the original dataset used to build the models (internal AUC) and an independent dataset of faunal distributions derived from photographic surveys (independent AUC). The 48 models are from eight published studies (see text for details) and cover 29 taxa: Alcyonacea (Alc); *Anthomastus robustus* (Ar); Antipatharia (Ant); Brisingidae (Bri); coral reef-forming taxa (REEF); Corallimorpharia (Cor); Crinoidea (Cri); Demospongiae (Dem); *Gracilechinus multidentatus* (Ech); Echinothuroida (TAM); *Flabellum* spp. (Fla); Galatheidae (Gal); *Goniocorella dumosa*; Hexactinellida (Hex); Gorgonacea (Gor); Hexactinellida (Hex); *Hyalinoecia tubicola* (Hya); *Metanephrops challengeri* (Met); Paguridae (Pag); Pennatulacea (Pen); *Radicipes* spp. (Rad); Scaphopoda (Sca); Scleractinia (Scl); Serolidae (Ser); small solitary corals (Sma); Spatangidae (Spa); Stylasteridae (Sty); *Taiaroa tauhou* (Tai); Volutidae (Vol), and Zoanthidae (Zoa). Anderson et al. (2016a) modeled REEF and *G. dumosa* with BRT and MaxEnt separately; MaxEnt results are indicated with a cross. Anderson et al. (2016b) calculated AUC in two ways: random selection of sites (black plots) and in longitudinal bands (white plots). Gray line shows a 1:1 relationship between internal and independent AUC scores, and gray shading indicates independent AUC scores less than 0.7, a value above which models are considered to be potentially useful for prediction, with darker shading to highlight the largest discrepancies between internal and independent AUC values.

(2016b) in which spatial banding was used to generate AUC_{kcv}. For these models, AUC_{kcv} values were less than 0.7 and closer to, but still higher than, their AUC_{ind} values.

The highest published cross-validation values ($AUC_{kcv} > 0.9$) were all from the three most recent studies (Anderson et al., 2016a,b; Georgian et al., 2019) but the corresponding AUC_{ind} values for these studies ranged widely, including both the highest (0.86 for *G. dumosa*) and lowest (0.52 for Alcyonacea) scores. Only two of the seven models from Georgian et al. (2019) and three of the eight from Anderson et al. (2016b) scored AUC_{ind} values of 0.7 or higher (*G. dumosa* and Antipatharia in both studies, and Stylasteridae in Anderson et al., 2016b) but both the MaxEnt and BRT models for REEF from Anderson et al. (2016a) scored above 0.7. There was a trend for AUC_{ind} to increase at finer taxonomic resolution but this was not matched in AUC_{kvc} values (**Supplementary Figure 1**), with strongly divergent values at Class level becoming more similar to AUC_{ind} values at finer resolutions. The increasing trend in AUC_{ind} at finer taxonomic resolution was associated with increases in the true positive rate (sensitivity), rather than the true negative rate, which showed no trend across taxonomic levels.

True skill statistic was strongly correlated with AUC_{ind} ($R^2 = 0.92$). Only six models yielded TSS values greater than 0.5, three of these scoring 0.6 or higher (**Table 3**), and the best-performing models were the same as identified by the AUC analysis.

t-Tests

For 35 of the 47 published models (74.5%), mean predicted probability of suitable habitat was significantly higher (TS, $p < 0.01$) at sites where the modeled taxon was present, rather than absent, in the independent data, with another four

TABLE 3 | Model assessment results.

Study	Taxon	AUC		TSS	*t*-test	Goodness-of-fit		Sites	
		Cross validation	Independent		$\bar{x}P > \bar{x}A$	Prevalence R^2 (mean ± 1 sd)	Density R^2	All	Presence
Tracey et al., 2011	*Goniocorella dumosa*	0.87 ± 0.012	0.79	0.55	TS	0.074 ± 0.008	0.071	237	53
Baird et al., 2013	Antipatharia	0.84	0.60	0.22	T	0.1 ± 0.03	0.022	296	31
	Stylasteridae	0.70	0.66	0.27	TS	0.057 ± 0.015	0.081	341	85
	Gorgonacea	0.81	0.48	0.05	NS	0.007 ± 0.004	0.019	341	149
	Scleractinia	0.76	0.64	0.30	TS	0.082 ± 0.040	0.182	341	232
	Coral-reef	0.86	0.75	0.41	TS	0.090 ± 0.008	0.235	238	65
	Small solitary	0.73	0.62	0.19	TS	0.049 ± 0.008	0.230	288	178
Compton et al., 2013	*Anthomastus robustus*	0.755 ± 0.043	0.71	0.42	TS	0.086 ± 0.021	0.040	204	61
	Corallimorpharia	0.849 ± 0.071	0.56	0.19	NS	0.030 ± 0.001	0.010	243	25
	Flabellum spp.	0.795 ± 0.042	0.55	0.10	NS	0.005 ± 0.003	<0.001	237	142
	Galatheidae	0.852 ± 0.041	0.82	0.50	TS	0.183 ± 0.030	0.070	243	180
	Hyalinoecia tubicola	0.760 ± 0.027	0.52	0.15	NS	0.011 ± 0.005	0.001	237	88
	Metanephrops challengeri	0.888 ± 0.031	0.70	0.39	TS	0.080 ± 0.002	0.077	237	76
	Radicipes spp.	0.798 ± 0.031	0.64	0.29	TS	0.046 ± 0.019	0.002	243	43
	Scaphopoda	0.814 ± 0.043	0.85	0.75	TS	0.099 ± 0.035	0.037	145	9
	Serolidae	0.808 ± 0.054	0.73	0.42	TS	0.174 ± 0.018	0.143	198	66
	Spatangidae	0.745 ± 0.042	0.69	0.32	TS	0.093 ± 0.021	<0.001	237	133
	Taiaroa tauhou	0.843 ± 0.042	0.70	0.43	TS	0.065 ± 0.031	<0.001	198	88
	Volutidae	0.760 ± 0.037	0.70	0.38	TS	0.055 ± 0.023	0.028	243	114
	Zoantharia	0.795 ± 0.042	0.72	0.38	TS	0.051 ± 0.014	0.068	204	42
	Paguridae	0.789 ± 0.048	0.61	0.22	T	0.021 ± 0.008	0.026	184	152
	Gracilechinus multidentatus	0.832 ± 0.038	0.81	0.59	TS	0.420 ± 0.034	0.186	190	48
	Echinothurioida	0.841 ± 0.054	0.59	0.27	TS	0.013 ± 0.002	0.038	204	79
Anderson et al., 2014	Coral-reef	0.68	0.63	0.33	TS	0.037 ± 0.001	0.023	236	63
	Goniocorella dumosa	0.97	0.77	0.54	TS	0.083 ± 0.004	0.069	232	53
	Antipatharia	0.98	0.55	0.26	NS	0.011 ± 0.005	0.021	296	31
Anderson et al., 2015	Coral-reef	0.884	0.76	0.48	TS	0.115 ± 0.027	0.105	237	65
	Goniocorella dumosa	0.876	0.75	0.43	TS	0.111 ± 0.011	0.144	231	53
	Antipatharia	0.800	0.63	0.33	TS	0.240 ± 0.038	0.339	295	30
Anderson et al., 2016a	Coral-reef (MXE)	0.880	0.72	0.43	TS	0.016 ± 0.021	0.027	237	65
	Coral-reef (BRT)	0.950	0.74	0.36	TS	0.014 ± 0.024	0.027	236	64
	Goniocorella dumosa (MXE)	0.97	0.76	0.50	TS	0.032 ± 0.006	0.050	231	53
Anderson et al., 2016b	Brisingidae	0.860 (0.680)	0.64	0.27	TS	0.013 ± 0.010	0.090	339	99
	Antipatharia	0.965 (0.803)	0.70	0.34	TS	0.043 ± 0.009	0.014	294	29
	Stylasteridae	0.950 (0.733)	0.70	0.31	TS	0.069 ± 0.011	0.067	339	83
	Crinoidea	0.942 (0.772)	0.55	0.11	NS	0.081 ± 0.071	0.708	339	70
	Demospongiae	0.965 (0.622)	0.59	0.16	TS	0.099 ± 0.019	0.201	339	234
	Goniocorella dumosa	0.963 (0.659)	0.86	0.61	TS	0.122 ± 0.033	0.208	230	53
	Hexactinellida	0.887 (0.696)	0.56	0.13	T	0.040 ± 0.018	0.079	339	116
	Pennatulacea	0.901 (0.674)	0.61	0.22	TS	0.121 ± 0.015	0.020	294	157
Georgian et al., 2019	Antipatharia	0.959	0.75	0.45	TS	0.089 ± 0.007	0.043	294	29
	Stylasteridae	0.956	0.68	0.27	TS	0.134 ± 0.006	0.097	339	82
	Demospongiae	0.962	0.55	0.11	NS	0.020 ± 0.006	0.071	339	233
	Goniocorella dumosa	0.942	0.86	0.60	TS	0.161 ± 0.013	0.105	229	53
	Hexactinellida	0.914	0.62	0.20	TS	0.046 ± 0.007	0.040	339	115
	Pennatulacea	0.872	0.58	0.17	T	0.071 ± 0.004	0.080	294	157
	Alcyonacea	0.972	0.48	0.03	NS	0.006 ± 0.004	0.030	339	233

Original cross validation AUC values extracted from the published studies; AUC values calculated against the independent dataset; correlation strength between predicted probability of presence and prevalence values from the independent dataset (R^2 mean ± 1sd calculated from prevalence results from three transect segment lengths; 1, 5, and 10 min); correlation strength between predicted probability of presence and population density values from the independent dataset (density R^2); significance of t-tests for the hypothesis that a taxon is more likely to be present than absent at sites where the published studies predict it to be present ($\bar{x}P > \bar{x}A$: NS, $p \geq 0.05$; T, $0.05 \geq p \geq 0.01$; TS, $p \leq 0.01$); the total number of sites available for each comparison (All), and the total number of sites at which each taxon was present in the test dataset (presence). Cross validation AUC values from Tracey et al. (2011) and Compton et al. (2013) are shown as means ± 1 standard error, and for Anderson et al. (2016b) as results of both random sample cross-validation and spatially banded cross-validation (in parentheses).

models included at the lower significance level (T, $p < 0.05$) (**Table 3**). For the remaining 8 models ("NS" results), the mean predicted probability of suitable habitat was never higher for absence sites than for presence sites. Twelve of the TS models were from Compton et al. (2013) but there were significant (TS) results for models in all studies and the proportions of significant results in each of the studies that modeled more than 3 taxa were broadly comparable: Baird et al. (2013), 66.6%; Compton et al. (2013), 75.0%; Anderson et al. (2016b), 75.0%, and Georgian et al. (2019), 57.1%. The models that scored as TS spanned a range of taxonomic levels, including species (*Goniocorella dumosa*, *Anthomastus robustus*, *Metanephrops challengeri*, *Taiaroa tauhou*, and *Gracilechinus multidentatus*), Genus (*Radicipes*), Family (Galatheidae, Serolidae, Spatangidae, Volutidae, and Stylasteridae), Order (Scleractinia, Alcyonacea, Antipatharia, Brisingida, Zoantharia, and Pennatulacea), and Class (Demospongiae and Hexactinellida). However, except for *G. dumosa*, which was modeled in six of the eight studies, all of the species-level models and the models for *Radicipes*, Galatheidae, Serolidae, Spatangoida, Volutidae, and Brisingida were from the Compton et al. (2013) study, which used occurrence data primarily from photographic sampling.

Correlations With Taxon Prevalence and Density

Correlations between predicted probability of suitable habitat in the published models and values of both prevalence and population density in the independent dataset were weak ($R^2_{prev} < 0.25$ for 46 of the 47 models, and $R^2_{dens} < 0.25$ for 45 of the 47 models). The three cases where correlation strength exceeded 0.25 were models for the echinoid *G. multidentatus* ($R^2_{prev} = 0.42$ in Compton et al., 2013), the black coral Order Antipatharia ($R^2_{dens} = 0.34$ in Anderson et al., 2014), and the Order Crinoidea ($R^2_{dens} = 0.71$ in Anderson et al., 2016b). Across all other models, the mean correlation strength against both prevalence and density values in the independent dataset was less than 0.1 (mean ± sd: $R^2_{prev} = 0.071 \pm 0.052$, $R^2_{dens} = 0.078 \pm 0.121$). The high R^2_{dens} value for Crinoidea in Anderson et al. (2016b) was driven by very high densities of crinoids on the Graveyard seamounts, which were within the area of high predicted probability of suitable habitat in this model.

Realized Rank Performance

When the models were ranked by their average rank results across all evaluation metrics, there was a broad spread of performance within and among studies. The 10 highest-ranked models were spread across 6 studies from 2013 to 2019, the 10 lowest-ranked models included two from the most recent study, and mean ranking by study showed no indication of a general trend of improvement over time (**Figure 4**). Of the four taxa that were modeled in more than two studies (*G. dumosa*, Coral reef, Stylasteridae, and Antipatharia), *G. dumosa* was the most consistently highly ranked, with five of its six models in the top fifteen. *G. dumosa* models also showed some indication of improving performance over time, as did Stylasteridae, with models from the two most recent studies ranking higher than those from earlier studies (**Figure 4**).

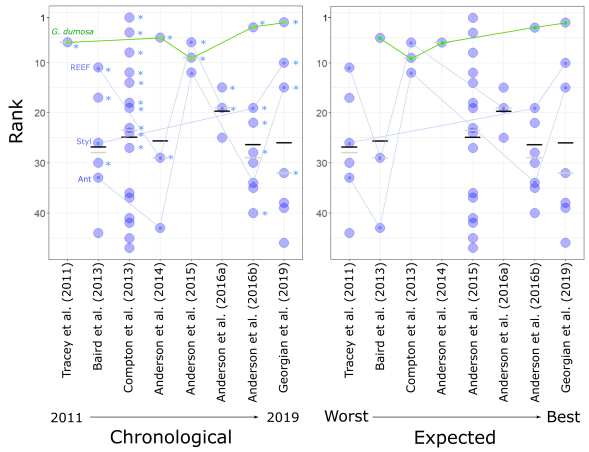

FIGURE 4 | Rank performance of 47 predictive models, showing average rank across four evaluation measures (TSS, *t*-test, R^2_{prev}, and R^2_{dens}) for each model (blue dots), and the mean (black bars) and median (gray bars) ranking for each study (S1–S8). **(Left)** studies in chronological order, asterisks indicate models for which the average predicted probability of suitable habitat for the target taxon was significantly higher (*t*-test, $p < 0.01$, "TS") at sites where that taxon was present in the independent dataset than at those where it was absent. Four taxa that were modeled in more than two studies are linked by lines (see **Figure 3** for full names). **(Right)** with studies ordered by their expected rankings as assessed by reference to the standards of Araujo et al. (2019).

The spread of high- and low-ranked models across studies was such that no overall ranking of the studies could be assigned with confidence (**Figure 4**). There was, however, neither evidence for substantially improved performance from earlier to later models nor support for the rankings assigned by reference to Araujo et al. (2019) prior to the evaluation exercise.

DISCUSSION

In this study, we have used independent data from seabed photographic surveys to explore the general utility of habitat suitability models that we have developed over more than 10 years with the aim of predicting distributions of seafloor taxa in the southwest Pacific, centered on New Zealand. The key results of our assessment are that (1) measured model performance was lower when assessed against independent data than by k-fold cross-validation for all but two of 47 models; (2) a trend of increasing model performance with time, which is seen in published cross-validation (AUC_{kcv}) values and is anticipated when the methods used in these studies are judged against objective criteria, is not supported when the models are tested against independent data; (3) for approximately 72% of the models, predicted probability of suitable habitat in the models was significantly higher at sites where a taxon was present in the independent data than where it was absent; and (4) correlation strengths between predicted probability of presence and observed taxon prevalence and density were weak.

While the third result here is the only statistic that offers support for the expectation that such models might be reliable for

predicting distributions, and then only at the level of prediction of presence, the results overall should also be viewed in the context of how realistic our expectations of such models are. A key aspect here is that Chatham Rise is in a highly dynamic oceanographic environment and encompasses a wide range of seafloor topographies within a relatively confined spatial extent (Nodder et al., 2012). Thus, although the Rise is one of the areas within New Zealand's EEZ that we are most interested in predicting to, because of its importance to commercial fisheries (Fisheries New Zealand, 2020) and potential mineral interests (Von Rad and Kudrass, 1987), it is also likely to be one of the most challenging. Perhaps more importantly for future work in this field, the evaluation exercise affords the chance to review our expectations and to explore which aspects of the models, in terms of input data, spatial scope, and modeling methods, contribute most to the observed differences in performance against the independent data. We use this evaluation to suggest future directions for model building that should produce models that can be used with greater confidence for environmental management.

Cross-Validation and Non-independence of Data

Several studies have demonstrated that model performance metrics generated by the common practice of cross-validation using withheld subsets of the input sample data will yield inflated values (e.g., Bahn and Mcgill, 2013; Valavi et al., 2019) because the withheld data are not independent of those used to train the model, particularly with respect to spatial autocorrelation (Bahn and Mcgill, 2007; Ploton et al., 2020). It is interesting here, however, that AUC values for most models in the two studies that made explicit allowance for spatial structure in the sample data, whether by withholding data in longitudinal bands (Anderson et al., 2016b) or by including spatial autocorrelation as a predictor variable (Georgian et al., 2019), were still inflated by comparison with those calculated against the independent data. This finding suggests that neither of these methods entirely overcame the issue of non-independence of data and, thus, that issues associated with using cross validation as a primary method for model performance are not easily overcome. Awareness of the need to account for spatial structuring of data in habitat suitability models is increasing, and availability of new, more flexible, tools now allows for more nuanced approaches that are likely to improve estimation of predictive performance by cross-validation (Valavi et al., 2019).

Regardless of the absolute values obtained from AUC analyses, our finding that the trend of increasing model performance with time was not supported against the independent data is concerning for two reasons: firstly, because our results show that we do not appear to be getting substantially better at describing distributions and, more importantly, because if apparent increases in performance encourage overconfidence in predictions from more recent models, it could lead to poor environmental management decisions (Regan et al., 2005). If our modeling methods have, indeed, improved over time, however, this result is also revealing because it suggests that the main impediments to accurate prediction are associated primarily with the quality and quantity of the input sample and environmental data, rather than with the details of specific modeling methods. This suggestion is further supported by the broad spread of AUC_{ind} values within individual studies in our results, with both the highest and lowest values being for models generated in the most recent, most technically sophisticated study (Georgian et al., 2019), and comparably high and low values recorded from earlier studies.

In our initial assessment of the studies against the standards of Araujo et al. (2019), the principal areas we characterized as being either "deficient" or "bronze" in all studies were understanding variability and uncertainty in the predictor variables (Issue 2C), and dealing with modeling and parameter uncertainty (Issue 3D). It was also clear, however, that the reliability and precision of taxon identification (Issue 1B) were likely to influence model performance. Thus, although our ranking was at the level of study, any assessment of how well taxon identification had been addressed in studies that covered multiple taxa would ideally be at the level of individual models, rather than the whole study, because of wide differences in how taxa were grouped.

Uncertainty in Predictor Variables

Uncertainty associated with the predictor variables used in the models was the area of greatest concern in the initial model assessments, with questions around the lack of some key ecologically relevant variables, limitations of spatial resolution, and the reliability of predictor layers that are themselves outputs from spatial modeling or interpolation processes (Davies and Guinotte, 2011). These issues affect all broad-scale habitat suitability modeling initiatives in the deep sea and present unique challenges by comparison with terrestrial studies, which often have the benefit of greater accessibility for direct sampling and full-coverage, high-resolution, remote sensing by satellite (e.g., Pearce and Ferrier, 2001; Parmentier et al., 2011; Ploton et al., 2020).

The lack of key variables is a fundamental issue affecting prediction of the distributions of seabed fauna. Substrate type in particular is a determinant of realized distributions for most benthic taxa, but our knowledge about the occurrence of substrate types in the deep sea at anything beyond highly local scales is of qualitatively the same type as our knowledge of the fauna we are interested in predicting: patchy, spatially auto-correlated, point sample records collated from multiple sources. Despite recent initiatives to generate continuous substrate-type layers by interpolation among point samples in our region (e.g., Bostock et al., 2019), these characteristics currently render such layers unreliable for use in predictive models (Georgian et al., 2019). In a study area that has been subject to modification by bottom-contact trawl fisheries for decades (Bowden and Leduc, 2017; Baird and Mules, 2019; Clark et al., 2019), it is also of note that none of the models assessed here included fishing effort as a predictor variable. While fishing might be expected to vary in location and intensity over finer spatial and temporal scales than most environmental variables, and thus have inconsistent influence on realized faunal distributions, it is also likely that any influence it does exert is likely to be strong in some habitats. The present distributions of cold-water scleractinian corals on seamounts and other topographic features that have been targeted by trawling, for instance, have been modified from

their natural state (Williams et al., 2010, 2020; Clark et al., 2019) and thus are unlikely to be predicted accurately by models that do not incorporate fishing effort as a predictor of presence.

While some environmental variables commonly used in deep-sea habitat suitability models are derived directly from full-coverage satellite remote sensing (e.g., sea surface temperature, chlorophyll a concentration), or acoustic remote sensing [e.g., multibeam echosounder (MBES) for smaller-scale studies], many others are derived indirectly from discrete sample data (e.g., single-beam acoustic soundings, CTD casts, and Argo floats), either by spatial interpolation (regional bathymetry and, thus, all topographic variables derived from it, including seabed slope, curvature, rugosity, and position index) or via modeling of physical (e.g., seabed currents and temperature), chemical (e.g., salinity and aragonite saturation), or biogeochemical (organic carbon flux to the seabed) processes. Furthermore, in modeling studies of seafloor fauna, values for individual grid cells are necessarily extracted by reference to the bathymetry layer (Davies and Guinotte, 2011), which as noted above, at spatial scales greater than local MBES surveys, carries its own unquantified uncertainty. Thus, all of the environmental data layers relied upon as predictor variables in habitat suitability modeling initiatives in the deep sea introduce some degree of additional, usually unquantified, uncertainty into the final predictions.

Formal analysis of the influence of inaccuracies in environmental variable layers used as predictors in SDM models is beyond the scope of this study, but some of the issues are illustrated by one of the studies assessed here (Anderson et al., 2016a), in which we ran a purpose-designed photographic survey of seamount features in the Louisville Seamount Chain to assess the reliability of predictions from habitat suitability models we had generated for the entire SPRFMO Convention Area. We found that our models were not successful at predicting occurrence of scleractinian corals at the scale of the survey, despite high AUC values for the models from internal cross-validation. We attributed this failure primarily to inaccuracies in the bathymetry layer at the spatial resolution of the model and to the lack of a predictor variable describing substratum type. Inaccuracies in the bathymetry later were compounded in all other predictor variables derived from it, including seafloor slope and rugosity, while the absence of hard substrata across large proportions of the seamount summits confounded predictions of high habitat suitability because hard substrata are a fundamental habitat requirement for the corals we were predicting. While these factors were probably exacerbated by the steep topography and isolated oceanic context of the Louisville seamounts, results in the present assessment indicate that the Anderson et al. (2016a) models fare no better against survey data from Chatham Rise, where bathymetric data are much more reliable and where much of their input faunal occurrence data were collected.

Modeling and Parameter Uncertainty

Acknowledging uncertainty in the predictor variables leads to the issue of how to deal with modeling and parameter uncertainty (Issue 3D in Araujo et al., 2019) because in deep-sea models the predictor variables are likely to be the largest source of uncertainty, for the reasons discussed above. Modeling uncertainty was considered explicitly in only three of the eight studies considered here (Anderson et al., 2016a,b; Georgian et al., 2019), but in each case it was quantified only in terms of the influence either of using different subsets of the response variable (taxon) data, or of using different modeling methods, or both, with no quantification of the uncertainty associated with the environmental predictor layers used. Thus, for all studies considered here, the largest potential source of uncertainty in the final model predictions remained unquantified. Our current inability to account for the uncertainty associated with the environmental layers used as predictors in habitat suitability models for the deep sea is a major impediment to increasing confidence in the predictions of such models (e.g., Kenchington et al., 2019).

Another rarely acknowledged source of uncertainty in habitat suitability models for the deep sea is that taxon occurrence data are, in most cases, compiled from sources that span periods of years, decades, or even centuries. This is a practical way to compensate for the general paucity of data available from deep-sea environments, which results from the logistical difficulty and cost of sampling at depth (Clark et al., 2016). All but one of the studies assessed here used data compiled over extended periods (e.g., Tracey et al., 2011 used coral occurrence data from 1955 to 2009), the exception being the study of Compton et al. (2013), in which models were trained on data from surveys conducted within 2 years of each other and only 2 years before the modeling work was undertaken. Two important assumptions are implicit when occurrence data are accumulated over extended periods: first, that patterns of occurrence will not have changed materially during the entire period from the date of the first occurrence record to production of the model predictions, and second, that the environmental characteristics used as predictors (the summaries for which are likely to represent somewhat different periods to those over which taxon records are accumulated) will not have changed materially either. For long-lived sessile taxa, such as cold-water corals, the first assumption might be reasonable in many cases. However, with increasing evidence of the effects of bottom-contact fishing and other anthropogenic and natural disturbances on realized occurrences (e.g., Clark et al., 2000, 2019, Mountjoy et al., 2018), and parallel increases in our understanding of the rates of large-scale environmental change resulting from global warming (Smith et al., 2009; Hoegh-Guldberg and Bruno, 2010; Doney et al., 2012), such assumptions become increasingly tenuous.

Taxonomic Resolution

Our study showed that there was a tendency for models of finer taxonomic levels to perform better against the independent data than those at a coarser level (**Supplementary Figure 2**). Aggregating records to coarser taxonomic groupings is common in studies of deep-sea benthic invertebrate distributions, where records are often collated from multiple sources at differing levels of identification and where available records at finer levels (species or genus) can be too sparse to inform habitat suitability models. This issue is not covered explicitly by Araujo et al. (2019) but is likely to have a strong influence on the predictive success of models because coarser taxonomic groupings encompass taxa with different adaptations and

environmental tolerances, which, when combined in a single model, may lead to predicted distributions that are too general to be useful. Effects of aggregating to coarser taxonomic levels are evident in our results, with the broadest groups modeled, including Alcyonacea (soft corals), Gorgonacea (gorgonian corals), Demospongiae (sponges), and Hexactinellida (glass sponges), generating among the lowest AUC_{ind} values and overall rank performances. An important observation here is that the models for these taxonomic groups are also potentially the most misleading because, in contrast to their performance against the independent data, they scored highly when assessed by cross validation—all models for the VME indicator taxa Demospongiae and Hexactinellida in the most recent studies (Anderson et al., 2016b; Georgian et al., 2019), for instance, yielding AUC_{kcv} values greater than 0.87, indicating "reliable" or "excellent" performance. While these contrasting patterns of model performance in relation to taxonomic resolution are intriguing and suggest an important direction for further investigation, the results here should be viewed with some caution because of confounding factors in the data available to this study. For instance: the numbers of taxa within taxonomic levels are unequal; taxonomic levels are not represented evenly across studies, thus introducing potential methodological bias; some reported taxonomic levels are potentially inaccurate (as we determined for the group Scaphopoda), and most of the species-level models are for a single taxon, *G. dumosa*, which was modeled using essentially the same response variable data in all models.

These issues notwithstanding, comparison between two taxa with the highest- and lowest-ranked models in our analysis, the scleractinian coral *G. dumosa* and sponges (Porifera, modeled as two Classes: Demospongiae and Hexactinellida, in Anderson et al., 2016b; Georgian et al., 2019), illustrates the probable influence of taxonomic resolution on model performance. *G. dumosa* is consistently identified to species level because it has a protected status in New Zealand and is of high conservation interest due to its provision of complex biogenic habitat. However, *G. dumosa* occurrence records are also clustered within relatively narrow environmental and spatial bounds, with the highest density of records used to inform all of the models assessed here coming from Chatham Rise itself (e.g., Tracey et al., 2011; Anderson et al., 2016b). Sponges, by contrast, are highly diverse and difficult to identify to species and thus are typically modeled at the coarse taxonomic level of Class (Demospongiae and Hexactinellida). Grouped at this level, occurrence records for sponges are spread much more widely across environmental gradients than would be the case for individual species. Given these differences in their input data, the task of modeling distributions is clearly simpler for *G. dumosa* than for the sponge classes and it is not surprising that their respective models ranked as they did, despite being modeled using the same methods and the same predictor variables.

The study of Compton et al. (2013) is interesting here because, unlike all the other studies, its models were based on data from two surveys designed specifically to sample seafloor invertebrate communities using high-resolution photographic transects and epibenthic sled samples. Only one of the surveys covered Chatham Rise (TAN0705), and the resulting density of sample points for both presences and absences was, therefore, lower than for the other studies. However, because photographic survey methods yield close to 100% detection of epibenthic invertebrates, a high proportion of taxa were identified reliably and consistently to species or genus level and it was possible to use true absence data, rather than random background absences, target-group absences, or pseudo-absences. We expected the fine taxonomic resolution and availability of true absence data to yield improvements in model predictions compared to other studies, but while some models from the study are among the highest ranked in our assessment (e.g., Scaphopoda, Galatheidae, and *Gracilechinus multidentatus*), others are among the lowest (e.g., *Hyalinoecia tubicola*, *Flabellum* spp.) and the overall range of results is comparable with other studies. Given that the sampling methods and taxonomic resolution scored highly against the evaluation criteria, the two key aspects remaining that might explain the overall performance are, again, uncertainty in the environmental predictor variables and the relatively low density of sampling for the response variables.

Measuring Prediction Success

The metrics used to evaluate models here were chosen to assess how useful the model predictions are likely to be in practical applications, the primary intended use for such predictions being to inform management and conservation decisions across a range of spatial scales (km to 100s km) within the model domain (Anderson et al., 2016a; Araujo et al., 2019; Winship et al., 2020). Thus, in addition to the well-established AUC and TSS metrics, we used the three simpler measures that were intended to reflect naïve questions about a taxon's distribution at differing levels of predictive skill: is it likely to be present at a given site? (*t*-tests); what is its prevalence likely to be at that site? (R^2_{prev}), and what is its abundance likely to be at that site? (R^2_{dens}). While these are simplistic measures of model performance, not least because any relationship between predicted probabilities and measured occurrences is unlikely to be linear and realized occurrences and densities are likely to be influenced by historical events, ecological interactions, and stochastic processes (Dayton and Hessler, 1972; Connell and Slatyer, 1977; Connolly and Roughgarden, 1999), they provide intuitively interpretable measures of how well the model predictions match the independent observations and, thus, our expectations of a model's predictive ability.

The models evaluated here can only predict the probability of suitable habitat being present at a given site, which in itself is of limited use for most applications (Bahn and Mcgill, 2013), but continuous maps based on these probabilities inevitably invite the interpretation that higher predicted habitat suitability should correspond with higher population densities of the target taxon (Lobo et al., 2008). This interpretation is not justifiable from a theoretical perspective, but it is, arguably, the way in which outputs from habitat suitability models are often viewed. Indeed, it is arguable that if such an interpretation is not at least partially justified, we should question what purpose such predictions serve, if not to indicate where a taxon is most likely to be found. In this context, an important result here is that most correlations between predicted probability of suitable habitat being present and observed densities of taxa on the seabed were weak. This is a practical demonstration that inferring the likelihood of taxon

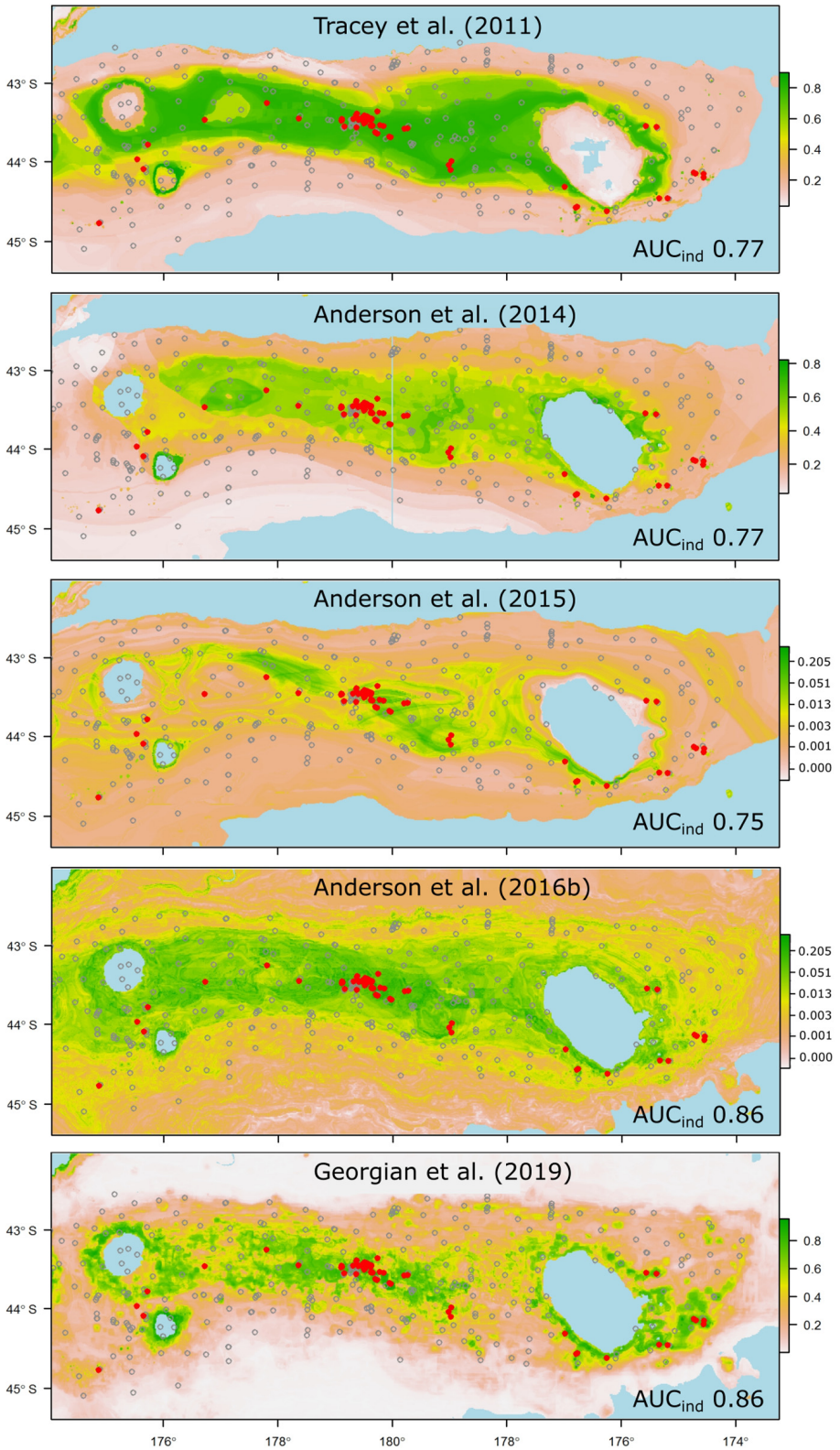

FIGURE 5 | Probability of suitable habitat for the branching scleractinian coral *Goniocorella dumosa* on Chatham Rise, as predicted by habitat-suitability models in five published studies. Probability values are scaled at right, and prediction maps are overlaid with presence (red dots) and absence (gray circles) locations from the independent photographic observation dataset. AUC values from testing against the independent dataset are shown for each model (AUC_{ind}).

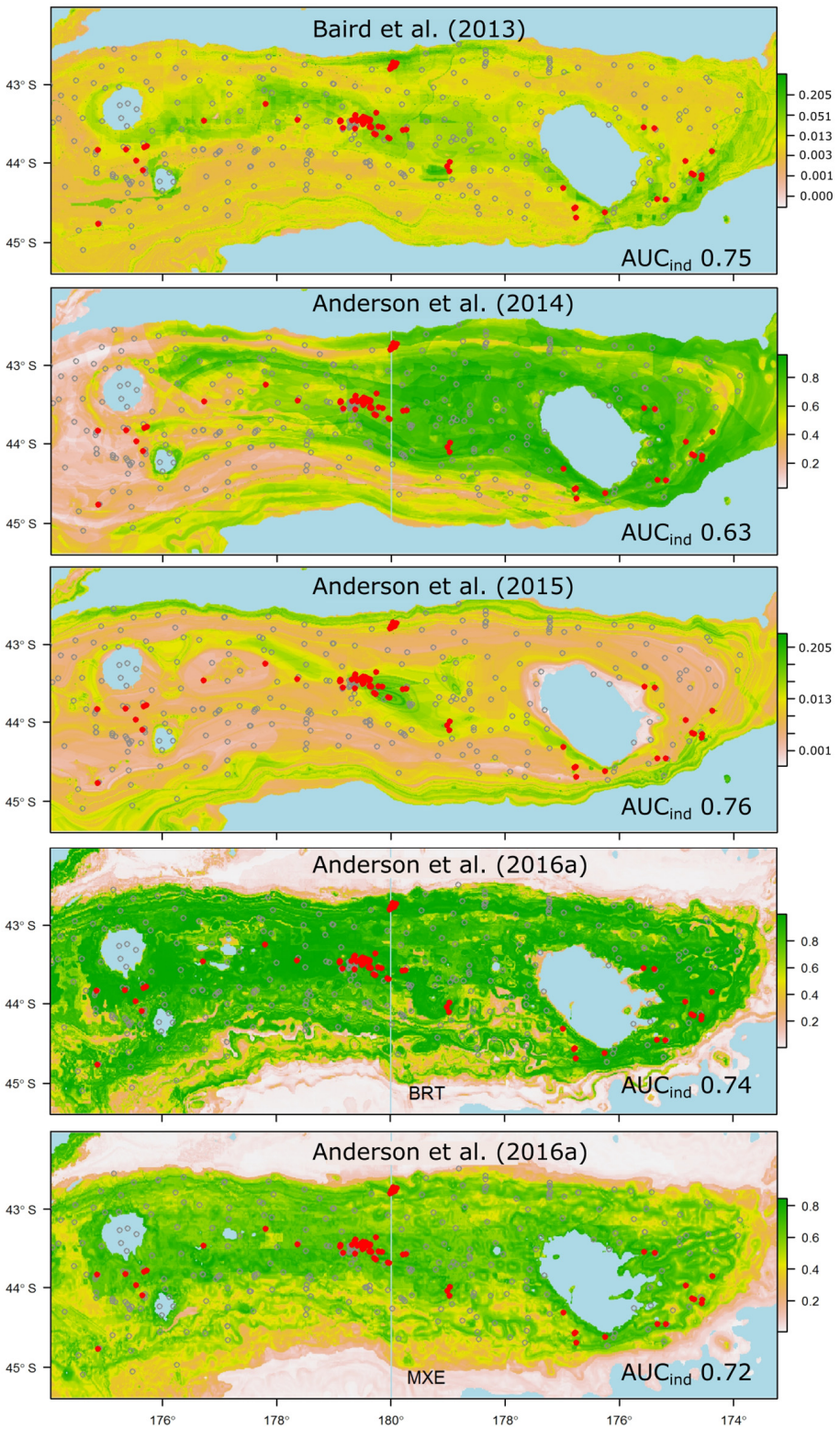

FIGURE 6 | Probability of suitable habitat for all thicket- or reef-building Scleractinian corals (*Solenosmilia variabilis*, *Madrepora oculata*, *Enallopsammia rostrata*, and *Goniocorella dumosa*) on Chatham Rise, as predicted by habitat-suitability models in four published studies. Anderson et al. (2016a) produced two models for reef-building corals, one using BRT, the other MaxEnt (MXE). Probability values are scaled at right, and prediction maps are overlaid with presence (red dots) and absence (gray circles) locations from the independent photographic observation dataset.

occurrence or population density from predictions of suitable habitat being present is, indeed, unlikely to be justified.

Prediction of presence or absence is less demanding than prediction of relative or absolute density (Bahn and Mcgill, 2013), so it is not surprising that models generally performed better against the presence–absence tests (AUC, TSS, and *t*-tests) than against the quantitative ones (R^2_{prev}, and R^2_{dens}), the most encouraging results being from tests of the simple expectation that modeled probabilities of suitable habitat would be significantly higher at sites where a given taxon was present in the independent data than at those where it was absent. At this level of predictive skill, 72% of models met the expectation at the more conservative level (TS, $p = 0.01$), offering some support for the utility of existing models in practical applications that require only prediction of likelihood of presence. However, that there was no pattern of general improvement against this test with increasing sophistication of modeling methods suggests that performance is limited more by the quality and quantity of input data than by analytical methods, and for the more interesting and potentially useful task of predicting taxon densities, there is no support.

Despite all but a few of the models performing less well than was anticipated from their original cross-validation scores, some emerge as being potentially useful for reliable prediction of taxon occurrence at unsampled sites. Thus, models that performed well against the independent data and predicted the occurrence of suitable habitat for taxa that have high management or conservation status might be used with some confidence in spatial management (Moilanen et al., 2006). In our case, these criteria would limit the set of potentially useful predictions to those for the scleractinian coral *G. dumosa* (**Figure 5**), and the combined grouping of reef-forming scleractinian corals (REEF, **Figure 6**). For the few taxa that have been modeled in more than one study, notably *G. dumosa*, models from the most recent studies (Anderson et al., 2016b; Georgian et al., 2019) performed better than earlier ones. However, the difference in performance between the earliest (Tracey et al., 2011) and latest (Georgian et al., 2019) models for *G. dumosa* was relatively minor, and the most recent studies also included some of the lowest-ranked models in our comparison. This finding suggests, again, that any benefits gained from refinement of modeling methods may be small by comparison with other aspects of the model-development process, including the quantity and spatial distribution of occurrence data, the taxonomic level and consistency of identifications, and the availability of reliable and appropriate environmental layers at ecologically relevant spatial scales.

CONCLUSION

For habitat suitability models to be useful in deep-sea environmental management applications, we need to have confidence that their predictions are reliable at appropriate spatial scales and taxonomic resolutions. A first step toward this should be routine use of cross-validation methods that account for spatial structuring in the input data. Reliability can only be confirmed, however, by assessing model predictions against independent data using methods that sample the target taxa effectively, our results demonstrating that such assessments can yield a very different picture of prediction success than is gained from cross-validation methods. While it is concerning that most of our current model predictions are apparently of limited use for their intended applications in management, the process of objective assessment helps to identify which aspects of the modeling process are most in need of improvement. Limitations in the quality and quantity of input data, for both response and predictor variables, appear to be the primary factors affecting prediction success, rather than details of the modeling methods used. If this is the case, increased confidence in the outputs from future models will probably be achieved only by greater investment in data collection and in quantifying the uncertainty associated with these data, for both response and predictor variables. Generating more reliable environmental data, particularly bathymetry, at spatial resolutions relevant to the habitat preferences of target taxa will be a critical component of this, with initiatives such as the Irish National Seabed Survey (O'toole et al., 2020) and Seabed 2030 Project (Mayer et al., 2018), exemplifying the scale of commitment required. In parallel with this, it seems likely that dedicated surveys of taxon distributions will always be necessary, both to validate existing models and to enhance data inputs for their successors.

Despite the generally disappointing performance of our models in this assessment, they can serve as useful heuristics if viewed as hypotheses to be tested. Approached in this way, we suggest that a practical long-term strategy to reduce uncertainty in model predictions can be structured around iterations of a 4-step cycle in which (1) initial habitat suitability modeling based on all available taxon occurrence data generates predictions, (2) these predictions are used to structure field validation surveys, (3) survey data are used for objective evaluation of prediction success, and (4) the survey data are then integrated with existing data and used to develop revised models. Predicting to overlapping seabed areas at each iteration of the cycle would progressively expand the environmental and spatial scope of the models while staying within the bounds of ecological credibility. The present study represents stage 3 of the first iteration of this cycle in New Zealand but as we note above, our results show that major sources of uncertainty in our models, including the quality, spatial resolution, and ecological relevance of the environmental predictor variables, have yet to be addressed adequately.

DATA AVAILABILITY STATEMENT

Data used in this study are publicly available via links in the published studies and reports available here: https://marinedata.niwa.co.nz/quantifying-benthic-biodiversity/.

AUTHOR CONTRIBUTIONS

DB, AR, and MC conceived the study. DB compiled independent test data. OA ran model assessments. OA and DB analyzed

results. DB led manuscript writing, with all authors contributing throughout the process.

FUNDING

This study was commissioned and funded by the New Zealand Ministry for Primary Industries (MPI), now known as Fisheries New Zealand (FNZ), under project ZBD201611 Quantifying Benthic Biodiversity, with additional support through NIWA Strategic Science Investment Fund project COES1801 Marine Food Web Dynamics. The following past and present agencies contributed to funding of the various models assessed in this study: MPI; the Foundation for Research Science and Technology; the Department of Conservation; and the Ministry of Business Innovation and Employment.

ACKNOWLEDGMENTS

We thank Mary Livingston for project governance at FNZ, Caroline Chin, Niki Davey, Alan Hart, Rob Stewart, and Megan Carter at NIWA for their meticulous image analysis work, and the officers and crew of RV Tangaroa on the camera survey voyages that made this analysis possible. Lastly, we acknowledge the initial thought- and action-provoking discussion with Alistair Dunn, Martin Cryer, Richard Ford, and Mary Livingston of MPI, who challenged us to demonstrate that the habitat suitability models we had developed for benthic fauna could be considered useful for fisheries management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.632389/full#supplementary-material

Supplementary Figure 1 | Illustration of prevalence calculation, showing the seabed track of a video transect (Voyage TAN0705, station 170) divided into 10-min segments (alternating blue–red), with individual observations of the taxon Paguridae (hermit crabs) indicated by black circles. Prevalence is calculated as the proportion of the total number of segments in which the taxon was observed, in this case 8/12 = 0.66. Also shown are the density of Paguridae as individuals 1,000 m^{-2}, and the probability of suitable habitat as predicted by the BRT model of Compton et al. (2013).

Supplementary Figure 2 | Variation in AUC and Sensitivity (true positive rate) with taxonomic resolution. **(Left)** AUC_{kvc} (by cross validation, blue) and AUC_{ind} (against independent data, red) values as described in **Figure 3**, plotted by taxonomic level. **(Right)** Sensitivity (true positive rate) for all models when assessed against the independent data. Horizontal bars show mean values in each taxonomic level.

Supplementary Table 1 | Author's ranking of eight published SDM studies by reference to assessment criteria and issues proposed by Araujo et al. (2019). DB, David Bowden; OA, Owen Anderson; AR, Ashley Rowden; FS, Fabrice Stephenson. Rankings from best to worst are: **g**old (G); **s**ilver (S); **b**ronze (B), and **d**eficient (D).

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., Clark, M. R., Mormede, S., Davies, A. J., et al. (2016a). Field validation of habitat suitability models for vulnerable marine ecosystems in the south pacific ocean: implications for the use of broad-scale models in fisheries management. *Ocean Coastal Manag.* 120, 110–126. doi: 10.1016/j.ocecoaman.2015.11.025

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., and Clark, M. R. (2016b). Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around new zealand. *Deep-Sea Res. Part I-Oceanogr. Res. Papers* 115, 265–292. doi: 10.1016/j.dsr.2016.07.006

Anderson, O. F., Mikaloff Fletcher, S. E., and Bostock, H. C. (2015). *Development of Models for Predicting Future Distributions of Protected Coral Species in the New Zealand region*. NIWA Client Report to Department of Conservation No. WLG2015-65, 28.

Anderson, O. F., Tracey, D. M., Bostock, H. C., Williams, M., and Clark, M. R. (2014). *Refined Habitat Suitability Modelling for Protected Coral Species in the New Zealand EEZ*. No. National Institute of Water and Atmsopheric Research, Report No. WLG2014-69, 46.

Araujo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., et al. (2019). Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5:eaat4858. doi: 10.1126/sciadv.aat4858

Araujo, M. B., and Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688. doi: 10.1111/j.1365-2699.2006.01584.x

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, 321–331. doi: 10.1111/j.1600-0706.2012.00299.x

Baird, S., Tracey, D. M., Mormede, S., and Clark, M. (2013). *The Distribution of Protected Corals in New Zealand waters*. NIWA Client Report WLG2012-43 to the Department of Conservation, 96.

Baird, S. J., and Mules, R. (2019). *Extent of Bottom Contact by New Zealand Commercial Trawl Fishing for Deepwater Tier 1 and Tier 2 Target Species Determined Using CatchMapper Software, Fishing Years 2008–17*. New zealand Aquatic Environment and Biodiversity Report No. 229, 106.

Black, J., and Tilney, R. (2015). *Monitoring New Zealand's Trawl Footprint for Deepwater Fisheries: 1989–90 to 2010–11*. New Zealand Aquatic Environment and Biodiversity Report No. 142, 56.

Black, J., Wood, R., Berthelsen, T., and Tilney, R. (2013). *Monitoring New Zealand's Trawl Footprint for Deepwater Fisheries: 1989–90 to 2009–10*. New Zealand Aquatic Environment and Biodiversity Report No. 110, 57.

Bostock, H., Jenkins, C., Mackay, K., Carter, L., Nodder, S., Orpin, A., et al. (2019). Distribution of surficial sediments in the ocean around new zealand/aotearoa. Part B: continental shelf. *New Zealand J. Geol. Geophys.* 62, 24–45. doi: 10.1080/00288306.2018.1523199

Bowden, D. A. (2011). *Benthic Invertebrate Samples and Data From the Ocean Survey 20/20 Voyages to Chatham Rise and Challenger Plateau, 2007*. New zealand Aquatic Environment and Biodiversity Report No. 65, 46.

Bowden, D. A., Anderson, O., Escobar-Flores, P., Rowden, A., and Clark, M. (2019a). *Quantifying Benthic Biodiversity: Using Seafloor Image Data to Build Single-Taxon and Community Distribution Models for Chatham Rise, New Zealand*. New Zealand Aquatic Environment and Biodiversity Report No. 235, 67.

Bowden, D. A., Davey, N., Fenwick, M., George, S., Macpherson, D., Ray, C., et al. (2017). *Quantifying Benthic Biodiversity: A Factual Voyage Report From RV Tangaroa Voyage TAN1701 to Chatham Rise, 4 January – 2 February 2017*. New zealand Aquatic Environment and Biodiversity Report No. 185, 194.

Bowden, D. A., and Jones, D. O. B. (2016). "Towed Cameras," in *Biological Sampling in The Deep Sea*, eds M. R. Clark, A. A. Rowden, and M. Consalvey, (Hoboken, NJ: Wiley & Sons), 260–284. doi: 10.1002/9781118332535.ch12

Bowden, D. A., and Leduc, D. (2017). *Ocean Survey 20/20, Chatham Rise Benthos: effects of seabed trawling on benthic communities*. New zealand Aquatic Environment and Biodiversity Report No. 183, 67.

Bowden, D. A., Rowden, A. A., Anderson, O. F., Clark, M. R., Hart, A., Davey, N., et al. (2019b). *Quantifying Benthic Biodiversity: Developing a Dataset of Benthic Invertebrate Faunal Distributions From Seabed Photographic Surveys of*

Chatham Rise. New Zealand Aquatic Environment and Biodiversity Report No. 221, 35.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recog.* 30, 1145–1159. doi: 10.1016/s0031-3203(96)00142-2

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Brodie, S., and Clark, M. R. (2003). "Seamount management strategy - steps towards conserving offshore marine habitat," in *Aquatic Protected Areas: what works best and how do we know? Proceedings of the World Congress on Aquatic Protected Areas*, eds J. P. Beumer, A. Grant, and D. C. Smith, (Cairns, Australia: Australian Society of Fish Biology), 664–673.

Clark, M. R., Anderson, O. F., Francis, R., and Tracey, D. M. (2000). The effects of commercial exploitation on orange roughy (Hoplostethus atlanticus) from the continental slope of the chatham rise, new zealand, from 1979 to 1997. *Fisheries Res.* 45, 217–238. doi: 10.1016/s0165-7836(99)00121-6

Clark, M. R., Bowden, D. A., Rowden, A. A., and Stewart, R. (2019). Little evidence of benthic community resilience to bottom trawling on seamounts after 15 years. *Front. Marine Sci.* 6:63. doi: 10.3389/fmars.2019.00063

Clark, M. R., Consalvey, M., and Rowden, A. (2016). *Biological Sampling in The Deep Sea*. Hoboken, NJ: Wiley & Sons, 481.

Clark, M. R., and Dunn, M. R. (2012). Spatial management of deep-sea seamount fisheries: balancing sustainable exploitation and habitat conservation. *Environ. Conserv.* 39, 204–214. doi: 10.1017/s0376892912000021

Compton, T. J., Bowden, D. A., Pitcher, R. C., Hewitt, J. E., and Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off new zealand. *J. Biogeogr.* 40, 75–89. doi: 10.1111/j.1365-2699.2012.02761.x

Connell, J. H., and Slatyer, R. O. (1977). Mechanisms of succession in natural communities and their role in community stability and organisation. *Am. Nat.* 111, 1119–1144. doi: 10.1086/283241

Connolly, S. R., and Roughgarden, J. (1999). Theory of marine communities: competition, predation, and recruitment-dependent interaction strength. *Ecol. Monogr.* 69, 277–296. doi: 10.1890/0012-9615(1999)069[0277:tomccp]2.0.co;2

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

Davies, A. J., and Guinotte, J. M. (2011). Global habitat suitability for framework-forming cold-water corals. *PLoS One* 6:e18483. doi: 10.1371/journal.pone.0018483

Dayton, P. K., and Hessler, R. R. (1972). Role of biological disturbance in maintaining diversity in the deep sea. *Deep-Sea Res.* 19, 199–204. doi: 10.1016/0011-7471(72)90031-9

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251. doi: 10.1890/0012-9658(2007)88[243:btfema]2.0.co;2

Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., et al. (2012). Climate change impacts on marine ecosystems. *Ann. Rev. Marine Sci.* 4, 11–37.

Elith, J., and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Ann. Rev. Ecol. Evol. Syst.* 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159

FAO. (2009). *International guidelines for the management of deep-sea fisheries in the high seas*. Rome: Food and Agriculture Organization, 73.

Fielding, A. H., and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49. doi: 10.1017/s0376892997000088

Fisheries New Zealand. (2020). *Annual review report for deepwater fisheries 2018/19. The Deepwater Team, Fisheries Management, Fisheries New Zealand*. Wellington, New Zealand: Fisheries New Zealand, 113.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals Stat.* 28, 337–374. doi: 10.1214/aos/1016218223

Georgian, S. E., Anderson, O. F., and Rowden, A. A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the south pacific ocean. *Fisheries Res.* 211, 256–274. doi: 10.1016/j.fishres.2018.11.020

Glover, T. A., and Vaughn, S. (2010). *Response to intervention; evaluating current science and practice*. New York: The Guilford Press, 322.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., et al. (2013). Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435.

Guisan, A., and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186. doi: 10.1016/s0304-3800(00)00354-9

Hanley, J. A., and Mcneil, B. J. (1982). The meaning and use of the area under the reveiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747

Helson, J., Leslie, S., Clement, G., Wells, R., and Wood, R. (2010). Private rights, public benefits: industry-driven seabed protection. *Marine Policy* 34, 557–566. doi: 10.1016/j.marpol.2009.11.002

Hill, P. (2009). Designing a deep-towed camera vehicle using single conductor cable. *Sea Technol.* 50, 49–51.

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

Huetten, E., and Greinert, J. (2008). Software controlled guidance, recording and post-processing of seafloor observations by ROV and other towed devices: the software package OFOP. *Geophysical Res. Abst.* 10.

Kaiser, M. J., Hilborn, R., Jennings, S., Amaroso, R., Andersen, M., Balliet, K., et al. (2016). Prioritization of knowledge-needs to achieve best practices for bottom trawling in relation to seabed habitats. *Fish Fisheries* 17, 637–663.

Kenchington, E., Callery, O., Davidson, F., Grehan, A., Morato, T., Appiott, J., et al. (2019). *Use of Species Distribution Modeling in the Deep Sea*. Canada: Canadian Technical Report of Fisheries and Aquatic Sciences, 76.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., and Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding new zealand: an analysis using boosted regression trees. *Marine Ecol. Prog. Series* 321, 267–281. doi: 10.3354/meps321267

Lobo, J. M., Jiménez-Valverde, A., and Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17, 145–151. doi: 10.1111/j.1466-8238.2007.00358.x

Marchal, P., Francis, C., Lallemand, P., Lehuta, S., Mahevas, S., Stokes, K., et al. (2009). Catch-quota balancing in mixed-fisheries: a bio-economic modelling approach applied to the new zealand hoki (macruronus novaezelandiae) fishery. *Aquatic Living Res.* 22, 483–498. doi: 10.1051/alr/2009033

Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., et al. (2018). The nippon foundation-GEBCO seabed 2030 project: the quest to see the world's oceans completely mapped by 2030. *Geosciences* 8:63. doi: 10.3390/geosciences8020063

McClatchie, S., Millar, R. B., Webster, F., Lester, P. J., Hurst, R., and Bagley, N. (1997). Demersal fish community diversity off New Zealand: Is it related to depth, latitude and regional surface phytoplankton? *Deep-Sea Res. Part I-Oceanogr. Res. Papers* 44, 647–667. doi: 10.1016/s0967-0637(96)00096-9

Moilanen, A., Runge, M. C., Elith, J., Tyre, A., Carmel, Y., Fegraus, E., et al. (2006). Planning for robust reserve networks using uncertainty analysis. *Ecol. Model.* 199, 115–124. doi: 10.1016/j.ecolmodel.2006.07.004

Mountjoy, J. J., Howarth, J. D., Orpin, A. R., Barnes, P. M., Bowden, D. A., Rowden, A. A., et al. (2018). Earthquakes drive large-scale submarine canyon development and sediment supply to deep-ocean basins. *Sci. Adv.* 4:eaar3748. doi: 10.1126/sciadv.aar3748

Nodder, S. D., Bowden, D. A., Pallentin, A., and Mackay, K. A. (2012). "Seafloor habitats and benthos of a continental ridge: chatham rise, New Zealand," in *Seafloor Geomorphology as Benthic Habitat*, ed. P. T. Harris (Amsterdam: Elsevier Inc.), 763–776.

O'Driscoll, R. L., MacGibbon, D., Fu, D., Lyon, W., and Stevens, D. (2011). *A Review of Hoki and Middle-Depths Trawl Surveys of the Chatham Rise, January 1992–2010*. New Zealand Fisheries Assessment Report 2011/47, No. 72.

O'toole, R., Judge, M., Sacchetti, F., Furey, T., Mac Craith, E., Sheehan, K., et al. (2020). Mapping Ireland's coastal, shelf and deep-water environments using illustrative case studies to highlight the impact of seabed mapping on the generation of blue knowledge. *Geol. Soc. London Spl. Public.* 505:207.

Parmentier, I., Harrigan, R. J., Buermann, W., Mitchard, E. T. A., Saatchi, S., Malhi, Y., et al. (2011). Predicting alpha diversity of African rain forests: models based

on climate and satellite-derived data do not perform better than a purely spatial model. *J. Biogeogr.* 38, 1164–1176. doi: 10.1111/j.1365-2699.2010.02467.x

Pearce, J., and Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* 133, 225–245. doi: 10.1016/s0304-3800(00)00322-7

Pearce, J., and Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biol. Conserv.* 98, 33–43. doi: 10.1016/S0006-3207(00)00139-7

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Pitcher, C. R., Ellis, N., Jennings, S., Hiddink, J. G., Mazor, T., Kaiser, M. J., et al. (2017). Estimating the sustainability of towed fishing-gear impacts on seabed habitats: a simple quantitative risk assessment method applicable to data-limited fisheries. *Methods Ecol. Evol.* 8, 472–480. doi: 10.1111/2041-210x.12705

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., et al. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* 11:4540.

Regan, H. M., Ben-Haim, Y., Langford, B., Wilson, W. G., Lundberg, P., Andelman, S. J., et al. (2005). Robust decision-making under severe uncertainty for conservation management. *Ecol. Appl.* 15, 1471–1477. doi: 10.1890/03-5419

Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., et al. (2015). Benthos distribution modelling and its relevance for marine ecosystem management. *Ices J. Marine Sci.* 72, 297–315. doi: 10.1093/icesjms/fsu107

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., and Lundquist, C. J. (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Front. Marine Sci.* 4:421. doi: 10.3389/fmars.2017.00421

Rowden, A., Leduc, D., Torres, L., Bowden, D., Hart, A., Chin, C., et al. (2014). *Benthic Epifauna Communities of the Central Chatham Rise crest.* NIWA Client Report to Chatham Rock Phosphate Ltd. No. WLG2014-9, 116.

Rowden, A. A., Stephenson, F., Clark, M. R., Anderson, O. F., Guinotte, J. M., Baird, S. J., et al. (2019). Examining the utility of a decision-support tool to develop spatial management options for the protection of vulnerable marine ecosystems on the high seas around new zealand. *Ocean Coastal Manag.* 170, 1–16. doi: 10.1016/j.ocecoaman.2018.12.033

Smith, K. L. Jr., Ruhl, H. A., Bett, B. J., Billett, D. S. M., Lampitt, R. S., and Kaufmann, R. S. (2009). Climate, carbon cycling, and deep-ocean ecosystems. *Proc. Natl. Acad. Sci. U S A.* 106, 19211–19218. doi: 10.1073/pnas.0908322106

Somodi, I., Lepesi, N., and Botta-Dukat, Z. (2017). Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecol. Evol.* 7, 863–872. doi: 10.1002/ece3.2654

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293. doi: 10.1126/science.3287615

Tracey, D. M., Rowden, A. A., Mackay, K. A., and Compton, T. (2011). Habitat-forming cold-water corals show affinity for seamounts in the new zealand region. *Marine Ecol. Prog. Series* 430, 1–59. doi: 10.3354/meps09164

Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* 10, 225–232. doi: 10.1111/2041-210x.13107

Verbyla, D. L., and Litvaitis, J. A. (1989). Resampling methods for evaluating clssification accuracy of wildlife habitat models. *Environ. Manag.* 13, 783–787. doi: 10.1007/bf01868317

Vierod, A. D. T., Guinotte, J. M., and Davies, A. J. (2014). Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. *Deep-Sea Res. Part ii Topical Studies Oceanogr.* 99, 6–18. doi: 10.1016/j.dsr2.2013.06.010

Von Rad, U., and Kudrass, H.-R. (1987). "Exploration and genesis of submarine phosphorite deposits from the Chatham Rise, New Zealand - a review," in *Marine minerals advance in research and resource assessement*, ed. P. Telekei, (Dordrecht: Reidel), 157–175. doi: 10.1007/978-94-009-3803-8_12

Warton, D. I., Blanchet, F. G., O'hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., et al. (2015). So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766–779. doi: 10.1016/j.tree.2015.09.007

Wiley, E. O., Mcnyset, K. M., Peterson, A. T., Robins, C. R., and Stewart, A. M. (2003). Niche modeling perspective on geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* 16, 120–127. doi: 10.5670/oceanog.2003.42

Williams, A., Althaus, F., Maguire, K., Green, M., Untiedt, C., Alderslade, P., et al. (2020). The fate of deep-sea coral reefs on seamounts in a fishery-seascape: what are the impacts, what remains, and what is protected? *Front. Marine Sci.* 7:567002. doi: 10.3389/fmars.2020.567002

Williams, A., Schlacher, T. A., Rowden, A. A., Althaus, F., Clark, M. R., Bowden, D. A., et al. (2010). Seamount megabenthic assemblages fail to recover from trawling impacts. *Marine Ecol.* 31, 183–199. doi: 10.1111/j.1439-0485.2010.00385.x

Winship, A. J., Thorson, J. T., Clarke, M. E., Coleman, H. M., Costa, B., Georgian, S. E., et al. (2020). Good practices for species distribution modeling of deep-sea corals and sponges for resource management: data collection, analysis, validation, and communication. *Front. Marine Sci.* 7:303. doi: 10.3389/fmars.2020.00303

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 21 April 2021
doi: 10.3389/fmars.2021.629230

Resource Occurrence and Productivity in Existing and Proposed Wind Energy Lease Areas on the Northeast US Shelf

Kevin D. Friedland[1], Elizabeth T. Methratta[2], Andrew B. Gill[3], Sarah K. Gaichas[4], Tobey H. Curtis[5], Evan M. Adams[6], Janelle L. Morano[7], Daniel P. Crear[8], M. Conor McManus[9] and Damian C. Brady[10]*

[1] National Marine Fisheries Service, Narragansett, RI, United States, [2] IBSS Corporation, in Support of NOAA National Marine Fisheries Service, Woods Hole, MA, United States, [3] Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, United Kingdom, [4] National Marine Fisheries Service, Woods Hole, MA, United States, [5] Atlantic Highly Migratory Species Management Division, National Marine Fisheries Service, Gloucester, MA, United States, [6] Biodiversity Research Institute, Portland, ME, United States, [7] Department of Natural Resources and the Environment, Cornell University, Ithaca, NY, United States, [8] ECS Federal, in Support of Atlantic Highly Migratory Species Management Division, National Marine Fisheries Service, Silver Spring, MD, United States, [9] Division of Marine Fisheries, Rhode Island Department of Environmental Management, Jamestown, RI, United States, [10] School of Marine Sciences, University of Maine, Walpole, ME, United States

OPEN ACCESS

Edited by:
Mary C. Fabrizio, College of William & Mary, United States

Reviewed by:
Ana B. Bugnot, The University of Sydney, Australia
Robert Isdell, College of William & Mary, United States

***Correspondence:**
Kevin D. Friedland
kevin.friedland@noaa.gov

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: *13 November 2020*
Accepted: *16 March 2021*
Published: *21 April 2021*

Citation:
Friedland KD, Methratta ET, Gill AB, Gaichas SK, Curtis TH, Adams EM, Morano JL, Crear DP, McManus MC and Brady DC (2021) Resource Occurrence and Productivity in Existing and Proposed Wind Energy Lease Areas on the Northeast US Shelf. Front. Mar. Sci. 8:629230. doi: 10.3389/fmars.2021.629230

States in the Northeast United States have the ambitious goal of producing more than 22 GW of offshore wind energy in the coming decades. The infrastructure associated with offshore wind energy development is expected to modify marine habitats and potentially alter the ecosystem services. Species distribution models were constructed for a group of fish and macroinvertebrate taxa resident in the Northeast US Continental Shelf marine ecosystem. These models were analyzed to provide baseline context for impact assessment of lease areas in the Middle Atlantic Bight designated for renewable wind energy installations. Using random forest machine learning, models based on occurrence and biomass were constructed for 93 species providing seasonal depictions of their habitat distributions. We developed a scoring index to characterize lease area habitat use for each species. Subsequently, groups of species were identified that reflect varying levels of lease area habitat use ranging across high, moderate, low, and no reliance on the lease area habitats. Among the species with high to moderate reliance were black sea bass (*Centropristis striata*), summer flounder (*Paralichthys dentatus*), and Atlantic menhaden (*Brevoortia tyrannus*), which are important fisheries species in the region. Potential for impact was characterized by the number of species with habitat dependencies associated with lease areas and these varied with a number of continuous gradients. Habitats that support high biomass were distributed more to the northeast, while high occupancy habitats appeared to be further from the coast. There was no obvious effect of the size of the lease area on the importance of associated habitats. Model results indicated that physical drivers and lower trophic level indicators might strongly control the habitat distribution of ecologically

and commercially important species in the wind lease areas. Therefore, physical and biological oceanography on the continental shelf proximate to wind energy infrastructure development should be monitored for changes in water column structure and the productivity of phytoplankton and zooplankton and the effects of these changes on the trophic system.

Keywords: wind energy, fisheries, habitat, monitoring, temperature, zooplankton

INTRODUCTION

The provision of ecosystem goods and services is intersecting with the rapid development of the offshore wind industry in continental shelf seas, an effort in energy generation designed in part to ameliorate the effects of climate change (Causon and Gill, 2018; UNEP, 2019). Shelf seas account for the majority of seafood production of the World Ocean (Costanza et al., 2014), which raises particular concern about the impacts that energy infrastructure will have on fisheries and dependent communities (Hooper et al., 2015; Carpenter, 2020). Therefore, we have a growing need to develop assessment methods to understand the intersection of fisheries resources with potential offshore wind energy areas.

Interactions between wind energy production and fisheries resources occur across trophic levels and life stages, during each phase of energy infrastructure development, and through biotic and abiotic pathways (Boehlert and Gill, 2010; Pezy et al., 2020). The exploration and construction phases of offshore wind development bring periodic elevated noise levels to the marine environment through increased vessel traffic, seismic survey methods, and, in most cases of fixed foundation turbines, impulse pile driving (Wahlberg and Westerberg, 2005; Hatch et al., 2008). Acoustic changes to the marine environment can cause sublethal physiological effects (Popper and Hawkins, 2019) and mortality, as well as changes in movement, behavior, habitat utilization, and migration patterns for numerous marine taxa (Boehlert and Gill, 2010; Brandt et al., 2018). Once installed and operating, wind turbine foundations create new hard bottom habitats that enhance the recruitment of native and non-native benthic invertebrates (De Mesel et al., 2015). The resulting artificial reefs attract fish species seeking food and refuge (Wilhelmsson et al., 2006). Shells and other biogenic materials associated with reef organisms are deposited in the surrounding environment, increasing sediment organic content and nutrient concentrations, and thus modifying benthic community composition (Wilding, 2014). Furthermore, the network of power cables associated with wind farms emits electromagnetic fields, which have the potential to affect behavior and movement of commercially valued and migratory species (Hutchison et al., 2020).

Localized hydrodynamic regime changes at the scale of individual turbines and wind farms occur as currents pass structures, modifying downstream turbulence, surface wave energy, and upwelling patterns (Bakhoday-Paskyabi et al., 2018). Much larger scale effects (∼80 km from structures) on hydrodynamics and vertical stratification are possible through the impact of wind wakes and dynamic coupling of the ocean and atmospheric systems (Carpenter et al., 2016). Physical and biological oceanographic processes are directly linked through numerous mechanisms, including the vertical and horizontal transport of macro- and micro- nutrients to primary producers, and changes in the distribution of suspended particulates, and the effect of this suspended matter on the depth of the photic zone. Altered hydrodynamic patterns could affect primary production as well as upper trophic levels. These conceptual linkages have been demonstrated with empirical data in the southern North Sea that revealed increased vertical mixing at an offshore wind farm resulting in the transport of nutrients to the surface mixed layer and subsequent uptake by phytoplankton in the photic zone (Floeter et al., 2017). Changes in water column properties (water temperature, dissolved oxygen, and suspended matter concentration) have also been linked to altered zooplankton community structure at offshore wind farms in China (Wang et al., 2018). Increased primary production could have important implications for the productivity of bivalves and other macro-benthic suspension feeders, representing a major component of artificial reef communities that form on turbine foundations (Slavik et al., 2019; Mavraki et al., 2020). In total, these effects may propagate to upper trophic levels, particularly predatory fish on and around the turbines (Pezy et al., 2020).

Assessing the effects of offshore wind on fisheries resources requires that we know what to measure, what survey designs to use, and how to coordinate that information with existing surveys that support regional stock assessments (Wilding, 2014; Methratta, 2020). Our traditional view of how fish habitat is defined is rapidly changing; there is expanding evidence suggesting that fish habitats can be determined by biological variables related to primary and secondary production patterns (Weber et al., 2018; Mazur et al., 2020). This highlights the need to continue, or expand, current sampling efforts related to the water column parameters such as phytoplankton and zooplankton and suspended sediment material.

Currently in the United States, the offshore wind energy developers are required by the Federal Agency responsible for permitting and management of the offshore waters [the Bureau of Ocean Management (BOEM)], to consider essential fish habitat (EFH) designations developed by the National Marine Fisheries Service (NMFS) for fishery management plans to help assess species and habitat impacts. While EFH has benefits in its availability and mostly standardized development across species, it is a comparatively coarse representation of a given species' distribution and habitat reliance, and can be based on sparse or discontinuously collected observations (Moore et al., 2016). EFH may encompass the broad range of a species' distribution and its habitats, but does not generally discern if there are focal or highly preferred habitat areas within that

distribution, or if there are seasonal shifts in distribution. Species distribution models (SDMs) can complement EFH by providing higher resolution distribution probabilities within EFH and identify influential environmental parameters and habitat suitability. Various types of SDMs have been developed using fishery-dependent or -independent data to both hindcast and forecast species distributions and habitat suitability relative to spatial fisheries management needs (Hazen et al., 2018; Crear et al., 2020).

Offshore wind energy capacity is developing rapidly across the Northeast US Continental Shelf ecosystem (NES), with most projects currently in a pre-construction planning and assessment phase. There are considerable data gaps and a need for additional baseline information to support the assessment of these impacts on marine resources in this highly productive region. The goal of this study was to characterize the use of wind energy lease areas by fish and macroinvertebrate species sampled in resource abundance surveys on the NES and identify species with a high dependency on lease area habitats. These species might be considered for prioritized attention and research by fisheries management. The characterization was based on habitat or SDMs developed using machine learning techniques. These models provided species occurrence probability and biomass productivity at spatial scales related to BOEM lease areas. The relative importance of habitats was characterized for use by each species, including the change in dependency on these habitats over time, and the biological and physical aspects of the ecosystem that shaped habitat. The models in this study draw on a range of lower trophic level variables to provide context for identifying ecosystem properties that would merit monitoring before, during, and after installing energy-generating structures.

MATERIALS AND METHODS

Study System

We studied the distribution of fish and macroinvertebrates occurring in the NES, a well-studied marine system along the western boundary of the North Atlantic Ocean. We fit SDMs estimating occupancy and biomass habitats onto a 0.1-degree grid, termed the estimation grid (**Figure 1A**). The boundaries of eleven wind energy lease areas are identified as either existing lease areas, E1–E7, or proposed areas, P1–P4. Each lease area will be composed of a number of parcels of varying size (boundaries as of May 2020). A single convex hull was drawn around the parcels of each lease area and was the basis for spatial data extraction representing that area; hence, this was exclusive of any cable corridors. Since depth distribution is an important factor, bathymetric relief and key depth contours are shown in **Figure 1B**, along with identifying areas within the NES commonly referred to in the text.

Survey Data Response Variables

The basis of the study was a series of SDMs incorporating habitat features for taxa captured in a fishery-independent bottom trawl survey conducted in the NES. The bottom trawl survey has been conducted by the Northeast Fisheries Science Center each year since 1963 in the fall and spring since 1968, occupying

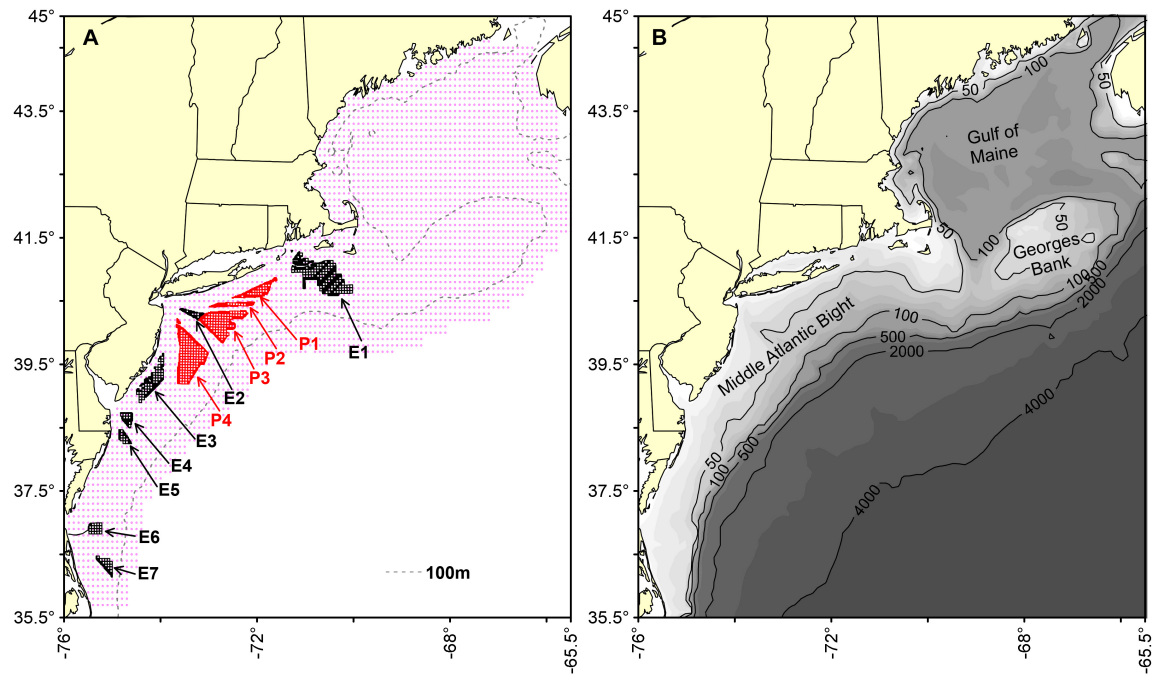

FIGURE 1 | Map of the study system showing the estimation grid (red dots) for habitat models **(A)** and bathymetry **(B)**. Existing (black outlines) and proposed (red outlines) energy lease sites on the Northeast US Shelf are labels E1 through E7 and P1 through P4, respectively. Bathymetric relief, key depth contours, areas of significance within the NES are also presented.

upward of 300 stations during each season and is based on a random stratified design (Desprespatanjo et al., 1988). Catches were standardized for various correction factors related to vessels and gears used in the time series (Miller et al., 2010). The survey catch provided the binary response of presence or absence for each taxa as the response variable in classification models, and catch per unit ($\log_{10}$[CPUE kg tow^{-1} + 1]) of biomass used as continuous variables in the regression models that can be thought of as a biomass habitat metric (see Data Availability Statement). Though time series of catch data extend back to the 1960s, the time series of the analysis was limited to the period 1976–2018 due the availability of other data described below.

Predictor Variables

Physical and biological environmental data used as predictor variables included dynamic variables that changed annually with recurring sampling and static variables that were held constant over time. The suite of predictors can be summarized over five general categories (**Table 1**). Physical environmental variables, including surface and bottom water temperature and salinity were collected contemporaneously with survey trawl samples with Conductivity/Temperature/Depth (CTD) instruments (see Data Availability Statement). Temperature and salinity were initially tested as dynamics variables; however, salinity was found to be a weak predictor and was applied as a static variable, which enabled training and fitting the models over the period 1976–2018. Depth of the survey station (meters) was a static variable in the analysis.

In addition to the dynamic station temperature variables, remote sensing sea surface temperature (SST) fields were used to derive a complementary set of static physical environment variables. SST fields from the MODIS Terra sensor were used to generate monthly mean SST data and monthly gradient magnitude, or frontal fields of the SST (see Data Availability Statement). There are many methods used to identify fronts (Belkin and O'Reilly, 2009) in oceanographic data that usually utilize some focal filter to reduce noise and then identify gradient magnitude with a Sobel filter. Calculations were performed in R using the "raster" package (version 2.6-7) by applying a three by three mean focal filter and a Sobel filter to generate x and y derivatives, which were then used to calculate gradient magnitude.

Benthic terrain descriptors included a series of static variables that characterize the shape and complexity of the substrate. Most benthic terrain variables were derived from the depth measurements, such as vector ruggedness, rugosity, and slope (**Table 2**). Other variables described the substrate itself, such as benthic sediment grain size. The vorticity of benthic currents was also considered a benthic terrain variable. These variables were sampled to match the position of a survey trawl.

Biological covariates included predictor variables representing lower trophic level primary and secondary production. Primary production variables were monthly chlorophyll concentration static variables developed from a multi-sensor remote sensing data source. These data were merged using the Garver, Siegel, Maritorena Model algorithm (Maritorena et al., 2010) obtained from the Hermes GlobColour website and provided data over the period of 1997–2016 (see Data Availability Statement). As with the remote sensing SST data, monthly gradient magnitude (i.e., chlorophyll frontal fields) were also developed. Both the SST and chlorophyll concentration variables were sampled to match the position of a survey trawl.

Secondary production variables were based on zooplankton abundances measured by the Ecosystem Monitoring Program (EcoMon), which conducts shelf-wide bimonthly random-stratified surveys of the NES (Kane, 2007). Zooplankton are collected obliquely through the water column to a maximum depth of 200 m using paired 61-cm Bongo samplers equipped with 333-micron mesh nets (see Data Availability Statement). We used the density estimates (number per 100 m^3) of the 18 most abundant taxonomic categories and a biomass indicator (settled bio-volume) as potential predictor variables (**Table 3**). The zooplankton time series has some missing values, which were ameliorated by summing data over 5-year time steps for each seasonal period and interpolating a complete field using ordinary kriging. For example, zooplankton data for spring 2000 would include the available data from tows made during the period 1998–2002. The zooplankton variables were sampled to match the date (season) and position of the survey trawl.

TABLE 1 | Summary of predictor variables used in the development of spring and fall presence/absence and biomass habitat models.

Predictor variable categories	Description	Number
Physical environment variables	Physical and oceanographic variables including depth (DEPTH), surface and bottom temperature (ST_SD, BT_SD), and surface and bottom salinity (SS_SD, BS_SD) derived from surveys.	5
Benthic terrain descriptors	A series of variables that characterize the structure of benthic habitats, most of which are based on bathymetry data. See **Table 2** for details.	19
Secondary production variables	Abundance of zooplankton taxa and a zooplankton biomass index (settled bio-volume) composed mostly of copepod species. Some taxa only identified to family or others to a general category. See **Table 3** for details.	19
Remote sensing Primary production variables	Remote sensed measurements of monthly mean chlorophyll concentration; and, the gradient magnitude or frontal data for the same fields [CHL_(R for raw data, F for frontal gradient magnitude)_(XX from 01 to 12 for month)].	24
Remote sensing Physical environment variables	Remote sensed measurements of monthly mean SST; and, the gradient magnitude or frontal data for the same fields [SST_(R for raw data, F for frontal gradient magnitude)_(XX from 01 to 12 for month)].	24

Number refers to number of variables.

TABLE 2 | Summary of benthic terrain predictor variables used in the development of spring and fall presence/absence and biomass habitat models.

Variable	Notes	References
COMPLEXITY	Terrain Ruggedness Index, the difference in elevation values from a center cell and the eight cells immediately surrounding it. Each of the difference values are squared to make them all positive and averaged. The index is the square root of this average.	Riley et al., 1999
NAMERA_BPI	BPI is a second order derivative of the surface depth using the TNC Northwest Atlantic Marine Ecoregional Assessment ("NAMERA") data with an inner radius = 5 and outer radius = 50.	Lundblad et al., 2006
NAMERA_VRM	Vector Ruggedness Measure (VRM) measures terrain ruggedness as the variation in three-dimensional orientation of grid cells within a neighborhood based the TNC Northwest Atlantic Marine Ecoregional Assessment ("NAMERA") data.	Hobson, 1972
PRCURV2KM, PRCURV10KM PRCURV 20KM	Benthic profile curvature at 2, 10, and 20 km spatial scales was derived from depth data.	Winship et al., 2018
RUGOSITY	A measure of small-scale variations of amplitude in the height of a surface, the ratio of the real to the geometric surface area.	Friedman et al., 2012
SEABEDFORMS	Seabed topography as measured by a combination of seabed position and slope.	http://www.northeastoceandata.org/
SLP2KM SLP10KM SLP20KM	Benthic slope at 2, 10, and 20 km spatial scales.	Winship et al., 2018
SLPSLP2KM SLPSLP10KM SLPSLP20KM	Benthic slope of slope at 2, 10, and 20 km spatial scales	Winship et al., 2018
SOFT_SED	Soft-sediments is based on grain size distribution from the USGS usSeabed: Atlantic coast offshore surficial sediment data.	http://www.northeastoceandata.org/
VORTFA VORTSP VORTSU VORTWI	Benthic current vorticity at a 1/6 degree (approx. 19 km) spatial scale in fall (fa), spring (sp), summer (su), and winter (wi).	Kinlan et al., 2016

TABLE 3 | Summary of zooplankton predictor variables used in the development of spring and fall presence/absence and biomass habitat models.

Variable name	Full name
ACARSP	*Acartia spp.*
CALFIN	*Calanus finmarchicus*
CHAETO	Chaetognatha
CHAMZZ	*Centropages hamatus*
CIRRZZ	Cirripedia
CTYPZZ	*Centropages typicus*
ECHINO	Echinodermata
EVADNE	*Evadne spp.*
GASZZZ	Gastropoda
HYPERZ	Hyperiidea
LARVAC	Appendicularians
MLUCEN	*Metridia lucens*
OITHSP	*Oithona spp.*
PARAZZ	*Paracalanus parvus*
PENILE	*Penilia* spp.
PSEUDO	*Pseudocalanus spp.*
SALPSZ	Salpa
TLONGZ	*Temora longicornis*
VOLUME	Plankton bio-volume

Occupancy and Biomass Habitat Models

Seasonal SDMs were developed using the approach as reported in Friedland et al. (2020); however, salinity variables were static fields as opposed to dynamic. Random Forest models based on occurrence were fit as classification models of the presence or absence of taxa in a trawl tow and yielded an occurrence probability; hereafter, these models are referred to as presence/absence models and the estimates of habitat from these models is referred to as occupancy habitat. Random Forest models based on biomass were fit as regression models of the catch rate of taxa in weight and yielded an index of biomass; hereafter, these models are referred to as biomass models and their output as biomass habitat. Random forest machine learning models were fit (Cutler et al., 2007) using the "randomForest" R package (version 4.6-14). Random forest models can achieve comparable predictive power to other statistical methods (Smolinski and Radtke, 2017). Prior to fitting the model, the independent variables were tested for multi-collinearity among the predictors using the multi-collinear command from R package "rfUtilities" (version 2.1-5) with a p-level of 0.1; highly correlated variables were eliminated from the analysis. From this reduced set of predictors, the final model variables were selected utilizing the model selection criteria of Murphy et al. (2010) as implemented in rfUtilities. This procedure ranks the importance of model variables based on the change in mean squared error as a ratio of the maximum model improvement error (termed MIR). A range of models are fit and all variables with MIRs above a given threshold are retained; the threshold is selected to minimize the number of variables in the model while minimizing the mean squared error and maximizing the variation explained (see **Supplementary Material** for example R code). The presence/absence models were evaluated for fit based on out-of-bag classification accuracy using the AUC or Area Under the ROC Curve index using the "irr"

package in R (version 0.84.1), applying an optimized classification threshold probability. Models with an AUC of at least 0.73 were deemed satisfactory, a level associated with the lowest performing model that was included in the study. Biomass model regressions were evaluated for fit using the root mean squared log error statistic based on the R package "Metrics" (version 0.1.4). The 96 candidate species selected were consistently abundant taxa from the survey, occurring in at least 150 trawl tows; separate seasonal (spring and fall/autumn) models were fit reflecting the two seasonal surveys. From this candidate list, a subgroup of species with satisfactory presence/absence model fits were used to estimate occupancy and biomass habitats over the estimation grid for the same period of the training data, 1976–2018.

Analysis Strategy

The study had four goals: (1) identify the species with habitats overlapping the wind energy lease areas; (2) characterize the relative importance of lease areas to species modeled in the study; (3) characterize the change over time in habitat value to species in the lease areas; and (4) determine which aspects of the ecosystem were critical in shaping habitat in the lease areas.

To achieve the first goal, an index representing the reliance of a species on lease area habitats was developed that utilized both seasonal habitat scores (limited to the occurrence probabilities only) for species and a relative score based on the ratio of habitat scores within a lease area to overall habitat score for the NES. A habitat score for a lease area was the median of occurrence probability for a species that fell within a convex hull that circumscribed a lease area, while the habitat score for the NES was the median of occurrence probability over the entire NES. Within a season, the index was incremented by one if the median habitat score across the lease areas averaged >0.1, by one if the average habitat score plus the 95% confidence interval was >0.1, and by one if the habitat scores in at least one lease area was >0.1. The 0.1 threshold was selected to represent those species with a clear presence in an area. Similarly, the index was incremented by one if the median ratio across the lease areas averaged >1, by one if the average ratio plus the 95% confidence interval was >1, and by one if the ratio in at least one lease area was >1. For a species with a satisfactory model fit in a single season, the index could range from 0 to 6, and for a species with satisfactory models in both seasons, the index could range from 0 to 12. The species were divided by their index scores into four groups: "high reliance" for species with indices from 12 to 10; "moderate reliance" for species with indices from 9 to 6; "low reliance" for species with indices from 5 to 1; "no reliance" for species with indices of 0. High reliance species consistently showed high utilization of lease areas in both seasons, moderate reliance species generally showed high utilization in at least one season, low reliance species generally showed low utilization in both seasons, and no reliance species did not show utilization in either season.

For the second goal, species occupancy and biomass habitat were compared across lease areas and with respect to the geographical position and size of the lease areas. For each lease area, the number of species for which the ratio exceeded 0.7 was summed, with the same exercise repeated for biomass habitat. The threshold value of 0.7 was determined by testing values from 1 to 0.5; the 0.7 level represented a breakpoint where lower thresholds did not dramatically increase the number of species included. The numbers of spring and fall species were averaged to represent the relative roles of lease areas in respect to species occurrence and biomass. In addition, species counts by model and season were associated with four properties of the lease areas: latitude, longitude, distance to the coast, and area of the lease area. The correlations between species number and the four properties were tested with Spearman rank order correlation.

For the third goal, trend in habitat use was evaluated using a non-parametric test of time series (1976–2018) trend using the R package "zyp" (version 0.10-1.1). We used the Yue et al. (2002) method to estimate Theil-Sen slopes and performed an auto-correlation corrected Mann–Kendall test of trend. The trend in occupancy and biomass habitat was evaluated for the high reliance species grouping and plotted against median occurrence probability and biomass habitat scores.

For the fourth goal, variable importance in the presence/absence models were evaluated for the models associated with the species in the high reliance grouping. Importance was based on five performance measures: the number of times a variable was the root variable (i.e., variable associated with the root node); the mean minimum node depth for the variable; Gini index of node impurity decreases; prediction accuracy decrease; and, the proportion of models for which the variable was among the 10 highest ranked variables. The first four of these indices was computed using the "randomForestExplainer" R package (version 0.10.0); as customary, the times a root variable was plotted against the mean minimum depth variable and the Gini index was plotted against the accuracy decrease variable. All the performance measures were used in a principal components (PCs) analysis to provide an overall rank of variables across species based on PC 1.

RESULTS

Species Models Included in the Analysis

The species modeled in this study included both finfish and macroinvertebrate taxa. Of the initial candidate species, 93 taxa had a seasonal presence/absence model with an AUC score of at least 0.73 and were thus included in the study (**Table 4**). Based on the performance of the seasonal presence/absence model, complementary biomass model results were also considered. Not all species with a satisfactory model fit in one season (i.e., spring or fall) had a satisfactory fit in the other season. Hence, we had model results for 83 taxa in the spring and 89 in the fall, with 80 taxa having models in both seasons.

Identification of Species Associated With Wind Lease Areas

Using our index based on habitat scores and ratios of habitat scores, we identified four groupings of species that reflected the importance of the lease area habitats. Twenty species fell within the criteria for the high reliance grouping (**Table 5** and

TABLE 4 | Species with presence/absence and biomass habitat models based on spring and fall survey data.

		AUC		RMSLE				AUC		RMSLE	
Species	**Abbr**	**Spring**	**Fall**	**Spring**	**Fall**	**Species**	**Abbr**	**Spring**	**Fall**	**Spring**	**Fall**
Sebastes fasciatus	ACARED	0.93	0.94	0.15	0.14	*Lophius americanus*	MONKFH	0.76	0.77	0.16	0.18
Alosa pseudoharengus	ALEWIF	0.75	0.87	0.15	0.10	*Triglops murrayi*	MOUSCL	0.84	0.84	0.01	0.02
Aspidophoroides monopterygius	ALLFSH	0.85	0.85	0.00	0.00	*Prionotus carolinus*	NORSEA	0.85	0.85	0.11	0.11
Hippoglossoides platessoides	AMEPLA	0.91	0.94	0.10	0.10	*Sphoeroides maculatus*	NPUFFR	0.87	0.90	0.00	0.03
Alosa sapidissima	AMESHA		0.81		0.06	*Macrozoarces americanus*	OCPOUT	0.79	0.82	0.15	0.08
Homarus americanus	AMLOBS	0.79	0.77	0.15	0.17	*Merluccius albidus*	OFFHAK	0.91	0.93	0.05	0.04
Squatina dumeril	ANGSHR	0.93	0.91	0.05	0.08	*Pollachius virens*	POLLOC	0.82	0.86	0.19	0.17
Peristedion miniatum	ARMSEA	0.88	0.88	0.02	0.02	*Cancer irroratus*	RCKCRA	0.80	0.76	0.06	0.04
Argentina silus	ATLARG	0.88	0.84	0.04	0.02	*Decapterus punctatus*	RDSCAD		0.83		0.03
Gadus morhua	ATLCOD	0.84	0.89	0.22	0.18	*Geryon quinquedens*	REDCRA	0.84	0.84	0.03	0.03
Micropogonias undulatus	ATLCRO	0.89	0.95	0.04	0.15	*Urophycis chuss*	REDHAK	0.80	0.82	0.15	0.16
Hippoglossus hippoglossus	ATLHAL	0.84	0.84	0.07	0.07	*Etrumeus teres*	RHERRI		0.80		0.13
Clupea harengus	ATLHER	0.76	0.90	0.21	0.14	*Trachurus lathami*	ROSCAD		0.84		0.03
Scomber scombrus	ATLMAC	0.77	0.77	0.19	0.10	*Leucoraja garmani*	ROSSKA	0.93	0.93	0.04	0.04
Brevoortia tyrannus	ATLMEN	0.85	0.85	0.04	0.03	*Dasyatis centroura*	RTSTIG		0.83		0.14
Melanostigma atlanticum	ATLPOU	0.86	0.84	0.01	0.00	*Ammodytes dubius*	SANDLA	0.82	0.75	0.09	0.05
Menidia menidia	ATLSIL	0.89		0.01		*Stenotomus chrysops*	SCUPZZ	0.90	0.91	0.10	0.16
Anarhichas lupus	ATLWOL	0.85	0.86	0.09	0.06	*Hemitripterus americanus*	SEARAV	0.80	0.81	0.14	0.11
Dipturus laevis	BARSKA	0.89	0.88	0.09	0.11	*Placopecten magellanicus*	SEASCA	0.84	0.84	0.11	0.13
Anchoa mitchilli	BAYANC	0.94	0.92	0.04	0.11	*Lumpenus maculatus*	SHANNY	0.90		0.01	
Centropristis striata	BLABAS	0.86	0.86	0.07	0.07	*Chlorophthalmus agassizi*	SHORTP	0.90	0.93	0.01	0.01
Helicolenus dactylopterus	BLAROS	0.90	0.90	0.06	0.06	*Illex illecebrosus*	SHTSQD	0.88	0.81	0.06	0.16
Callinectes sapidus	BLUCRA	0.74	0.89	0.00	0.02	*Merluccius bilinearis*	SILHAK	0.81	0.82	0.17	0.17
Pomatomus saltatrix	BLUEFI	0.88	0.85	0.05	0.18	*Etropus microstomus*	SMAFLO	0.87	0.79	0.01	0.01
Alosa aestivalis	BLUHER	0.76	0.88	0.09	0.05	*Mustelus canis*	SMODOG	0.92	0.89	0.09	0.17
Zenopsis conchifera	BUCDOR	0.89	0.91	0.04	0.04	*Malacoraja senta*	SMOSKA	0.89	0.88	0.07	0.07
Peprilus triacanthus	BUTTER	0.86	0.77	0.12	0.23	*Majidae*	SPICRA		0.73		0.01
Scyliorhinus retifer	CHADOG	0.95	0.94	0.04	0.04	*Squalus acanthias*	SPIDOG	0.79	0.80	0.32	0.28
Scomber japonicus	CHUBMA	0.74		0.00		*Urophycis regia*	SPOHAK	0.88	0.84	0.09	0.13
Raja eglanteria	CLESKA	0.92	0.92	0.07	0.09	*Leiostomus xanthurus*	SPOTZZ	0.84	0.94	0.02	0.13
Conger oceanicus	CONGEL	0.82		0.03		*Anchoa hepsetus*	STRANC		0.93		0.10
Tautogolabrus adspersus	CUNNER	0.84	0.87	0.05	0.05	*Morone saxatilis*	STRBAS	0.90	0.87	0.10	0.08
Brosme brosme	CUSKZZ	0.89	0.88	0.09	0.09	*Prionotus evolans*	STRSEA	0.89	0.89	0.06	0.08
Lepophidium profundorum	FAWMEL	0.89	0.89	0.04	0.04	*Paralichthys dentatus*	SUMFLO	0.84	0.90	0.11	0.13
Monacanthus hispidus	FILEFS		0.75		0.01	*Tautoga onitis*	TAUTOG		0.84		0.03
Paralichthys oblongus	FOUFLO	0.87	0.83	0.11	0.13	*Amblyraja radiata*	THOSKA	0.88	0.89	0.12	0.14
Enchelyopus cimbrius	FRBERO	0.89	0.88	0.03	0.02	*Lopholatilus chamaeleonticeps*	TILEFI	0.92	0.85	0.03	0.01
Citharichthys arctifrons	GULFLO	0.87	0.86	0.02	0.03	*Cynoscion regalis*	WEAKFI	0.84	0.93	0.03	0.12
Melanogrammus aeglefinus	HADDOC	0.85	0.84	0.20	0.21	*Maurolicus weitzmani*	WEITZP	0.78	0.75	0.01	0.01
Myxine glutinosa	HAGFIS	0.84	0.87	0.02	0.02	*Urophycis tenuis*	WHIHAK	0.87	0.88	0.13	0.14
Cancer borealis	JONCRA	0.74	0.74	0.05	0.05	*Scophthalmus aquosus*	WINDOW	0.83	0.85	0.12	0.13
Ovalipes ocellatus	LADCRA	0.86	0.87	0.01	0.04	*Pseudopleuronectes americanus*	WINFLO	0.88	0.87	0.13	0.14

(Continued)

TABLE 4 | Continued

Species	Abbr	AUC		RMSLE		Species	Abbr	AUC		RMSLE	
		Spring	Fall	Spring	Fall			Spring	Fall	Spring	Fall
Urophycis chesteri	LGFINH	0.92	0.92	0.02	0.02	*Leucoraja ocellata*	WINSKA	0.81	0.88	0.23	0.17
Leucoraja erinacea	LITSKA	0.83	0.86	0.23	0.19	*Glyptocephalus cynoglossus*	WITFLO	0.84	0.90	0.09	0.09
Myoxocephalus octodecemspinosus	LONSCU	0.89	0.88	0.13	0.13	*Cryptacanthodes maculatus*	WRYMOU	0.88	0.88	0.02	0.02
Doryteuthis pealeii	LONSQD	0.89	0.85	0.12	0.22	*Limanda ferruginea*	YELFLO	0.87	0.88	0.12	0.12
Cyclopterus lumpus	LUMPFI	0.79	0.83	0.04	0.04						

Presence/absence model performance statistic is Area under the ROC curve (AUC) index, biomass model performance statistic is Root Mean Squared Log Error (RMSLE) index. Species name abbreviations (Abbr) referred to elsewhere.

see **Supplementary Material**). Since the minimum score for this grouping was 10, all high reliance species had both spring and fall model data included in the study. The high reliance grouping included demersal and pelagic fish and invertebrate species, including cephalopods (squid) and decapods (crab). There were 31 species in the moderate reliance group, some of which were only considered for one season. Like the high reliance group, there were finfish and invertebrates among the taxa in the moderate reliance group and many of the single season species were small pelagic taxa. Twenty-six species were in the low use grouping, many of which fell within the single season category. Finally, there were 16 taxa in the no-reliance group; while they can occur in the lease areas their reliance was below the required threshold.

Utilization of Lease Areas

Habitat in the lease areas varied widely between depictions based on the output of the presence/absence and biomass models. Averaged over season and lease areas and with ratios exceeding 0.7, approximately 38 species used the lease areas as occupancy habitat compared to 42 species that utilized the areas as biomass habitat (**Figures 2A,B**). The areas with the highest number of species making use of the lease areas as occupancy habitat included E1 and P3 and among the lowest were areas E2 through E5, which were located inshore. The areas with the highest number of species using the lease areas as biomass habitat included E1 and P1, two of the more northerly lease areas. The lease areas with the lowest number of species making use of the lease areas as biomass habitat included E2 through E5, the same areas with low numbers of species using the lease areas as occupancy habitat.

When species counts were related to parameters reflecting the position and size of the lease areas, distinct trends emerged. The number of species with ratios > 0.7 for presence/absence models were uncorrelated with latitude in the spring but correlated in the fall season (**Figures 3A,C**). However, there appears to be a stronger relationship between the number of species and latitude for the biomass habitat, which is significant in both seasons (**Figures 3B,D**). Though not explicitly tested with the correlation coefficient, biomass model responses appeared non-linear and suggested lower counts at middle latitudes. There was significant correlation with longitude of the lease areas for both presence/absence and biomass models in both seasons (**Figures 4A–D**), suggesting higher species counts with more eastern lease areas. Counts from spring presence/absence models were positively correlated with distance to the coast of the lease areas (**Figures 5A,C**); however, the relationship in the fall data is less developed. The biomass model counts were also positively related to distance to the coast, although these relationships were non-significant (**Figures 5B,D**). Size of the lease area may be playing a role in species counts; however, the correlations for presence/absence and biomass models were relatively weak (**Figure 6**).

Retrospective Change in Lease Area Habitat Use by Species

Most taxa classified as high reliance species with respect to lease area utilization were also increasingly dependent on the lease area habitats. For presence/absence model output, 80 and 89% of species had medians of significant habitat time series trends that were positive for spring and fall models, respectively (**Figures 7A,B**). Notably, among the species with the greatest increases in occurrence probability in the spring were *Urophycis regia*, *Leucoraja ocellata, Paralichthys dentatus,* and *Squalus acanthias*; in fall, *Citharichthys arctifrons, Raja eglanteria, U. regia,* and *Centropristis striata* were among the species with the greatest increases in occurrence probability. In both spring and fall, *Glyptocephalus cynoglossus* and *Limanda ferruginea* had negative trends in occurrence probability. For biomass model output, 65 and 60% of species in spring and fall, respectively, had increasing trends in biomass habitat index scores (**Figures 8A,B**). In spring, there were exceptional increases in biomass habitat for *Leucoraja erinacea* and *S. acanthias*; however, many species had moderate increases in biomass habitat scores in fall, including *L. erinacea, P. carolinus, S. acanthias, R. eglanteria, P. dentatus, Placopecten magellanicus,* and *U. regia*.

Model Variable Importance as Indicators of Habitat

The different variable classes contributed to presence/absence models in a hierarchical fashion when considering the times a root, mean minimum depth, Gini decreases, and accuracy decrease variable performance measures. Physical and biological variables had a larger influence on the model fits than benthic

TABLE 5 | Species groups and their relative reliance on wind lease areas, based on the sum of habitat indices where Spring is the sum of habitat score indices triggered for spring distributions and Fall, likewise to spring.

High				Moderate				Low				No			
Species	**Spring**	**Fall**	**Total**	**Species**	**Spring**	**Fall**	**Total**	**Species**	**Spring**	**Fall**	**Total**	**Species**	**Spring**	**Fall**	**Total**
BUTTER	6	6	12	ATLCRO	3	6	9	ATLSIL	5	0	5	ACARED	0	0	0
CLESKA	6	6	12	ATLMEN	6	3	9	RTSTIG	0	5	5	ALLFSH	0	0	0
FOUFLO	6	6	12	LADCRA	3	6	9	SPICRA	0	5	5	AMEPLA	0	0	0
GULFLO	6	6	12	OCPOUT	6	3	9	BLUCRA	1	3	4	ATLHAL	0	0	0
LITSKA	6	6	12	SCUPZZ	3	6	9	FAWMEL	1	3	4	ATLWOL	0	0	0
LONSQD	6	6	12	SMAFLO	5	4	9	JONCRA	1	3	4	CUSKZZ	0	0	0
NORSEA	6	6	12	NPUFFR	2	6	8	OFFHAK	1	3	4	HAGFIS	0	0	0
RCKCRA	6	6	12	SILHAK	4	4	8	ROSCAD	0	4	4	LUMPFI	0	0	0
SPOHAK	6	6	12	SPOTZZ	2	6	8	TILEFI	3	1	4	MOUSCL	0	0	0
SUMFLO	6	6	12	AMLOBS	3	4	7	ATLPOU	0	3	3	POLLOC	0	0	0
WINDOW	6	6	12	ATLHER	6	1	7	BARSKA	1	2	3	REDCRA	0	0	0
WINFLO	6	6	12	ATLMAC	6	1	7	CHUBMA	3	0	3	SHANNY	0	0	0
WINSKA	6	6	12	BLUEFI	1	6	7	ATLCOD	2	0	2	SMOSKA	0	0	0
BAYANC	5	6	11	BLUHER	6	1	7	CONGEL	2	0	2	THOSKA	0	0	0
BLABAS	5	6	11	LONSCU	4	3	7	RDSCAD	0	2	2	WHIHAK	0	0	0
SANDLA	6	5	11	MONKFH	4	3	7	WEITZP	2	0	2	WITFLO	0	0	0
YELFLO	6	5	11	REDHAK	4	3	7	WRYMOU	1	1	2				
SEASCA	5	5	10	SHTSQD	4	3	7	AMESHA	0	1	1				
SMODOG	4	6	10	STRSEA	1	6	7	ATLARG	0	1	1				
SPIDOG	6	4	10	ALEWIF	5	1	6	BLAROS	0	1	1				
				ANGSHR	0	6	6	CUNNER	0	1	1				
				ARMSEA	3	3	6	FILEFS	0	1	1				
				BUCDOR	3	3	6	FRBERO	1	0	1				
				CHADOG	3	3	6	HADDOC	0	1	1				
				RHERRI	0	6	6	LGFINH	1	0	1				
				ROSSKA	3	3	6	TAUTOG	0	1	1				
				SEARAV	4	2	6								
				SHORTP	3	3	6								
				STRANC	0	6	6								
				STRBAS	6	0	6								
				WEAKFI	0	6	6								

Total is the sum of spring and fall indices. Groupings from high to no reliance are divisions intended to reflect the role of lease area habitats to individual species.

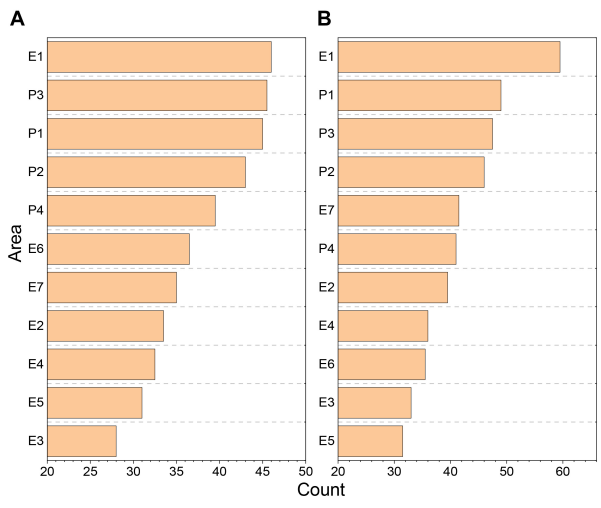

FIGURE 2 | Mean of spring and fall species counts with lease area to NES ecosystem ratio of >0.7 for occurrence **(A)** and biomass **(B)** model output. See **Figure 1** for area abbreviations.

terrain variables. Variable classes with a higher times a root and lower mean minimum depth score (upper left quadrant, **Figure 9A**) are indicative of variables of greater importance. Likewise, variable classes with large Gini decrease and accuracy decrease values (upper right quadrant, **Figure 9B**), are indicative of variables that are more important in explaining species' variation in occupancy. The spring models suggest physical, primary production, and secondary production variable classes were more important than the benthic terrain habitat variables. For the fall models, however, primary and secondary production variables were of greater significance than the physical variables since they occur more frequently in the key quadrants (**Figures 9C,D**). For both sets of seasonal models, the terrain variables made the lowest contribution.

The PC analysis supported the role of temperature and depth in defining fish and macroinvertebrate habitat, but also suggested primary and secondary production variables were critical. The first dimension of the spring and fall PCAs explained 76.9 and 83.2% of the variance in variable importance indices, respectively. For both models, the second dimension explained <15% of the variance, and was not considered in

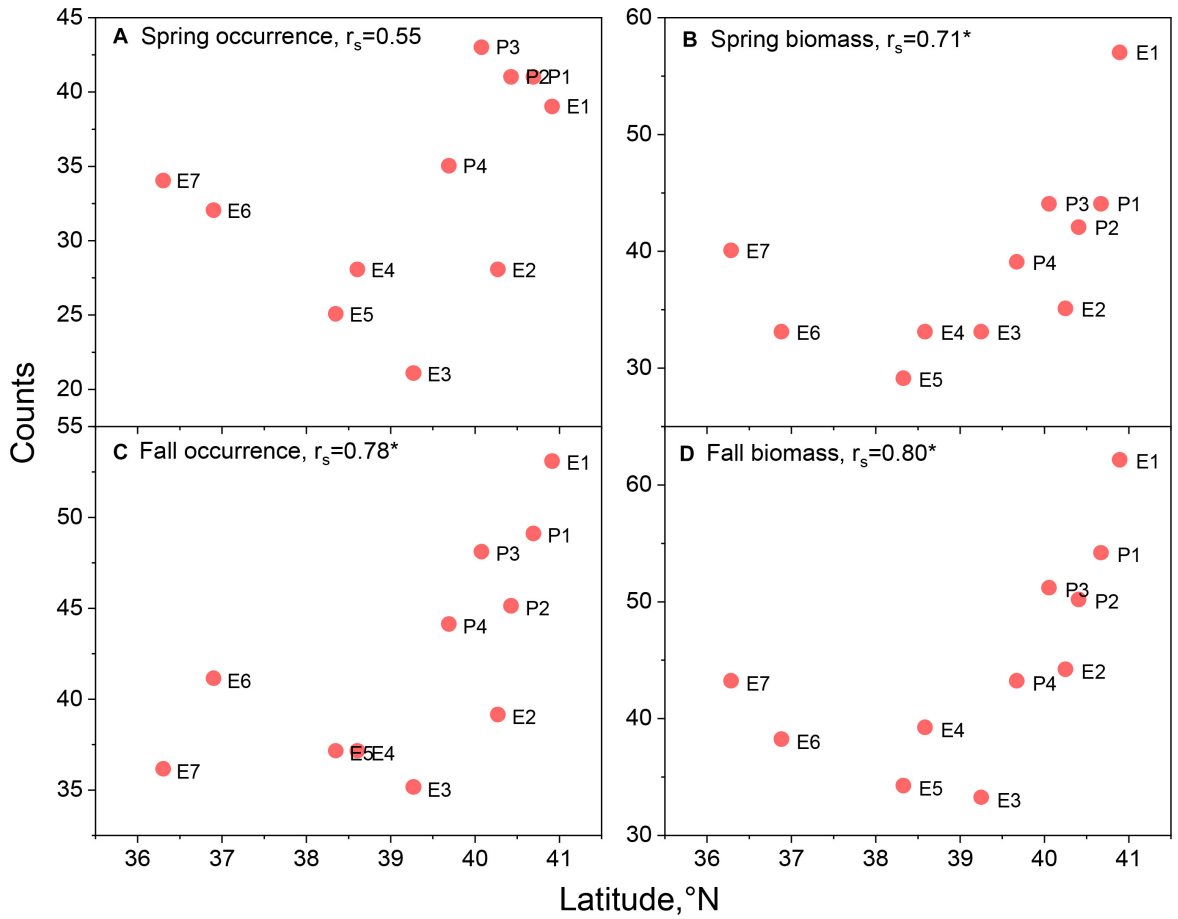

FIGURE 3 | Species counts with lease area to ecosystem ratio of >0.7 versus centroid latitude of the respective lease area for spring occurrence **(A)** and biomass **(B)** models and fall occurrence **(C)** and biomass **(D)** models. Spearman rank order correlations, r_s, noted in panel titles with significant correlations marked with asterisk. See **Figure 1** for area abbreviations.

further analyses. For spring models, the highest PC scores were found with bottom temperature and depth; however, among the top twenty variables were 15 primary and secondary production variables (**Figure 10A**). Only one terrain variable was among the top variables. For fall models, depth and bottom temperature were the top variables, with 17 primary and secondary production variables among the top 20 variables (**Figure 10B**). No terrain variables were among the top 20 variables from the fall models.

DISCUSSION

The characterization of occupancy and biomass habitat of fish and macroinvertebrate species allowed the identification of taxa with high dependency on habitats that overlap the wind lease areas in the NES ecosystem. This raises the potential for impact of offshore wind development on these candidate taxa, where impact is broadly defined as the potential direct effects for some species on their habitat and subsequently their growth and reproduction (Stenberg et al., 2015; Raoux et al., 2017). However, it may also represent indirect effects in that wind farms may change use patterns in the lease areas, such as limiting fishery access and thus reducing fishing mortality for select species (Ashley et al., 2014; Coates et al., 2016). Among the species with the highest fisheries landings in the Middle Atlantic Bight (MAB) in the most recent decade were Atlantic menhaden (*B. tyrannus*), sea scallops (*P. magellanicus*), squids (*Doryteuthis pealeii, Illex illecebrosus*), Atlantic croaker (*Micropogonias undulatus*), scup (*Stenotomus chrysops*), spiny dogfish (*S. acanthias*), summer flounder (*P. dentatus*), and striped bass (*Morone saxatilis*), which were all species with a high to moderate reliance on the lease area habitats. This information can aid ongoing efforts to characterize cumulative risk to fishery species by a range of environmental factors and human activities, including non-fisheries uses (Hobday et al., 2011; Holsman et al., 2017). In particular, summer flounder, scup, and spiny dogfish are among federally managed species evaluated annually in an ecosystem-level risk assessment, and summer flounder fisheries have been found to face multiple other risks (Gaichas et al., 2018).

Despite the importance of multiple species to specific fisheries in the MAB, some were not reliant on the lease areas. For

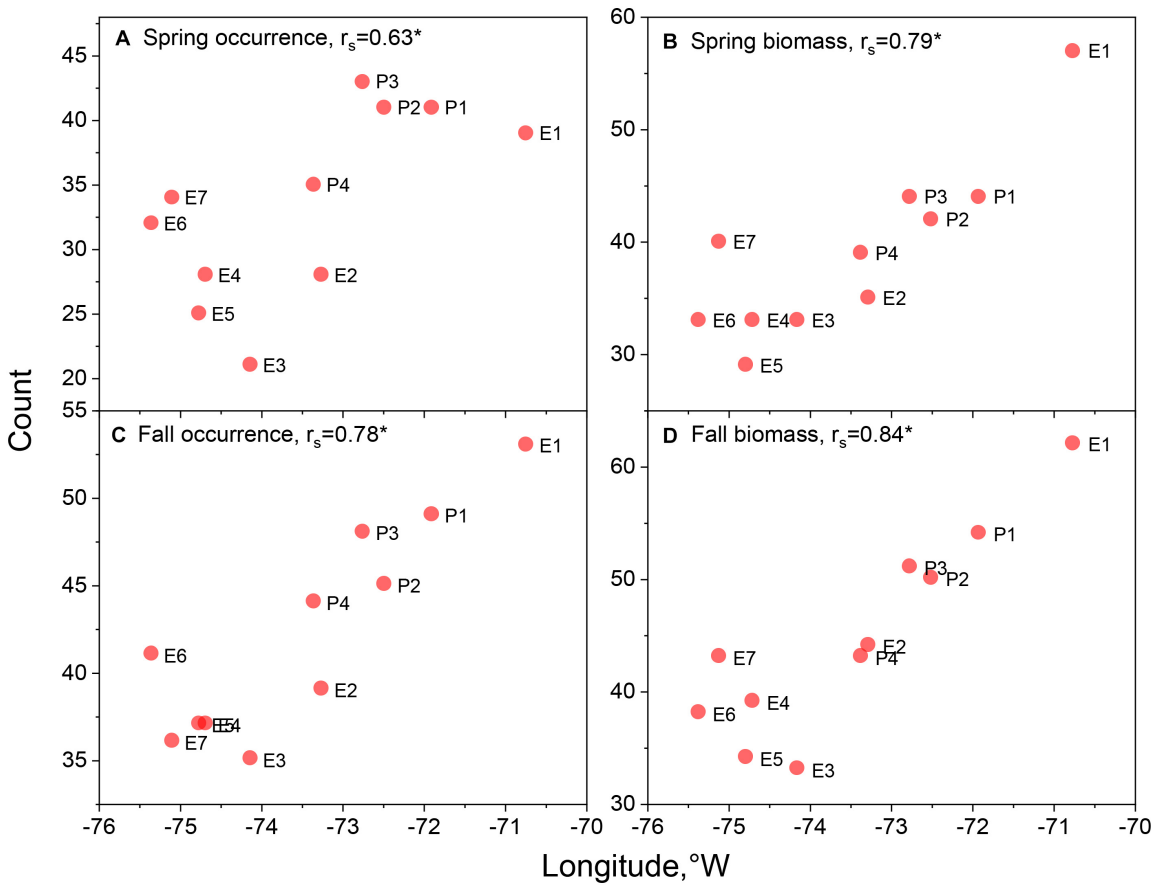

FIGURE 4 | Species counts with lease area to ecosystem ratio of >0.7 versus centroid longitude of the respective lease area for spring occurrence **(A)** and biomass **(B)** models and fall occurrence **(C)** and biomass **(D)** models. Spearman rank order correlations, r_s, noted in panel titles with significant correlations marked with asterisk. See **Figure 1** for area abbreviations.

example, three taxa with high landings in the MAB, blue crab (*Callinectes sapidus*), surf clam (*Spisula solidissima*), and ocean quahogs (*Arctica islandica*), occur in the lease areas, but do not appear in either the high or moderate reliance use groups. This is likely because the clam species are infauna and not effectively sampled in bottom trawl surveys, and blue crab provide high regional landings, but are mostly an inshore estuarine taxon. Despite low habitat use across most of the lease areas, some taxa play pivotal roles in the NES ecosystem. For example, Atlantic cod (*Gadus morhua*), which only meets the criteria of low reliance, has been an ecologically pivotal species supporting regional groundfish fisheries and has declined in recent years due to overfishing and changing climate conditions (Pershing et al., 2015). Hence, even if cod interactions were limited to a small segment of the overall lease area, it would seem prudent to monitor and study the effects of habitat change on cod biology.

Ongoing monitoring is comprehensive in the lease areas for specific fisheries species and ecosystem aspects; however, other components are characterized less well, such as benthic infauna, birds, bats, marine mammals, or highly migratory megafauna. Results here for finfish could potentially be expanded using data from other dedicated surveys. Advances in acoustic and satellite tracking can now identify important foraging and migration habitats for highly migratory tunas, billfishes, and sharks (Wilson et al., 2005; Curtis et al., 2018) and multiple species of seabirds (Montevecchi et al., 2012), and bats, which have been detected up to 21 km offshore (Sjollema et al., 2014). Expanding analyses could be particularly valuable for cetaceans, which were found to occur in higher than expected numbers in lease areas off southern New England (Stone et al., 2017). Habitat models provide critical baselines for assessing potential overlap of many species with wind lease areas (Roberts et al., 2016); in particular, understanding habitat use by endangered North Atlantic right whales in proximity to lease areas (Davis et al., 2017). In the absence of standardized fishery-independent survey data, fishery-dependent catch data may be a useful alternative for developing distribution models for species not well-represented in this study (Hazen et al., 2018).

Climate-driven changes in species distribution may play an important role in changing dependency of species on lease area habitats over time. Center of gravity distribution for most species on the NES have shifted to higher latitudes, which in a practical sense, represents an along-shelf movement from the southwest to the northeast (Nye et al., 2009; Kleisner et al.,

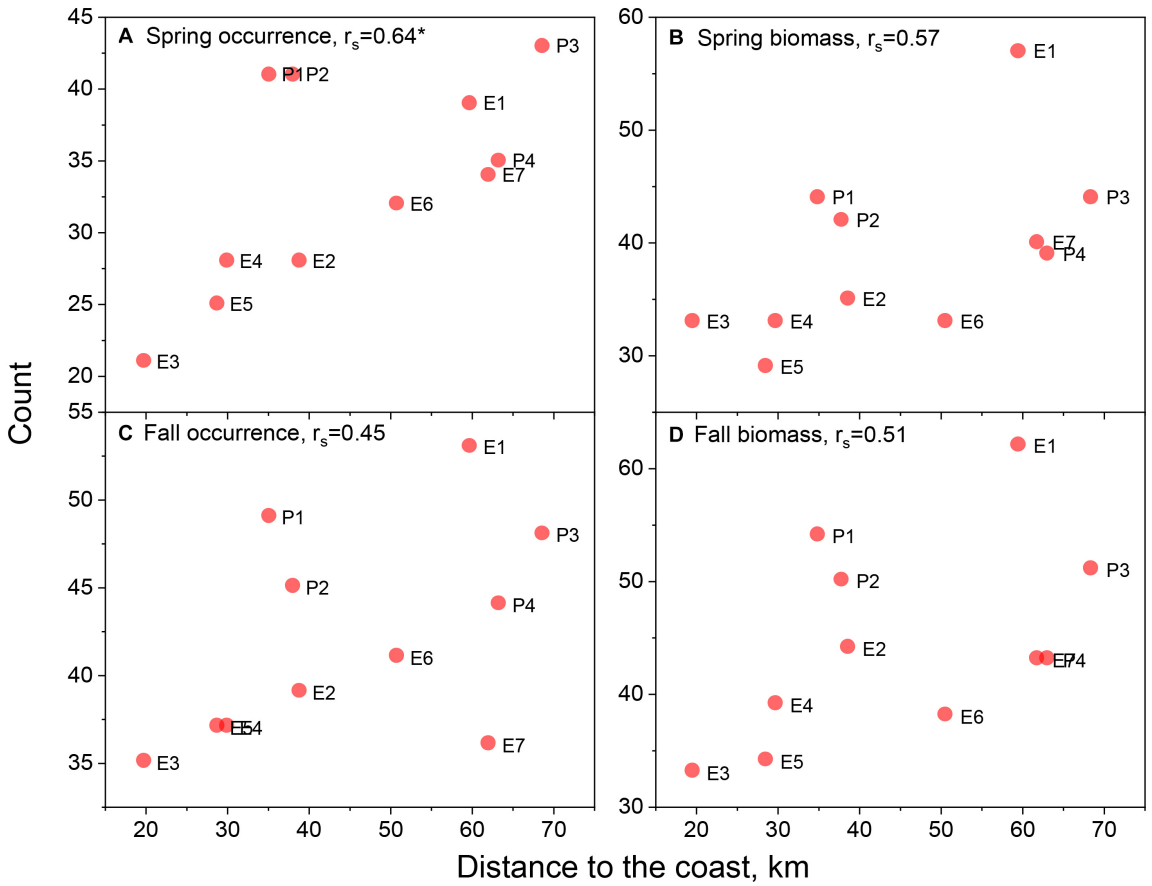

FIGURE 5 | Species counts with lease area to ecosystem ratio of >0.7 versus centroid distance to the coast of the respective lease area for spring occurrence **(A)** and biomass **(B)** models and fall occurrence **(C)** and biomass **(D)** models. Spearman rank order correlations, r_s, noted in panel titles with significant correlations marked with asterisk. See **Figure 1** for area abbreviations.

2016). Concomitant with shifting distributions, habitat use for most taxa have expanded over recent decades (Friedland et al., 2020), suggesting changing climate conditions have expanded the range of useful, or at least tolerable, habitat for many species. Occupancy habitat, reflected by the change in occurrence probability, has increased for most high reliance species within the lease areas. The increasing importance of these habitats is likely the result of ranging dynamic environmental variables like SST and secondary productivity that likely represent currently optimal climate refugia (Poloczanska et al., 2013; Ban et al., 2016). Moreover, these changes could shift more southerly species into these communities (Lurgi et al., 2012; Barceló et al., 2016). Secondary production variables reflect climatological production zones, with some taxa displaying similar distribution shifts as fish. Specifically, many copepod species have experienced a shift in their center of gravity to the northeast, which would reinforce the putative effects of temperature in increased habitat values on a latitudinal basis (Friedland et al., 2019). Finally, factors further complicating our understanding of climate impacts are changes to the habitat and the related ecosystem components within the lease blocks due to energy development itself that will have species-specific outcomes (Langhamer, 2012).

Current marine spatial planning for offshore wind considers the wind resource, the seabed type for installation, the location of designated areas and other key marine users, such as navigation and military uses. Fisheries grounds and vessel transit routes are considered to varying degrees. However, the relative importance of those locations to the actual fisheries species (as demonstrated here) and life history association (Barbut et al., 2020) is generally lacking. As more offshore wind locations are chosen and the extent of the lease areas increases, there will be a need to also understand the temporal use of the areas by the fisheries species and the life stage that may be dependent on the areas (Birchenough and Degraer, 2020). The migratory connections between life history stage habitats (Buscher et al., 2016) and the availability of alternative habitat require consideration when assessing potentially significant impact on species, such as population or community change (Boehlert and Gill, 2010). Determining the changes that do occur will need to be considered over the appropriate time scale for cohort recruitment of species and spatially may be within the jurisdiction of other states, or countries. Larger developments will necessarily have more subsea cables, longer installation times and larger turbine

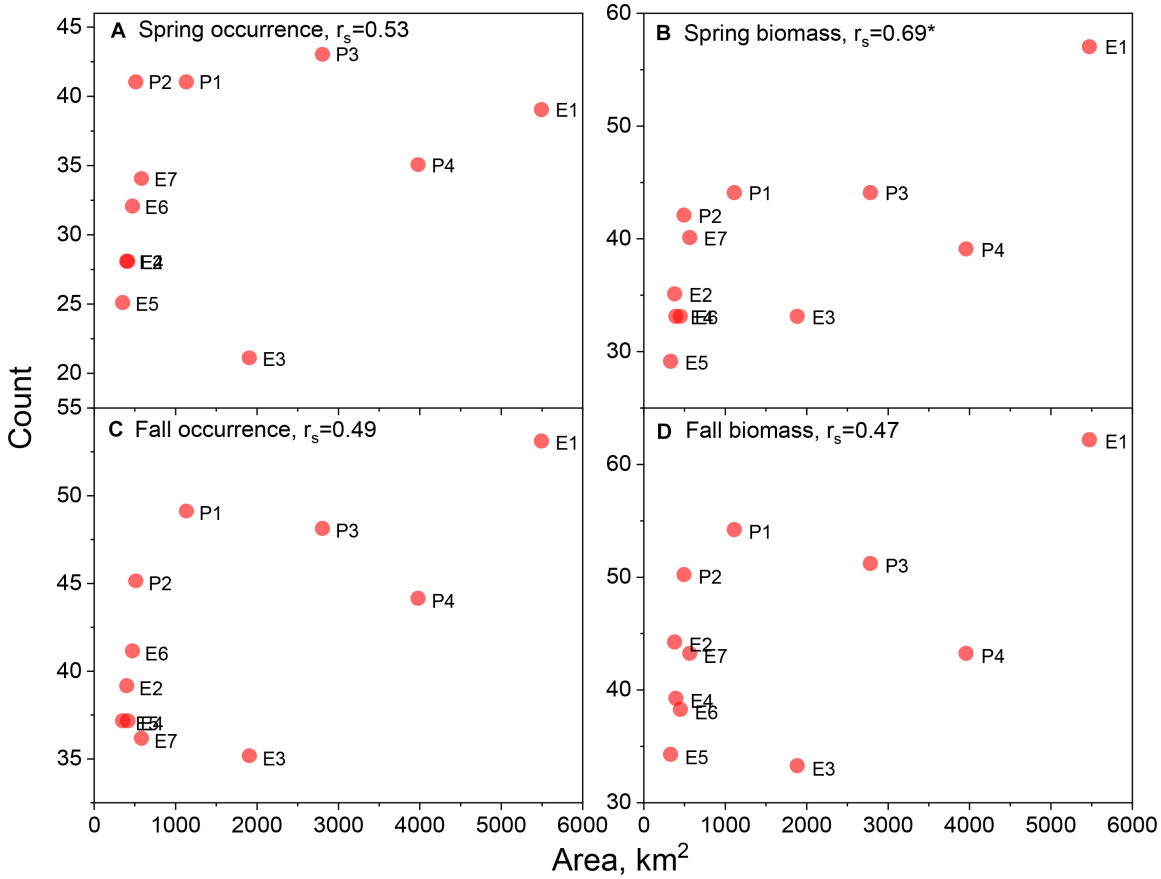

FIGURE 6 | Species counts with lease area to ecosystem ratio of >0.7 versus area of the respective lease area for spring occurrence **(A)** and biomass **(B)** models and fall occurrence **(C)** and biomass **(D)** models. Spearman rank order correlations, r_s, noted in panel titles with significant correlations marked with asterisk. See **Figure 1** for area abbreviations.

structures, spaced further apart than currently sized turbines. Hence, the installation of the offshore wind infrastructure will need to be incorporated into the future analysis of habitat association and productivity of fisheries species in terms of spatial extent and the length of time that the fisheries may be affected.

With the advent of floating turbines, the future interactions between marine resources and energy development may change more toward offshore fisheries, particularly in the Gulf of Maine. With moving offshore, there will be gradients of potential habitat importance and perhaps differential impacts. With this in mind, the assessment of occurrence and productivity of fisheries species will remain a requirement to provide explanatory variables that determine the species association in these new offshore areas. Fish attraction (of different life stages) to the structures and connectivity linked to prevailing currents will likely need determination and occurrence of pelagic predators would be expected. Moving wind turbine structures offshore also increases the likelihood of cables intersecting with migratory species routes, including electromagnetically sensitive fauna such as elasmobranchs, sturgeons, and marine turtles (Hutchison et al., 2020).

Physical (temperature and depth) and biological (primary and secondary production) oceanographic variables were important for determining habitat occupancy for upper trophic level species in this study. Temperature and depth are well known to be strong determinants of fish distribution and abundance in marine ecosystems (Murawski and Finn, 1988). Primary and secondary production form the base of the marine food web and are associated with fisheries yields (Friedland et al., 2012; Stock et al., 2017). Both physical and biological oceanographic variables are likely to be affected by the operation of offshore wind farms (Floeter et al., 2017; Bakhoday-Paskyabi et al., 2018). However, few studies have directly demonstrated linkages between physical and biological oceanographic variables at these facilities. One of the first studies to empirically take on these questions reported increased vertical mixing, doming of the thermocline (i.e., rising of the thermocline to replace the surface mixed layer), and transport of nutrients to the surface mixed layer followed by uptake by phytoplankton in the photic zone (Floeter et al., 2017). Coupled with our findings, this would suggest that wind farm effects on phytoplankton and zooplankton might extend to upper trophic level impacts, potentially modifying the distribution and abundance of finfish and invertebrates. The spatial scale of

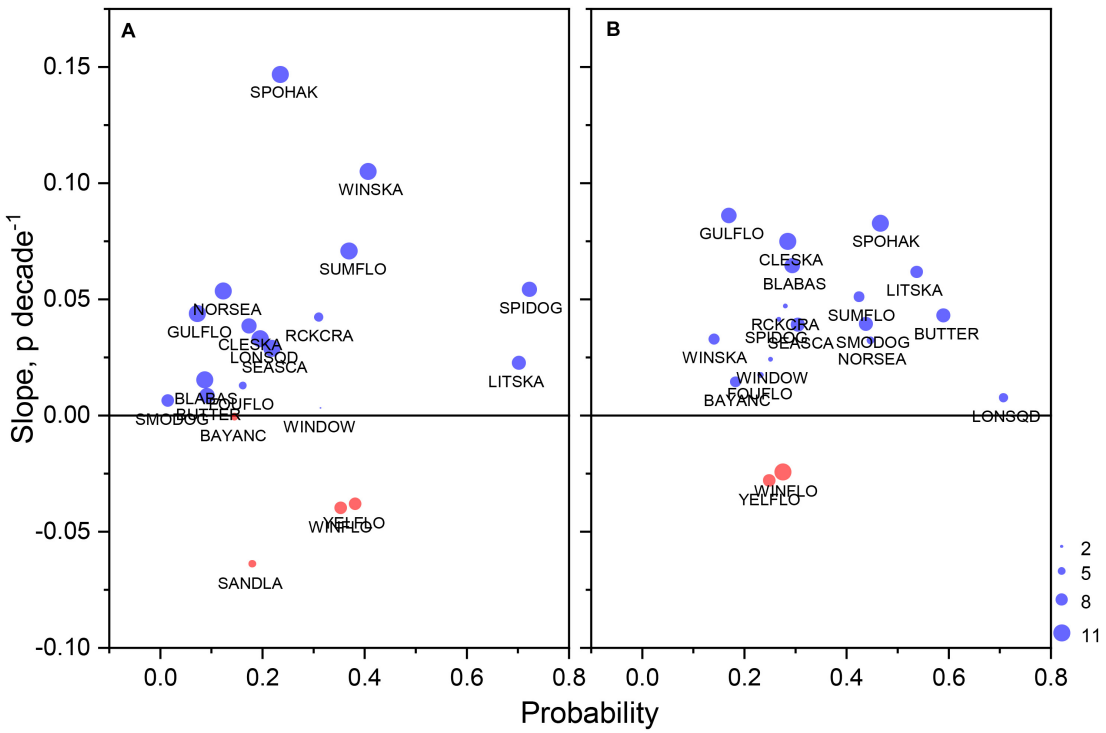

FIGURE 7 | Median occurrence probability of species during the study period versus change in probability as Theil–Sen slope for essential species in existing and proposed lease areas during spring **(A)** and fall **(B)**. Only data associated with significant slopes were used in the calculations; sample size expressed with the size of plotting symbol with maximum of 11 reflecting total number of lease areas in the study. See **Table 4** for species abbreviations.

these effects remains unknown but could range from localized within individual farms to broader spatial scales (Carpenter et al., 2016; Bakhoday-Paskyabi et al., 2018). The implications of these effects for fisheries production and commercial fisheries economics could be significant; hence, we suggest that the effect of wind farms on oceanographic variables be considered in environmental impact analyses.

A well-designed regional wind farm monitoring program is essential to understanding how offshore wind development affects marine ecosystems, and the design of strategic and meaningful assessments to properly contextualize impacts is required (Lindeboom et al., 2015). With respect to our case study, there is currently much discussion about how to conduct monitoring focusing on variables that are sensitive to change, that are sampled with current gear and technologies, and that are indicative of wind farm effects on the biological community and the ecosystem as a whole. To date, limited attention in the research community has been given to how changes in oceanographic variables due to offshore wind farm operation are linked to effects on biological processes. Given the importance of oceanographic variables in determining habitat occupancy of upper trophic levels revealed in our study, we recommend that comprehensive monitoring of these variables be conducted in conjunction with biological monitoring. Specifically, to be cognizant of the factors related to primary and secondary production and how they are linked to patterns of natural resource abundance and distribution.

In addition to regional monitoring, research programs are also under development in the Northeast US led by a new science entity, the Responsible Offshore Science Alliance, which seeks to address research questions through cross-sector collaboration and hypothesis-driven questions development (Dannheim et al., 2020). The findings of this study could inform the development of such research priorities and questions. For example, we can foresee questions related to feeding behavior and diet of species at various life stages and sizes and the associated impacts of wind farm structures. The results of this study show black sea bass (*Centropristis striata*) are highly dependent on habitat in wind areas in both the spring and fall. Because of its attraction to structural habitat and reef formations, black sea bass has previously garnered attention as a species that may benefit from the installation of turbine foundations. As phytoplankton and zooplankton are important drivers of habitat occupancy, species that are planktivorous such as Atlantic menhaden (*Brevoortia tyrannus*), may be sensitive to any changes in biological oceanography caused by wind farm operation. These findings could be used to develop ecosystem simulation models that couple changes in physical and biological oceanography to explore a range of bottom up forcing scenarios caused by wind farm operation and how they might affect upper trophic levels (Pezy et al., 2020). Knowing how the range of primary and secondary productivity values are linked to upper trophic levels would be extremely useful in parameterizing and testing such models.

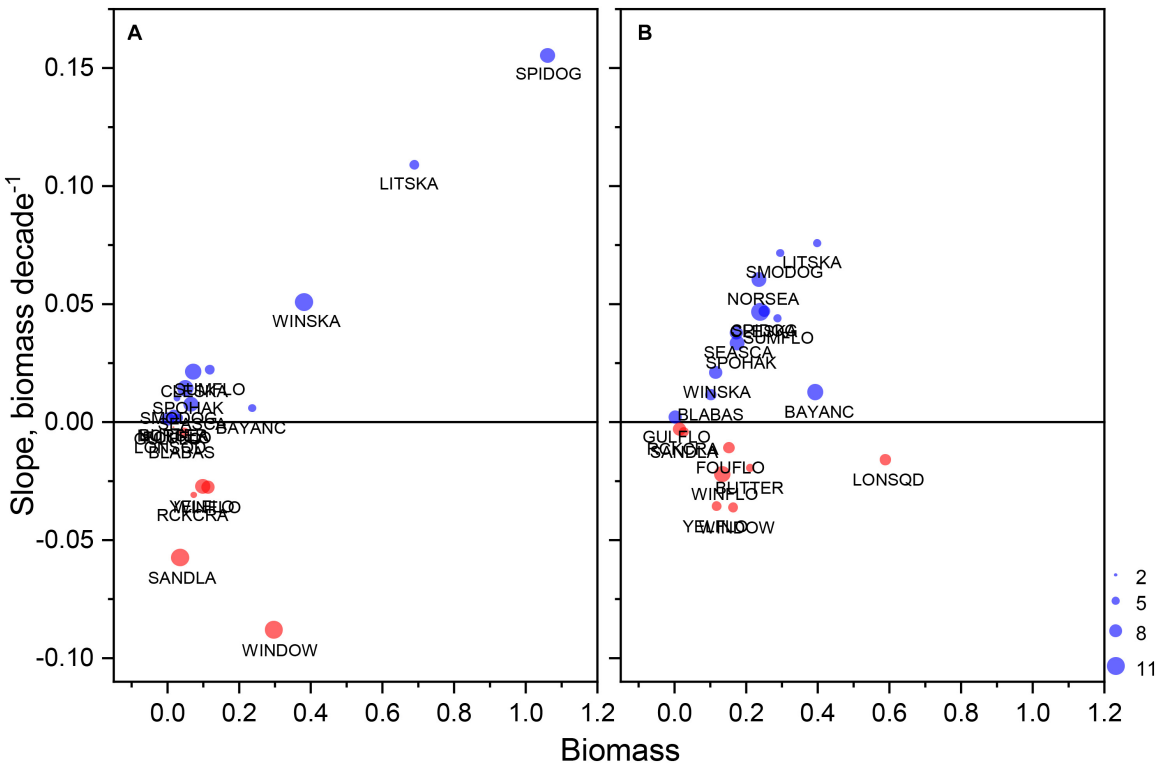

FIGURE 8 | Median biomass index during the study period versus change in biomass index as Theil–Sen slope for essential species in existing and proposed lease areas during spring **(A)** and fall **(B)**. Only data associated with significant slopes were used in the calculations; sample size expressed with the size of plotting symbol with maximum of 11 reflecting total number of lease areas in the study. See **Table 4** for species abbreviations.

Species' occupancy and biomass spatial predictions provide a basis for quantifying baseline conditions for wind energy areas of the NES. However, our methodological framework could have broad applicability to other parts of the world with comparable data availability that are also experiencing similar growth in offshore energy development or other impacts. As part of developers' construction plans, they are charged with understanding the environment they plan to utilize. Their pre-construction monitoring needs to take into account the variability in baseline estimates (Witt et al., 2012). The model predictions in our study provide a method for testing the tradeoffs of defining baseline year lengths for species of high exploitation and ecosystem significance. The results also highlight how the relationship between areas and species can change through time, as many of the high reliance species have increased dependency over time (**Figure 7**). Many of these taxa are experiencing regional shifts, often northeastward (Bell et al., 2015; Kleisner et al., 2017), suggesting that the historical significance of these areas for species should not be the sole metric used to understand prospective wind farm impacts. Using SDMs with forecasted environmental data and fit with the appropriate rigor (Briscoe et al., 2019), may provide increased insight into future interactions between offshore wind development and the fish community.

The study presented here utilized data from NOAA's long-term bottom trawl survey and ecosystem monitoring programs. In the Northeast US, wind development areas overlap with these, and a number of scientific surveys representing more than 315 years of cumulative survey effort, which are executed by NOAA ships and aircraft. Information gathered from these surveys represents one of the most comprehensive data sets on marine ecosystems in the world (Desprespatanjo et al., 1988). In addition to making predictive modeling studies such as this one possible, these surveys support fisheries assessment and management process, protected species assessment and remediation, ecosystem-based fisheries management, and regional and national climate assessments, as well as a number of regional, national, and international science activities (Smith, 2008). Within offshore wind facility areas, survey operations will be curtailed or eliminated under current vessel and aircraft capacities, safety requirements, and monitoring protocols. For example, in the case of the bottom-trawl survey, a stratified-random sampling methodology will no longer be possible because wind energy areas will not be sampled with the current vessel and gear specifications. The same limitation will affect ecosystem survey work. The inability to conduct sampling inside of wind areas will lead to survey bias, a reduction in information, increased uncertainty in stock assessments, and poorly informed

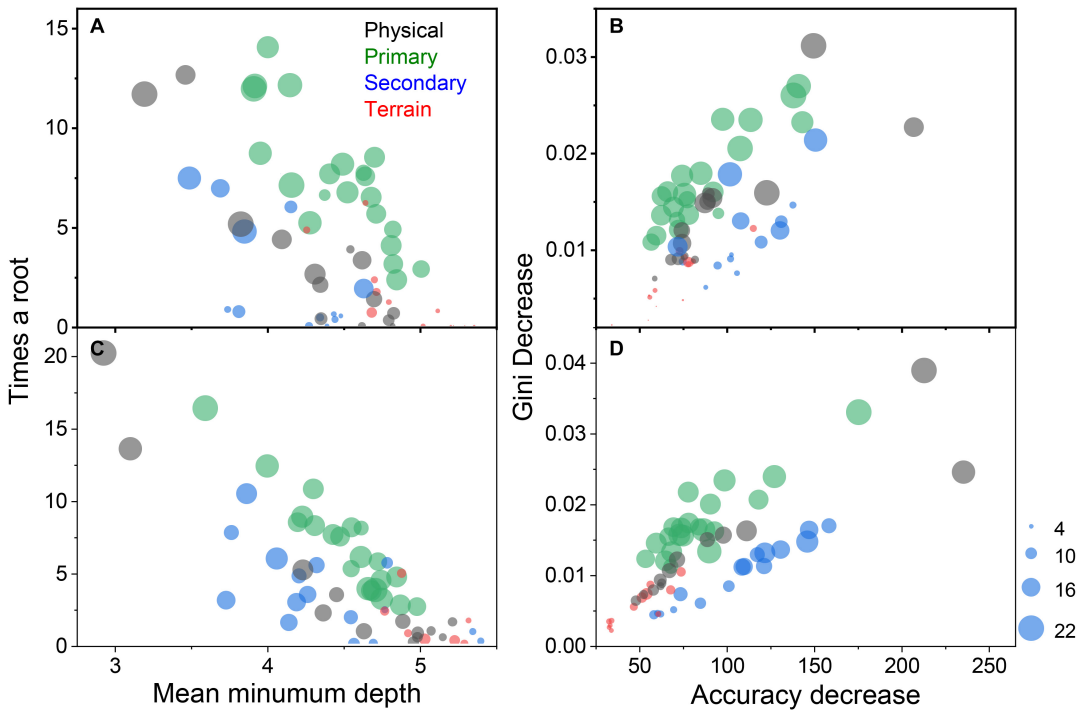

FIGURE 9 | The mean number of times a variable was the root node variable versus the mean minimum depth of a variable in a tree for spring **(A)** and fall **(C)** models; and, the mean decrease in the Gini index of node impurity versus the mean accuracy decrease if a variable were to be removed for spring **(B)** and fall **(D)** models. Variables are color coded by their class: physical (black), primary production (green), secondary production (blue), and benthic terrain complexity (red). Symbol size is scaled by the number of species models the variable was selected.

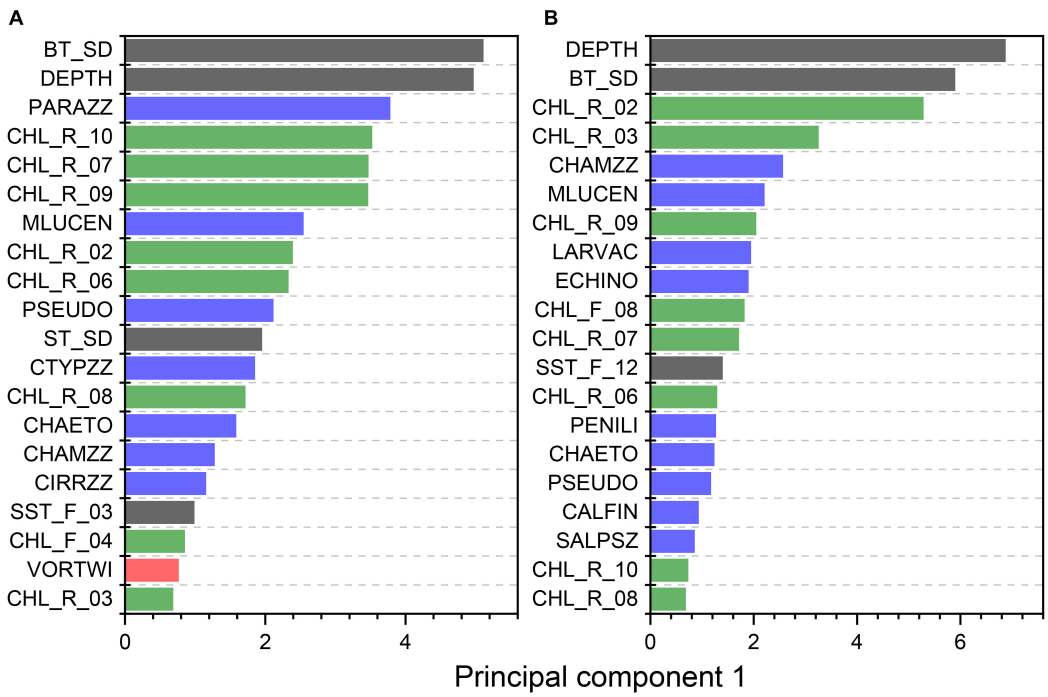

FIGURE 10 | The top twenty variables across all spring **(A)** and fall **(B)** models of high reliance species based on principal component 1 variable scores as an index of importance. Physical, primary production, secondary production, and terrain variables are represented in black, green, blue, and red, respectively. See **Tables 1–3** for variable abbreviations.

management decisions (Boenish et al., 2020). Maintaining the integrity of long-term scientific assessments through development and implementation of novel survey design and gear types is essential to ensure the compatibility of data collected inside and outside of wind farms and of historical data with future data sets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

KF led the analysis with primary drafting of the manuscript by KF, EM, AG, SG, and TC. EA, JM, DC, MM, and DB contributed to the drafting and editing of the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.629230/full#supplementary-material

REFERENCES

Ashley, M. C., Mangi, S. C., and Rodwell, L. D. (2014). The potential of offshore windfarms to act as marine protected areas – A systematic review of current evidence. *Mar. Policy* 45, 301–309. doi: 10.1016/j.marpol.2013.09.002

Bakhoday-Paskyabi, M., Fer, I., and Reuder, J. (2018). Current and turbulence measurements at the FINO1 offshore wind energy site: analysis using 5-beam ADCPs. *Ocean Dyn.* 68, 109–130. doi: 10.1007/s10236-017-1109-5

Ban, S. S., Alidina, H. M., Okey, T. A., Gregg, R. M., and Ban, N. C. (2016). Identifying potential marine climate change refugia: a case study in Canada's Pacific marine ecosystems. *Glob. Ecol. Conserv.* 8, 41–54. doi: 10.1016/j.gecco.2016.07.004

Barbut, L., Vastenhoud, B., Vigin, L., Degraer, S., Volckaert, F. A. M., and Lacroix, G. (2020). The proportion of flatfish recruitment in the North Sea potentially affected by offshore windfarms. *ICES J. Mar. Sci.* 77, 1227–1237. doi: 10.1093/icesjms/fsz050

Barceló, C., Ciannelli, L., Olsen, E. M., Johannessen, T., and Knutsen, H. (2016). Eight decades of sampling reveal a contemporary novel fish assemblage in coastal nursery habitats. *Glob. Change Biol.* 22, 1155–1167. doi: 10.1111/gcb.13047

Belkin, I. M., and O'Reilly, J. E. (2009). An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. *J. Mar. Syst.* 78, 319–326. doi: 10.1016/j.jmarsys.2008.11.018

Bell, R. J., Richardson, D. E., Hare, J. A., Lynch, P. D., and Fratantoni, P. S. (2015). Disentangling the effects of climate, abundance, and size on the distribution of marine fish: an example based on four stocks from the Northeast US shelf. *ICES J. Mar. Sci.* 72, 1311–1322. doi: 10.1093/icesjms/fsu217

Birchenough, S. N. R., and Degraer, S. (2020). Science in support of ecologically sound decommissioning strategies for offshore man-made structures: taking stock of current knowledge and considering future challenges. *ICES J. Mar. Sci.* 77, 1075–1078. doi: 10.1093/icesjms/fsaa039

Boehlert, G., and Gill, A. (2010). Environmental and ecological effects of ocean renewable energy development – a current synthesis. *Oceanog* 23, 68–81. doi: 10.5670/oceanog.2010.46

Boenish, R., Willard, D., Kritzer, J. P., and Reardon, K. (2020). Fisheries monitoring: perspectives from the United States. *Aquac. Fish.* 5, 131–138. doi: 10.1016/j.aaf.2019.10.002

Brandt, M. J., Dragon, A.-C., Diederichs, A., Bellmann, M. A., Wahl, V., Piper, W., et al. (2018). Disturbance of harbour porpoises during construction of the first seven offshore wind farms in Germany. *Mar. Ecol. Prog. Ser.* 596, 213–232. doi: 10.3354/meps12560

Briscoe, N. J., Elith, J., Salguero−Gómez, R., Lahoz−Monfort, J. J., Camac, J. S., Giljohann, K. M., et al. (2019). Forecasting species range dynamics with process-explicit models: matching methods to applications. *Ecol. Lett.* 22, 1940–1956. doi: 10.1111/ele.13348

Buscher, E., Olson, A. M., Pascoe, E. S., Weil, J., and Juanes, F. (2016). David H. Secor: migration ecology of marine fishes. *Rev. Fish. Biol. Fish.* 26, 609–610. doi: 10.1007/s11160-016-9423-4

Carpenter, J. R., Merckelbach, L., Callies, U., Clark, S., Gaslikova, L., and Baschek, B. (2016). Potential impacts of offshore wind farms on North Sea Stratification. *PLoS One* 11:e0160830. doi: 10.1371/journal.pone.0160830

Carpenter, S. (2020). *Offshore Wind Companies Are Racing To Develop America's East Coast. First They Must Appease The Fishermen. Forbes*. Available online at: https://www.forbes.com/sites/scottcarpenter/2020/06/16/offshore-wind-companies-are-racing-to-develop-americas-east-coast-first-they-must-appease-the-fishermen/(accessed July 7, 2020).

Causon, P. D., and Gill, A. B. (2018). Linking ecosystem services with epibenthic biodiversity change following installation of offshore wind farms. *Environ. Sci. Policy* 89, 340–347. doi: 10.1016/j.envsci.2018.08.013

Coates, D. A., Kapasakali, D.-A., Vincx, M., and Vanaverbeke, J. (2016). Short-term effects of fishery exclusion in offshore wind farms on macrofaunal communities in the Belgian part of the North Sea. *Fish. Res.* 179, 131–138. doi: 10.1016/j.fishres.2016.02.019

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., et al. (2014). Changes in the global value of ecosystem services. *Glob. Environ. Change* 26, 152–158. doi: 10.1016/j.gloenvcha.2014.04.002

Crear, D. P., Watkins, B. E., Saba, V. S., Graves, J. E., Jensen, D. R., Hobday, A. J., et al. (2020). Contemporary and future distributions of cobia, *Rachycentron canadum*. *Divers. Distrib.* 26, 1002–1015. doi: 10.1111/ddi.13079

Curtis, T. H., Metzger, G., Fischer, C., McBride, B., McCallister, M., Winn, L. J., et al. (2018). First insights into the movements of young-of-the-year white sharks (*Carcharodon carcharias*) in the western North Atlantic Ocean. *Sci. Rep.* 8:10794. doi: 10.1038/s41598-018-29180-5

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., and Hess, K. T. (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792. doi: 10.1890/07-0539.1

Dannheim, J., Bergström, L., Birchenough, S. N. R., Brzana, R., Boon, A. R., Coolen, J. W. P., et al. (2020). Benthic effects of offshore renewables: identification of knowledge gaps and urgently needed research. *ICES J. Mar. Sci.* 77, 1092–1108. doi: 10.1093/icesjms/fsz018

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Bort Thornton, J., et al. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. *Sci. Rep.* 7:13460. doi: 10.1038/s41598-017-13359-3

De Mesel, I., Kerckhof, F., Norro, A., Rumes, B., and Degraer, S. (2015). Succession and seasonal dynamics of the epifauna community on offshore wind farm foundations and their role as stepping stones for non-indigenous species. *Hydrobiologia* 756, 37–50. doi: 10.1007/s10750-014-2157-1

Desprespatanjo, L. I., Azarovitz, T. R., and Byrne, C. J. (1988). 25 years of fish surveys in the northwest atlantic - the NMFS northeast fisheries centers bottom trawl survey program. *Mar. Fish. Rev.* 50, 69–71.

Floeter, J., van Beusekom, J. E. E., Auch, D., Callies, U., Carpenter, J., Dudeck, T., et al. (2017). Pelagic effects of offshore wind farm foundations in the stratified North Sea. *Prog. Oceanog.* 156, 154–173. doi: 10.1016/j.pocean.2017.07.003

Friedland, K. D., Langan, J. A., Large, S. I., Selden, R. L., Link, J. S., Watson, R. A., et al. (2020). Changes in higher trophic level productivity, diversity and niche

space in a rapidly warming continental shelf ecosystem. *Sci. Total Environ.* 704:135270. doi: 10.1016/j.scitotenv.2019.135270

Friedland, K. D., McManus, M. C., Morse, R. E., and Link, J. S. (2019). Event scale and persistent drivers of fish and macroinvertebrate distributions on the Northeast US Shelf. *ICES J. Mar. Sci.* 76, 1316–1334. doi: 10.1093/icesjms/fsy167

Friedland, K. D., Stock, C., Drinkwater, K. F., Link, J. S., Leaf, R. T., Shank, B. V., et al. (2012). Pathways between primary production and fisheries yields of large marine ecosystems. *PLoS One* 7:e0028945. doi: 10.1371/journal.pone.0028945

Friedman, A., Pizarro, O., Williams, S. B., and Johnson-Roberson, M. (2012). Multi-scale measures of rugosity, slope and aspect from benthic stereo image reconstructions. *PLoS One* 7:e0050440. doi: 10.1371/journal.pone.0050440

Gaichas, S. K., DePiper, G. S., Seagraves, R. J., Muffley, B. W., Sabo, M. G., Colburn, L. L., et al. (2018). Implementing ecosystem approaches to fishery management: risk assessment in the US Mid-Atlantic. *Front. Mar. Sci.* 5:442. doi: 10.3389/fmars.2018.00442

Hatch, L., Clark, C., Merrick, R., Van Parijs, S., Ponirakis, D., Schwehr, K., et al. (2008). Characterizing the relative contributions of large vessels to total ocean noise fields: a case study using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. *Environ. Manag.* 42, 735–752. doi: 10.1007/s00267-008-9169-4

Hazen, E. L., Scales, K. L., Maxwell, S. M., Briscoe, D. K., Welch, H., Bograd, S. J., et al. (2018). A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Sci. Adv.* 4:eaar3001. doi: 10.1126/sciadv.aar3001

Hobday, A. J., Smith, A. D. M., Stobutzki, I. C., Bulman, C., Daley, R., Dambacher, J. M., et al. (2011). Ecological risk assessment for the effects of fishing. *Fish. Res.* 108, 372–384. doi: 10.1016/j.fishres.2011.01.013

Hobson, R. D. (1972). "Surface roughness in topography: quantitative approach," in *Spatial Analysis in Geomorphology*, ed. R. J. Chorley (New York, NY: Harper and Row), 221–245. doi: 10.4324/9780429273346-8

Holsman, K., Samhouri, J., Cook, G., Hazen, E., Olsen, E., Dillard, M., et al. (2017). An ecosystem-based approach to marine risk assessment. *Ecosyst. Health Sust.* 3:e01256. doi: 10.1002/ehs2.1256

Hooper, T., Ashley, M., and Austen, M. (2015). Perceptions of fishers and developers on the co-location of offshore wind farms and decapod fisheries in the UK. *Mar. Policy* 61, 16–22. doi: 10.1016/j.marpol.2015.06.031

Hutchison, Z. L., Gill, A. B., Sigray, P., He, H., and King, J. W. (2020). Anthropogenic electromagnetic fields (EMF) influence the behaviour of bottom-dwelling marine species. *Sci. Rep.* 10:4219. doi: 10.1038/s41598-020-60793-x

Kane, J. (2007). Zooplankton abundance trends on Georges Bank, 1977-2004. *ICES J. Mar. Sci.* 64, 909–919. doi: 10.1093/icesjms/fsm066

Kinlan, B. P., Winship, A. J., White, T. P., and Christensen, J. (2016). *Modeling At-Sea Occurrence and Abundance of Marine Birds to Support Atlantic Marine Renewable Energy Planning: Phase I Report*. Washington, DC: U.S. Department of the Interior, xvii+113. OCS Study BOEM 2016-039.

Kleisner, K. M., Fogarty, M. J., McGee, S., Barnette, A., Fratantoni, P., Greene, J., et al. (2016). The effects of sub-regional climate velocity on the distribution and spatial extent of marine species assemblages. *PLoS One* 11:e0149220. doi: 10.1371/journal.pone.0149220

Kleisner, K. M., Fogarty, M. J., McGee, S., Hare, J. A., Moret, S., Perretti, C. T., et al. (2017). Marine species distribution shifts on the US northeast continental shelf under continued ocean warming. *Prog. Oceanogr.* 153, 24–36. doi: 10.1016/j.pocean.2017.04.001

Langhamer, O. (2012). Artificial reef effect in relation to offshore renewable energy conversion: state of the art. *Sci. World J.* 2012:e386713. doi: 10.1100/2012/386713

Lindeboom, H., Degraer, S., Dannheim, J., Gill, A. B., and Wilhelmsson, D. (2015). Offshore wind park monitoring programmes, lessons learned and recommendations for the future. *Hydrobiologia* 756, 169–180. doi: 10.1007/s10750-015-2267-4

Lundblad, E., Wright, D. J., Miller, J., Larkin, E. M., Rinehart, R., Anderson, S. M., et al. (2006). A Benthic Terrain Classification Scheme for American Samoa. *Mar. Geod.* 29, 89–111. doi: 10.1080/01490410600738021

Lurgi, M., López, B. C., and Montoya, J. M. (2012). Novel communities from climate change. *Philos. Trans. R. Soci. B Biol. Sci.* 367, 2913–2922. doi: 10.1098/rstb.2012.0238

Maritorena, S., d'Andon, O. H. F., Mangin, A., and Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: characteristics, benefits and issues. *Remote Sens. Environ.* 114, 1791–1804. doi: 10.1016/j.rse.2010.04.002

Mavraki, N., De Mesel, I., Degraer, S., Moens, T., and Vanaverbeke, J. (2020). Resource niches of Co-occurring invertebrate species at an offshore wind turbine indicate a substantial degree of trophic plasticity. *Front. Mar. Sci.* 7:379. doi: 10.3389/fmars.2020.00379

Mazur, M. D., Friedland, K. D., McManus, M. C., and Goode, A. G. (2020). Dynamic changes in American lobster suitable habitat distribution on the Northeast U.S. Shelf linked to oceanographic conditions. *Fish. Oceanogr.* 29, 349–365. doi: 10.1111/fog.12476

Methratta, E. T. (2020). Monitoring fisheries resources at offshore wind farms: BACI vs. BAG designs. *ICES J. Mar. Sci.* 77, 890–900. doi: 10.1093/icesjms/fsaa026

Miller, T. J., Das, C., Politis, P. J., Miller, A. S., Lucey, S. M., Legault, C. M., et al. (2010). *Estimation of Albatross IV to Henry B. Bigelow calibration factors*. Northeast Fisheries Science Center Reference Document 10-05. Washington, DC: NOAA, 238.

Montevecchi, W. A., Hedd, A., McFarlane Tranquilla, L., Fifield, D. A., Burke, C. M., Regular, P. M., et al. (2012). Tracking seabirds to identify ecologically important and high risk marine areas in the western North Atlantic. *Biol. Conserv.* 156, 62–71. doi: 10.1016/j.biocon.2011.12.001

Moore, C., Drazen, J. C., Radford, B. T., Kelley, C., and Newman, S. J. (2016). Improving essential fish habitat designation to support sustainable ecosystem-based fisheries management. *Mar. Policy* 69, 32–41. doi: 10.1016/j.marpol.2016.03.021

Murawski, S. A., and Finn, J. T. (1988). Biological bases for mixed-species fisheries: species co-distribution in relation to environmental and biotic variables. *Can. J. Fish. Aquat. Sci.* 45, 1720–1735. doi: 10.1139/f88-204

Murphy, M. A., Evans, J. S., and Storfer, A. (2010). Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91, 252–261. doi: 10.1890/08-0879.1

Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Mar. Ecol. Prog. Ser.* 393, 111–129. doi: 10.3354/Meps08220

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science* 350, 809–812. doi: 10.1126/science.aac9819

Pezy, J.-P., Raoux, A., and Dauvin, J.-C. (2020). An ecosystem approach for studying the impact of offshore wind farms: a French case study. *ICES J. Mar. Sci.* 77, 1238–1246. doi: 10.1093/icesjms/fsy125

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., et al. (2013). Global imprint of climate change on marine life. *Nat. Clim. Chang.* 3, 919–925. doi: 10.1038/Nclimate1958

Popper, A. N., and Hawkins, A. D. (2019). An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes. *J. Fish Biol.* 94, 692–713. doi: 10.1111/jfb.13948

Raoux, A., Tecchio, S., Pezy, J.-P., Lassalle, G., Degraer, S., Wilhelmsson, D., et al. (2017). Benthic and fish aggregation inside an offshore wind farm: which effects on the trophic web functioning? *Ecol. Ind.* 72, 33–46. doi: 10.1016/j.ecolind.2016.07.037

Riley, S. J., DeGloria, S. D., and Elliot, R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Int. J. Sci.* 5, 23–27.

Roberts, J. J., Best, B. D., Mannocci, L., Fujioka, E., Halpin, P. N., Palka, D. L., et al. (2016). Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Sci. Rep.* 6:22615. doi: 10.1038/srep22615

Sjollema, A. L., Gates, J. E., Hilderbrand, R. H., and Sherwell, J. (2014). Offshore activity of bats along the mid-atlantic coast. *Northeast. Nat.* 21, 154–163. doi: 10.1656/045.021.0201

Slavik, K., Lemmen, C., Zhang, W., Kerimoglu, O., Klingbeil, K., and Wirtz, K. W. (2019). The large-scale impact of offshore wind farm structures on pelagic primary productivity in the southern North Sea. *Hydrobiologia* 845, 35–53. doi: 10.1007/s10750-018-3653-5

Smith, T. D. (2008). "A History of fisheries and their science and management," in *Handbook of Fish Biology and Fisheries*, eds P. J. B. Hart and J. D. Reynolds

(Hoboken, NJ: John Wiley & Sons, Ltd), 61–83. doi: 10.1002/9780470693919.ch4

Smolinski, S., and Radtke, K. (2017). Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques. *ICES J. Mar. Sci.* 74, 102–111. doi: 10.1093/icesjms/fsw136

Stenberg, C., Støttrup, J. G., van Deurs, M., Berg, C. W., Dinesen, G. E., Mosegaard, H., et al. (2015). Long-term effects of an offshore wind farm in the North Sea on fish communities. *Mar. Ecol. Prog. Ser.* 528, 257–265. doi: 10.3354/meps11261

Stock, C. A., John, J. G., Rykaczewski, R. R., Asch, R. G., Cheung, W. W. L., Dunne, J. P., et al. (2017). Reconciling fisheries catch and ocean productivity. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1441–E1449. doi: 10.1073/pnas.1610238114

Stone, K. M., Leiter, S. M., Kenney, R. D., Wikgren, B. C., Thompson, J. L., Taylor, J. K. D., et al. (2017). Distribution and abundance of cetaceans in a wind energy development area offshore of Massachusetts and Rhode Island. *J. Coas.t Conserv.* 21, 527–543. doi: 10.1007/s11852-017-0526-4

UNEP (2019). *Global Trends in Renewable Energy Investment 2019*. Nairobi: UNEP-UN Environment Program.

Wahlberg, M., and Westerberg, H. (2005). Hearing in fish and their reactions to sounds from offshore wind farms. *Mar. Ecol. Prog. Ser.* 288, 295–309. doi: 10.3354/meps288295

Wang, T., Yu, W., Zou, X., Zhang, D., Li, B., Wang, J., et al. (2018). Zooplankton community responses and the relation to environmental factors from established offshore wind farms within the Rudong Coastal Area of China. *J. Coast. Res.* 34, 843–855. doi: 10.2112/JCOASTRES-D-17-00058.1

Weber, E. D., Chao, Y., and Chai, F. (2018). Performance of fish-habitat classifiers based on derived predictors from a coupled biophysical model. *J. Mar. Syst.* 186, 105–114. doi: 10.1016/j.jmarsys.2018.06.012

Wilding, T. A. (2014). Effects of man-made structures on sedimentary oxygenation: extent, seasonality and implications for offshore renewables. *Mar. Environ. Res.* 97, 39–47. doi: 10.1016/j.marenvres.2014.01.011

Wilhelmsson, D., Malm, T., and Öhman, M. C. (2006). The influence of offshore windpower on demersal fish. *ICES J. Mar. Sci.* 63, 775–784. doi: 10.1016/j.icesjms.2006.02.001

Wilson, S. G., Lutcavage, M. E., Brill, R. W., Genovese, M. P., Cooper, A. B., and Everly, A. W. (2005). Movements of bluefin tuna (*Thunnus thynnus*) in the northwestern Atlantic Ocean recorded by pop-up satellite archival tags. *Mar. Biol.* 146, 409–423. doi: 10.1007/s00227-004-1445-0

Winship, A. J., Kinlan, B. P., White, T. P., Leirness, J. B., and Christensen, J. (2018). *Modeling At-Sea Density of Marine Birds to Support Atlantic Marine Renewable Energy Planning: Final Report*. Washington, DC: U.S. Department of the Interior, x+67. OCS Study BOEM 2018-010.

Witt, M. J., Sheehan, E. V., Bearhop, S., Broderick, A. C., Conley, D. C., Cotterell, S. P., et al. (2012). Assessing wave energy effects on biodiversity: the Wave Hub experience. *Philos. Trans. R. So. A Math. Phys. Eng. Sci.* 370, 502–529. doi: 10.1098/rsta.2011.0265

Yue, S., Pilon, P., Phinney, B., and Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Process.* 16, 1807–1829. doi: 10.1002/hyp.1095

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect those of their agencies.

Conflict of Interest: DC was employed by ECS Federal, and EM was employed by IBSS Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 29 April 2021
doi: 10.3389/fmars.2021.631262

A Novel Framework to Predict Relative Habitat Selection in Aquatic Systems: Applying Machine Learning and Resource Selection Functions to Acoustic Telemetry Data From Multiple Shark Species

Lucas P. Griffin[1], Grace A. Casselberry[1], Kristen M. Hart[2], Adrian Jordaan[1], Sarah L. Becker[1], Ashleigh J. Novak[1], Bryan M. DeAngelis[3], Clayton G. Pollock[4], Ian Lundgren[5], Zandy Hillis-Starr[6], Andy J. Danylchuk[1†] and Gregory B. Skomal[7†]*

[1] Department of Environmental Conservation, University of Massachusetts Amherst, Amherst, MA, United States, [2] Wetland and Aquatic Research Center, United States Geological Survey, Davie, FL, United States, [3] The Nature Conservancy, Narragansett, RI, United States, [4] National Park Service, Key West, FL, United States, [5] Office of Habitat Conservation, NOAA Fisheries, Silver Spring, MD, United States, [6] National Park Service, Christiansted, US Virgin Islands, [7] Massachusetts Division of Marine Fisheries, New Bedford, MA, United States

OPEN ACCESS

Edited by:
Mark J. Henderson,
United States Geological Survey,
United States

Reviewed by:
Ross Dwyer,
University of the Sunshine Coast,
Australia
Vinay Udyawer,
Australian Institute of Marine Science
(AIMS), Australia

***Correspondence:**
Lucas P. Griffin
lucaspgriffin@gmail.com

[†] These authors have contributed equally to this work and share last authorship

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: 19 November 2020
Accepted: 19 March 2021
Published: 29 April 2021

Citation:
Griffin LP, Casselberry GA, Hart KM, Jordaan A, Becker SL, Novak AJ, DeAngelis BM, Pollock CG, Lundgren I, Hillis-Starr Z, Danylchuk AJ and Skomal GB (2021) A Novel Framework to Predict Relative Habitat Selection in Aquatic Systems: Applying Machine Learning and Resource Selection Functions to Acoustic Telemetry Data From Multiple Shark Species. Front. Mar. Sci. 8:631262. doi: 10.3389/fmars.2021.631262

Resource selection functions (RSFs) have been widely applied to animal tracking data to examine relative habitat selection and to help guide management and conservation strategies. While readily used in terrestrial ecology, RSFs have yet to be extensively used within marine systems. As acoustic telemetry continues to be a pervasive approach within marine environments, incorporation of RSFs can provide new insights to help prioritize habitat protection and restoration to meet conservation goals. To overcome statistical hurdles and achieve high prediction accuracy, machine learning algorithms could be paired with RSFs to predict relative habitat selection for a species within and even outside the monitoring range of acoustic receiver arrays, making this a valuable tool for marine ecologists and resource managers. Here, we apply RSFs using machine learning to an acoustic telemetry dataset of four shark species to explore and predict species-specific habitat selection within a marine protected area. In addition, we also apply this RSF-machine learning approach to investigate predator-prey relationships by comparing and averaging tiger shark relative selection values with the relative selection values derived for eight potential prey-species. We provide methodological considerations along with a framework and flexible approach to apply RSFs with machine learning algorithms to acoustic telemetry data and suggest marine ecologists and resource managers consider adopting such tools to help guide both conservation and management strategies.

Keywords: resource selection, space use, acoustic telemetry, machine learning, random forest, marine protected area, sharks, predator-prey

Abbreviations: BIRNM, Buck Island Reef National Monument; BBMM, Brownian bridge movement model; COAs, centers of activity; GLMM, generalized linear mixed model; MPA, marine protected area; RF, random forest; RSF, resource selection function.

INTRODUCTION

Habitat loss and degradation are two of the largest drivers of loss in global biodiversity (Hoekstra et al., 2005), making identifying important habitats critical for resource managers to prioritize habitat protection for species of concern (Morris and Kingston, 2002; Chetkiewicz and Boyce, 2009; Heinrichs et al., 2017). Habitat selection is driven by the physical, chemical, and biological composition and condition of an area that is occupied by a given animal (Block and Brennan, 1993). Thus, behavioral choices related to selection are ultimately determined by a wide range of coupled and uncoupled abiotic and biotic factors, such as energetic demands and tradeoffs from foraging opportunities, predation risk, and competition (Rosenzweig, 1974; Craig and Crowder, 2002). Understanding how species select habitats across heterogeneous landscapes provides key information regarding occupancy patterns that contribute to survival and reproductive success (Kramer et al., 1997; McGarigal et al., 2016). Such information could then be used to identify, protect, and restore specific ecologically valuable habitats and corridors (Kramer and Chapman, 1999; Beier et al., 2008; Fraschetti et al., 2009; Zeller et al., 2017).

Resource selection functions (RSFs), defined as a function that produces values that are proportional to the probability of use by an animal (Manly et al., 2007), are a popular method to determine and predict relative habitat selection by animals (e.g., Nielsen et al., 2003; Johnson et al., 2004; Ciarniello et al., 2007). These functions evaluate the relationships between resource use (i.e., the units of area selected by an animal) and the environmental characteristics associated with each unit of area (Boyce et al., 2002). Animal spatial data, from sources such as telemetry, can be incorporated into RSFs to define the relative habitat selection strengths among animal space use and a given set of environmental covariates, such as habitat type, substrate, elevation, or water depth (Boyce and McDonald, 1999). When the true absences are unknown, as generated by presence only data derived from sources such as telemetry approaches, RSFs are implemented within a use/availability framework where known presences (1) are compared with a random sample across 'available' resource units, also known as pseudo-absences or background points (0) (Boyce, 2006; Pearce and Boyce, 2006). Alternative to use/availability (e.g., from telemetry), data from observations collected from survey methods, often without timestamps, are typically referred to as presence-background and are fitted as species distribution models (Fieberg et al., 2018). Using RSFs to derive the relative probability of selection, rather than the absolute probability (see Lele et al., 2013; Avgar et al., 2017), telemetry data are then typically fitted using logistic regression models (Johnson et al., 2006; Manly et al., 2007) or, as of more recently, with machine learning algorithms [e.g., random forest (RF), boosted regression trees] (Shoemaker et al., 2018; Heffelfinger et al., 2020).

While RSFs have been largely applied in terrestrial ecology, such as with wolves (Ordiz et al., 2020), birds (Meager et al., 2012), grizzly bears (McLoughlin et al., 2002), and deer (Godvik et al., 2009), the application of RSFs within aquatic environments has been limited comparatively, likely due to technological challenges related to continuously tracking animals through water (Hussey et al., 2015). Today, passive acoustic telemetry has become one of the most common practices to quantify aquatic animal space use (Cooke et al., 2004; Donaldson et al., 2014; Hussey et al., 2015). This technique involves tagging an animal with an acoustic transmitter that periodically emits an ultrasonic ping with a unique identification number (ID code). When in range and with sufficient detection efficiency the ping is detected by an acoustic receiver that registers both the unique ID code and the time the transmitter was detected (Hussey et al., 2015). Depending on the scope and extent of both research questions and available funding, acoustic receivers are strategically arranged in fixed locations with either non-overlapping detection ranges (Heupel et al., 2006; Brownscombe et al., 2019b), or with overlapping detection ranges that can produce high resolution positioning estimates of space use (Espinoza et al., 2011). While both methods are limited to the available detection coverage (presence only data), the former is often used to examine space use across a given study area at much larger spatial extents (Carlisle et al., 2019) and, thus, is well catered to exploring relative habitat selection.

Although the application of RSFs in combination with acoustic telemetry has been limited (see Freitas et al., 2016; Harrison et al., 2016; Gutowsky et al., 2017; Selby et al., 2019; Griffin et al., 2020), much needed information on animal habitat selection in the marine environment can be derived. For example, Selby et al. (2019) and Griffin et al. (2020) applied RSFs to acoustic telemetry data for hawksbill (*Eretmochelys imbricata*) and juvenile green turtle (*Chelonia mydas*), respectively, from St. Croix, United States Virgin Islands, and determined that the size and extent of a marine protected area (MPA) being used by these sea turtles was sufficient to meet conservation goals. In addition to providing insights on potential drivers of relative habitat selection, RSFs were also extended to predict movements in areas that did not have acoustic receivers to provide potential locations where fine-scale habitat protection may be further prioritized (Griffin et al., 2020). Such results have important implications within marine environments especially for resource managers seeking to incorporate animal movement data to generate effective conservation strategies (Cooke, 2008; Knip et al., 2012; Allen and Singh, 2016; Hays et al., 2016, 2019; Lea et al., 2016). Since management and conservation efforts often rely on spatial management techniques (Peel and Lloyd, 2004; Sequeira et al., 2019), including MPAs (Gell and Roberts, 2003; Lubchenco et al., 2003; Gleason et al., 2010; Lubchenco and Grorud-Colvert, 2015; Weeks et al., 2017; Feeley et al., 2018; Keller et al., 2020; Gallagher et al., 2021), habitat selection predictions should help managers meet conservation endpoints and play a role in evaluating management alternative strategies for both species and for the habitats on which they rely on.

In this study, we provide a framework to implement RSFs using machine learning algorithms to examine and accurately predict relative habitat selection for tracking data collected using acoustic telemetry. Specifically, we apply RSFs to evaluate the relative habitat and resource selection of four

shark species: Caribbean reef (*Carcharhinus perezi*), lemon (*Negaprion brevirostris*), nurse (*Ginglymostoma cirratum*), and tiger (*Galeocerdo cuvier*) sharks within a Caribbean MPA. In the Caribbean Sea, these four species occupy a wide range of environments from nearshore reef and seagrass habitats to offshore pelagic habitats (Pikitch et al., 2005; Legare et al., 2015; Pickard et al., 2016; Casselberry et al., 2020; Gallagher et al., 2021). Considering Caribbean reef sharks are listed as "Endangered" by the IUCN Red List (Carlson et al., 2021a), lemon and nurse sharks are listed as "Vulnerable" (Carlson et al., 2021b,c), and tiger sharks are listed as "Near Threatened" (Ferreira and Simpfendorfer, 2019), all with decreasing population trends, conservation and management efforts would benefit from understanding and incorporating findings surrounding their spatial ecology.

Successful management is especially needed since it has been suggested losses in shark abundance may disrupt food web dynamics that would lead to reduced ecosystem health (Baum and Worm, 2009; Ferretti et al., 2010; Heupel et al., 2014; Hammerschlag et al., 2019). Indeed, food web simulations for Caribbean coral reefs show sharks, as top predators, are members of strongly interacting tri-trophic food chains whose loss could result in trophic cascades (Bascompte et al., 2005). This is supported by *in situ* studies of mesopredator fish populations in Australia that found shark-depleted coral reefs have reduced fish diversity, species abundance, and biomass, with individual species showing changes in diet and body condition when compared to reefs with healthy shark populations (Barley et al., 2017a,b). Information about shark habitat use and selection could lead to proactive management strategies to mitigate non-sustainable or illegal harvest (White et al., 2017; Jacoby et al., 2020) and/or protect and restore important habitat (Speed et al., 2016; Daly et al., 2018). Because spatial management techniques, such as MPAs, can provide protection for multiple species across a variety of life stages, understanding resource selection across species should help to tailor effective conservation strategies and, specifically, ensure adequate coverage of ecologically vital habitats and areas (Lea et al., 2016).

Considering MPA design may benefit from the inclusion and understanding of predator-prey dynamics (Micheli et al., 2004; Cashion et al., 2020), we also demonstrate how RSFs can be extended to examine spatially explicit relationships between marine predators and their prey. This was accomplished by deriving and averaging overlapping selection values from tiger sharks and from their potential prey, including juvenile green turtles, juvenile Caribbean reef sharks, juvenile lemon sharks, great barracuda (*Sphyraena barracuda*), horse-eye jack (*Caranx latus*), yellowtail snapper (*Ocyurus chrysurus*), and mutton snapper (*Lutjanus analis*) (Lowe et al., 1996; Simpfendorfer et al., 2001; O'Shea et al., 2015; Aines et al., 2018; Gallagher et al., 2021). Herein, we provide a framework for studies wishing to investigate animal relative habitat selection and predator-prey relationships with acoustic telemetry in marine environments.

Ultimately, these collective RSF findings provide insights into shark spatial ecology and is useful for the conservation of Caribbean reef, lemon, nurse, and tiger sharks and their habitats. In addition, we have included an R code vignette (**Appendix A**), to improve accessibility and application of RSFs and machine learning.

MATERIALS AND METHODS

Study Area and Field Data Collection

Buck Island Reef National Monument (BIRNM), a 77 km^2 no-take MPA, is located on the northeast shelf of St. Croix, United States Virgin Islands (Pittman et al., 2008). Buck Island is an uninhabited, 0.7 km^2 island that is situated in the middle of the MPA, and 2.5 km northeast of St. Croix. This MPA ranges from shallow-water habitats (<10 m) near the island to deep-water habitats (>1,000 m) off the continental shelf. Generally, benthic habitats range from lagoon habitat (50–150 m wide, around the island excluding the west and southwest sides of the island), linear reef (south side of island and wrapping toward northwest corner), patch reef systems (northwest and north of the island, and south of the southern linear reef), seagrass patches (*Thalassia* sp., *Syringodium* sp., and *Halophila* sp.) and sand flats (south and southwest) (Pittman et al., 2008; Costa et al., 2012).

Between 2011 and 2019, a total of 147 VEMCO VR2W receivers (Innovasea Systems Inc., Nova Scotia, Canada) were deployed as a passive acoustic receiver array within BIRNM to study multiple species (Becker et al., 2016, 2020; Bryan et al., 2019; Selby et al., 2019; Casselberry et al., 2020; Griffin et al., 2020; Novak et al., 2020a,b) (**Supplementary**). Receivers were deployed, in depths ranging from 2 to 40 m, either on sand screws or cement block anchors around the island with receiver downloads occurring twice a year (see Becker et al., 2016; Selby et al., 2016; Casselberry et al., 2020 for mooring details). Among years, the receiver array design changed in extent through the addition of new receiver stations or decommissioning old stations, due to the availability of receivers and evolving project goals, while maintaining a core set of receiver stations through the duration of the project. The array began with 17 receivers in 2011 and reached its greatest coverage with 147 receivers in 2017. For this study, we collected and analyzed acoustic telemetry data from only 2013–2019 when detection coverage and tag deployment was most substantial across BIRNM. Animal tracking data were collected from surgically implanted V13 or V16 transmitters (delay 60–180 s, battery life 360–3,217 days, Innovasea Systems Inc., Nova Scotia, Canada) in 14 Caribbean reef sharks (between 2013 and 2019), 10 lemon sharks (between 2013 and 2019), 11 nurse sharks (between 2013 and 2019), and eight tiger sharks (between 2015 and 2019). In addition, to examine prey habitat selection in relation to tiger shark selection, data were also collected from 58 juvenile green turtles (between 2013 and 2014), 25 great barracuda (hereinafter referred to as barracuda) (between 2014 and 2015), five horse-eye jack (between 2016 and 2017), eight yellowtail snapper (between 2015 and 2017), and four mutton snapper (between 2015 and 2016). All detection data were reviewed and filtered to remove false detections (Simpfendorfer et al., 2015), including detections that occurred within 60 s of each other for a given individual, singular detections occurring within 12 h, and detections that indicated unrealistic movements (>3 m per second). Tagging locations

and methods, including additional specific detection filtering processes, can be found for sharks in Casselberry et al. (2020), barracuda in Becker et al. (2016), horse-eye jack in Novak et al. (2020a), yellowtail snapper in Novak et al. (2020b), and for green turtles in Griffin et al. (2020).

Framework to Apply RSFs Using Acoustic Telemetry and Machine Learning

To derive and predict relative selection values of each species, we describe four important components that include defining available resource units, aggregating habitat information, implementing, evaluating, and interpreting RSFs with machine learning algorithms, and, ultimately, predicting habitat selection for sharks across BIRNM. All analyses were conducted in R version 3.6.2 (R Core Team, 2019). We describe each step in detail below and in the included R code vignette (**Appendix A**).

Defining Available Resource Units and Presence/Background Points

To estimate fine resolution space use away from the exact location of receivers, detection data was first converted into short-term centers of activity (COAs) using the mean position algorithm (Simpfendorfer et al., 2002). To disaggregate detection data from receiver locations, this method, using the detections across multiple receivers, provides position estimates that are based on the weighted means of the number of detections among each receiver during a specified time window (Simpfendorfer et al., 2002). Here, using the VTRACK package (Campbell et al., 2012), this algorithm was implemented with 90-min time bins to provide animal positioning data across BIRNM (Selby et al., 2019; Griffin et al., 2020). In addition to disaggregating data from receiver locations, constructed COAs provide an approach to potentially reduce issues with autocorrelation by subsampling data into defined time steps (e.g., 90-min bins) (Matley et al., 2017). Autocorrelation, an inherent problem with tracking data, occurs when sequential locations are obtained from the same individual and can lead to biased parameter estimates of animal habitat/space use (Legendre, 1993; Johnson et al., 2013; Fleming et al., 2015).

Consistent with Selby et al. (2019) and Griffin et al. (2020), we defined our available resource units by deriving 400 m buffers around each receiver for each year it was deployed (i.e., 2013–2019) (**Supplementary Figure 1**). While detection range was variable across BIRNM habitats with an average of 58.2% (95% confidence interval: 44–73% CI) probability of detection at 100 m distance from a receiver (Selby et al., 2016), we decided to extend the buffer size to 400 m since COAs are able to provide approximate positioning estimates even outside of receiver detection range.

To implement RSFs within a use/availability framework and to account for variable receiver coverage across years, we restricted both the COAs (presences) and the randomly distributed background points (pseudo-absences) to our defined available resource units (400 m receiver buffers at the year level) only. Background points were randomly distributed equal to the number of observed COAs (see Barbet-Massin et al., 2012) per individual, diel period (night vs. day), and year across all available resource units (**Figure 1**). Diel period was calculated using the maptools package (Bivand and Lewin-Koh, 2013). Only using COAs and background points that were within the 400 m buffer from any receiver, they were then collapsed into 200 m × 200 m raster cells.

Aggregating Habitat Information

Using habitat mapping data provided by the National Oceanic and Atmospheric Administration (NOAA) (Costa et al., 2012), we converted available and relevant shapefile data into 200 m × 200 m raster cells using the raster (Hijmans et al., 2015) and the sp (Pebesma et al., 2012) packages. Derived habitat raster files included classifications aggregated by zone (fore reef, reef flat, lagoon, etc.) (**Supplementary Figure 2A**), fine-scale structure (aggregate reef, sand, pavement with sand channels, etc.) (**Supplementary Figure 2B**), fine-scale cover [seagrass patchy (10%–<50%), seagrass patchy (50–<90%), seagrass continuous (90–100%), etc.], broad-scale cover (algae, live coral, seagrass, etc.), and percent coral cover (i.e., 0–<10%, 10–<50%). In addition, we generated two relevant habitat raster files including distance to land (m) (Buck Island) and depth (m) (**Supplementary Figure 3**).

Subsequently, corresponding habitat and depth information were extracted from each raster cell and assigned to each COA and background point using the rasterize function from the raster package (Hijmans et al., 2015) (see **Appendix A**). Habitat information was converted into factors with depth (m) and distance to land (m) remaining as continuous variables.

Applying RSFs With Machine Learning

Resource selection functions were applied using RF models, a commonly utilized machine learning algorithm, to evaluate the relative habitat selection of each species within BIRNM. RF, using binary recursive partitioning to fit multiple data trees with randomly selected predictor subsets (Breiman, 2001), effectively reduces variance and model overfitting while optimizing predictive accuracy (James et al., 2013; Hengl et al., 2018; Schratz et al., 2018). To increase the prediction of the response variable (presences/background points), RF models were fit for each species with 500 trees, replacement, and with 60% of the data. The 40% remainder of each dataset, known as the holdout dataset, were then used to test model performance. RF models were implemented in the ranger (Wright and Ziegler, 2015) and mlr (Bischl et al., 2016) packages.

Characteristic of RF models, a user may tune how trees are generated and fitted from the data. These controls, referred to as hyperparameters, are set prior to fitting an RF by running multiple iterations of values (see Probst et al., 2019). While default values for hyperparameters lead to relatively high performance alone, tuning can often lead to overall model improvement (Lovelace et al., 2019; Probst et al., 2019). Ultimately, hyperparameter settings control the degree of randomness across trees and may include the number of predictors that should be used in each tree (mtry), the fraction of observations to be used in each tree (sample.fraction)

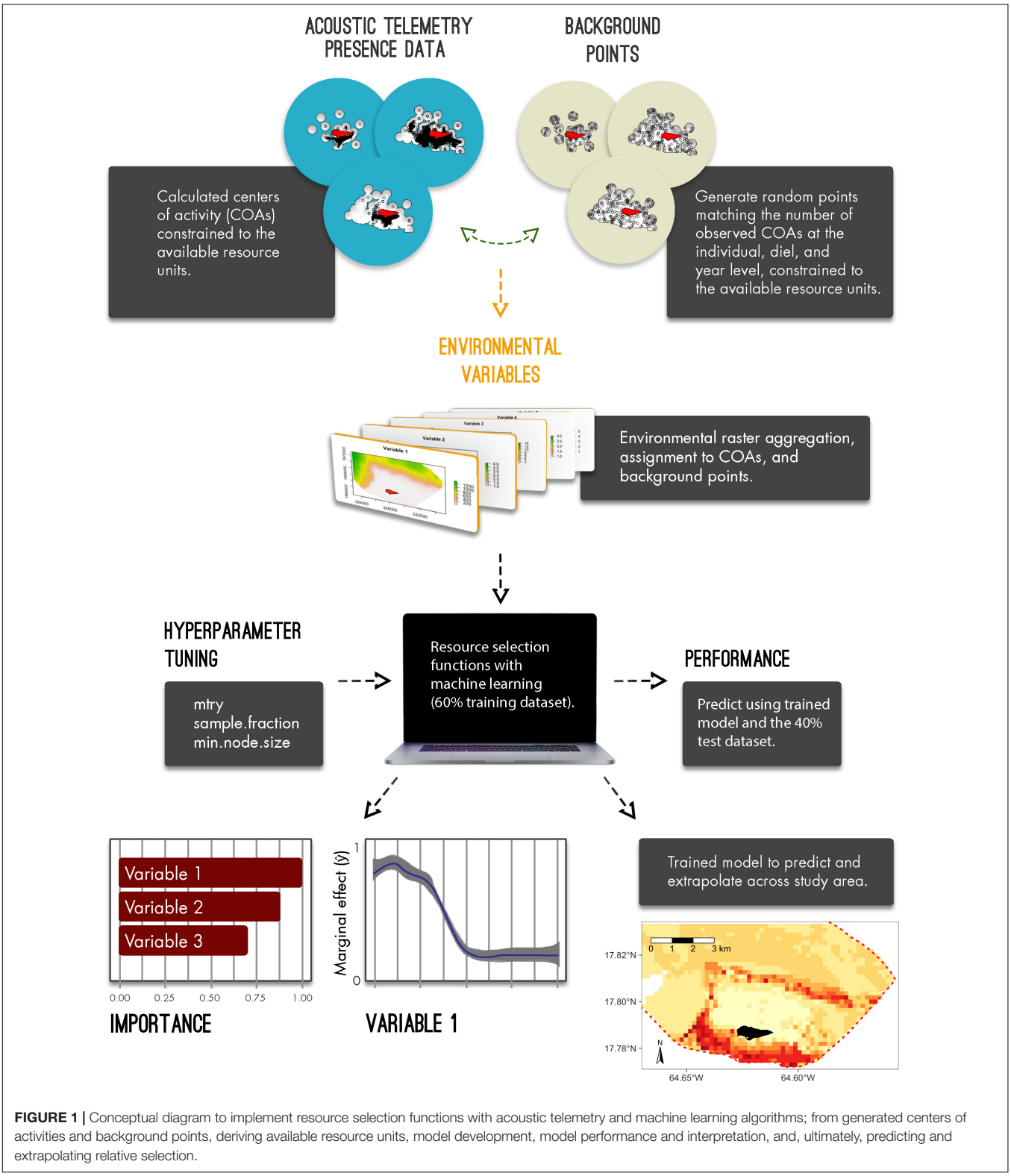

FIGURE 1 | Conceptual diagram to implement resource selection functions with acoustic telemetry and machine learning algorithms; from generated centers of activities and background points, deriving available resource units, model development, model performance and interpretation, and, ultimately, predicting and extrapolating relative selection.

with lower fractions leading to lower correlation across trees, and the number of observations a terminal node (within a tree) should at least have (min.node.size) (Lovelace et al., 2019). To find the optimal hyperparameter values, we first partitioned the training dataset into five distinct geographic sections and then for each partition, we generated 50 random combinations of hyperparameters and subsequently chose the optimal combination (see Lovelace et al., 2019) using the

tuneParams function from the mlr package (Bischl et al., 2016). Thus, 50 iterations of random hyperparameter values across each of the five partitions resulted in 250 models in total. Subsequently, the optimal hyperparameter combinations were used to tune and train the final RF for each species, using the 60% training dataset.

Performance, Interpretation, and Prediction

With the trained model, we predicted across the 40% holdout dataset to evaluate performance. The functions calculateConfusionMatrix and calculateROCMeasures from the mlr package (Bischl et al., 2016) were used to determine overall and class (present versus background point) accuracy, error rate, and performance of each RF. Specifically, performance measures were derived from the confusion matrix table that compared the true observations versus the model predictions, these metrics included overall accuracy, i.e., correct number of predictions/total number of predictions, sensitivity (true positive rate), specificity (true negative rate), fall-out (false positive rate), miss rate (false negative rate), and precision (positive predictive value). Predictor, also known as feature, importance was also assessed using the permutation importance method (mean decrease in accuracy) where predictors were evaluated based on the increase or decrease in prediction error after permutation (Breiman, 2001). For interpretation across RF models, all importance values were normalized (min-max normalization). To identify which variables generated the greatest two-way interaction strengths, we derived the H-statistic (Friedman and Popescu, 2008) using the Interaction function from the iml package (Molnar et al., 2018). This calculation, which can be extremely computationally intensive when examining every possible interaction, was implemented after rerunning each RF model but with only 25% of the training dataset. After identifying the top three variables that led to the greatest interaction strengths, this function was used again to assess those other variables with which each top variable interacted. However, this was performed using the original model, with the entire training dataset, since computation times were greatly reduced when examining single two-way interactions as opposed to each possible interaction.

To assess the marginal effect of covariates on the predicted outcome ($\hat{y}$), i.e., predictor probabilities for each RF, we constructed partial dependency plots, using the pdp package (Greenwell, 2017), for the most important feature as identified by the mean decrease in accuracy approach. Partial dependency plots were also generated for the top three two-way interactions as identified from the calculated H-statistic values. Discrete and continuous predictors were shown with 95% confidence intervals. To visualize marginal effect variation within each continuous predictor, we used a generalized additive model smoother via the ggplot2 package (Wickham, 2011). All partial dependency plots were restricted to depths of 50 m or less so to avoid extrapolating outside the shelf of BIRNM where no receivers were located. Finally, the trained RFs were then used to predict relative habitat selection at the species level in BIRNM. Model extrapolation across the MPA was constrained to the maximum depth observed for the given species based on acoustic detections.

RSFs and Predator-Prey Relationships

Resource selection functions were also extended to explore selection overlap values between large juvenile and mature tiger sharks (n = 8, >200 cm FL) and their potential prey species, including juvenile green turtles, juvenile Caribbean reef sharks (n = 12, <120 cm FL), juvenile lemon sharks (8, <120 cm FL), barracuda, horse-eye jacks, yellowtail snapper, and mutton snapper. First, relative habitat selection values were calculated and extrapolated across BIRNM for each potential prey species following the steps outlined in Sections "Defining Available Resource Units and Presence/Background Points," "Aggregating Habitat Information," "Applying RSFs With Machine Learning," and "Performance, Interpretation, and Prediction." Second, to explore areas of potential predator-prey overlap, relative habitat selection values across BIRNM were averaged between tiger sharks and each potential prey species. Lastly, by removing raster cells where relative selection values of potential prey were <0.5, we examined specific high overlap areas between tiger sharks and each potential prey species.

Kernel Density Estimates

To compare the predicted relative habitat selection values to observed animal space use, we fit kernel utilization distributions to the COAs at the species level. Each species' kernel utilization distribution, representing a bivariate probability density function of animal use (Worton, 1989; Lichti and Swihart, 2011), was then used to extract the 50 and 95% kernel density estimates to produce space use estimates. Kernel utilization distributions and subsequent kernel density estimates were constructed using the adehabitatHR package (Calenge, 2006) with 200 m smoothing parameters. While species level kernel density estimates were plotted along with all predicted relative habitat selection values, it should be noted these estimates were used for broad comparison since they are likely biased to some extent due to unequal sample sizes across individuals.

Additional Methodological Considerations

To explore model sensitivity to varying parameter inputs, we also implemented RF models using COA data binned at 60-min timesteps. These models and their outputs were derived from using the same procedures as outlined in Sections "Defining Available Resource Units and Presence/Background Points," "Aggregating Habitat Information," "Applying RSFs With Machine Learning," and "Performance, Interpretation, and Prediction." To examine how models performed under different available habitat extents, we again ran RF models but with available habitat defined using either 200 or 600 m buffers. Hyperparameter inputs were kept consistent with original respective models. Further, to avoid extrapolating predictive models beyond the range of our measured data (Mesgaran et al., 2014), we explored and mapped extrapolation reliability in BIRNM using the dsmextra package (Bouchet et al., 2020). Using the presence/background locations and their depth (m) and distance to land (m) values, the compute_extrapolation function evaluated areas across BIRNM

that fell within the sampled covariate space (Mesgaran et al., 2014; Bouchet et al., 2019, 2020). This multivariate statistical tool highlighted areas of univariate extrapolation (when predictions are considered outside the range of covariates), combinational extrapolation (predications made within range of covariates but in novel combinations), and areas of geographical interpolation [predications made within our range of covariates and in analogous conditional space (see Mesgaran et al., 2014)]. Subsequently, using the map_extrapolation function (Bouchet et al., 2020), we visually assessed extrapolation reliability within BIRNM (see **Appendix A**).

Lastly, we examined individual shark space use variation by constructing Brownian bridge movement models (BBMMs) (Horne et al., 2007) and by implementing RSFs using generalized linear mixed models (GLMMs) with individual as the random effect. BBMMs incorporate movement paths into the modeling process (Horne et al., 2007) and has been recommended when evaluating individual space use with COAs since it can better account for temporal autocorrelation (Udyawer et al., 2018). Here, using all available detection data, we constructed individual BBMMs and plotted individual space use within BIRNM.

While BBMMs highlight variation in space use across individuals, RF models are currently unable to easily incorporate random effects to account for individual level effects, thus, model outputs are potentially biased to some extent. Alternatively, RSFs used in-combination with mixed effect models, can include a random effect for individual ID to explicitly account for individual variability and, in turn, provide measures of inference for the entire population (Gillies et al., 2006; Aarts et al., 2008; Hebblewhite and Merrill, 2008). Here, we implemented RSF GLMMs with individual ID as the random effect for each shark species using the top three most important variables as fixed effects that were identified by RF models. All variables were examined for correlation issues using variance inflation scores and continuous variables were standardized. As a simplified approach and for the purpose of examining the relative contribution of individual ID on each model, no temporal or spatial autocorrelation dependency structures, interaction terms, or non-linear relationships were included. All models were implemented and assessed using 60% of the dataset via the glmmTMB (Magnusson et al., 2017) and performance (Lüdecke et al., 2019) packages. Two goodness-of-fit metrics, marginal R^2 and conditional R^2, were calculated for each model. While marginal R^2 evaluates the variance explained by fixed effects, conditional R^2 evaluates the variance explained by both fixed and random effects, allowing us to assess the relative contribution of the random effect on each model (Nakagawa and Schielzeth, 2013). In addition to both goodness-of-fit metrics, we also used the 40% holdout datasets to test GLMM performance, as was done with the RF models above.

RESULTS

Using the converted COA tracking data (**Table 1**), RF model accuracy and model performance varied across species with overall accuracy ranging from 80 to 95% and sensitivity (true

TABLE 1 | Tagging and detection data for shark species and the potential prey species for tiger sharks, including a subset of juvenile Caribbean reef and lemon sharks, monitored within Buck Island National Monument.

Species	Sci. name	ID count	Avg. size (cm)	*SD* size (cm)	Total COA	Avg. COA	*SD* COA	Avg. tracking. dur. (d)	*SD* tracking dur. (d)
Caribbean Reef	*Carcharhinus perezi*	14	97.79	24.90	115475	8248.21	8665.22	991.43	822.96
Lemon	*Negaprion brevirostris*	10	81.30	35.58	60509	6050.90	10680.64	592.20	799.34
Nurse	*Ginglymostoma cirratum*	11	128.45	30.77	48920	4447.27	5223.02	541.82	537.57
Tiger	*Galeocerdo cuvier*	8	231.50	26.69	19120	2390.00	2280.74	720.62	610.98
Caribbean reef (juv.)	*Carcharhinus perezi*	12	90.58	18.30	91055	7587.92	8728.62	906.17	860.23
Great barracuda	*Sphyraena barracuda*	19	83.74	10.51	17041	896.89	1236.71	342.74	113.77
Green turtle	*Chelonia mydas*	36	48.40	7.86	45787	1271.86	1180.64	156.86	117.22
Horse-eye jack	*Caranx latus*	5	57.20	8.98	13708	2741.60	2049.30	335.00	180.69
Lemon (juv.)	*Negaprion brevirostris*	8	67.38	22.03	27167	3395.88	9338.61	383.00	735.18
Mutton snapper	*Lutjanus analis*	4	52.62	7.48	7310	1827.50	1373.26	153.75	102.25
Yellowt ail Snapper	*Ocyurus chrysurus*	8	29.44	4.25	15363	1920.38	1069.84	443.62	166.56

TABLE 2 | Confusion matrix performance metrics derived from using the trained random forest model to predict across the 40% holdout dataset.

Species	Accuracy	Sensitivity (true positive rate)	Specificity (true negative rate)	Fall-out (false positive rate)	Miss rate (false negative rate)	Precision (positive predictive value)
Caribbean reef	0.90	0.90	0.89	0.11	0.10	0.89
Lemon	0.94	0.97	0.91	0.09	0.03	0.92
Nurse	0.85	0.91	0.79	0.21	0.09	0.81
Tiger	0.83	0.91	0.76	0.24	0.09	0.79
Caribbean ref (juv.)	0.89	0.90	0.89	0.11	0.10	0.89
Great brracuda	0.87	0.85	0.89	0.11	0.15	0.89
Green turtle	0.95	0.97	0.93	0.07	0.03	0.94
Horse-eye jack	0.80	0.89	0.72	0.28	0.11	0.76
Lemon (juv.)	0.94	0.97	0.91	0.09	0.03	0.91
Mutton snapper	0.88	0.98	0.78	0.22	0.02	0.81
Yellowtail snapper	0.92	0.92	0.93	0.07	0.08	0.93

Accuracy (i.e., correct number of predictions/total number of predictions), sensitivity (i.e., true positive rate), specificity (i.e., true negative rate), fall-out (i.e., false positive rate), miss rate (i.e., false negative rate), and precision (i.e., positive predictive value) were reported for shark species and the potential prey species for tiger sharks, including a subset of juvenile Caribbean reef and lemon sharks, monitored within Buck Island National Monument.

positive rate) from 85 to 98% (**Table 2**). Predictor importance and rank varied across shark species, with depth (m) as either the most important or within the top two most important predictors for all four species (**Figure 2**). Overall, $\hat{y}$ values generally decreased as depth increased for all sharks (**Figure 3**). While $\hat{y}$ values decreased for Caribbean reef, nurse, and tiger sharks in areas >3 km from Buck Island, tiger sharks appeared to have higher $\hat{y}$ values farther away from the island at distances approximately between 500 and 2,000 m. Lemon shark $\hat{y}$ values decreased rapidly as distance from land increased, with lowest values occurring >1,000 m.

Caribbean reef sharks were more likely to select for coral or coral-containing habitats with higher $\hat{y}$ values observed within coral habitats (sand with scattered coral and rock, aggregate reef, and aggregated patch reefs) (**Figure 3A**). Caribbean reef shark two-way predictor interactions highlighted relatively high $\hat{y}$ values in depths of 20–30 m, areas <2 km away from land, and in areas of sand with scattered coral and rock (**Figure 4A**). Lemon sharks were more likely to select for shallow areas directly adjacent to land, specifically in shallow (0–5 m) habitats classified as channel, lagoon, and reef crest (**Figures 3B**, **4B**). While nurse sharks followed a similar pattern, $\hat{y}$ values indicated they were more likely to select for habitats between 0 and 2,000 m away from land but within <15 and 25–30 m of depth. In addition, $\hat{y}$ values were higher in areas of sand with scattered coral and rock located within bank/shelf, bank/shelf escarpment, fore reef, and reef crest zones (**Figures 3C**, **4C**). Lastly, tiger sharks exhibited the greatest $\hat{y}$ values away from land (500–2,000 m), in <30 m depth, and in aggregate reef, sand, sand with scattered coral and rock, and pavement habitats (**Figures 3D**, **4D**).

Extrapolated relative habitat selection values across BIRNM, as computed from the trained RF models, followed similar patterns to kernel density estimates (**Figure 5**). Specifically, 50% kernel density estimates largely overlapped with extrapolated areas of high relative selection. However, for Caribbean reef, nurse, and tiger sharks, levels of high extrapolated relative selection also extended beyond 50% kernel density estimates along the western shelf and the eastern side of BIRNM, where receiver coverage was limited. While Caribbean reef, lemon, and tiger sharks exhibited more targeted habitat selection with greater affinities to specific areas and habitats, nurse sharks exhibited a more generalist approach to relative habitat selection across BIRNM (**Figure 5**). Caribbean reef sharks showed strong affinity to habitats with reef-containing structure, including areas of linear reef around the island and in the aggregated patch reef system that is characteristic north of the island (**Figure 5A**). In addition, Caribbean reef sharks exhibited higher relative selection values along the western shelf near adjacent deep water habitats (>50 m). Alternatively, lemon shark relative habitat selection values were tightly located around the island, within the reef sheltered lagoon, with lower values along the southwest side of the island where less lagoon and structure habitat exist (**Figure 5B**). Nurse shark extrapolated relative selection values were wide ranging with the densest cluster of higher values surrounding the island (reef habitats), to the southwest of the island along the bank (sand and seagrass habitats), and to the far eastern side of BIRNM (reef, pavement, and sand habitats) (**Figure 5D**). Lastly, similar to nurse sharks, tiger sharks primarily have highest selection values extrapolated south of the island along banks containing both seagrass and sand habitats, leading to the continental shelf break in the west. Relative selection values were also expected to be high along the western shelf and in some locations around the north/northwest shelf. While low relative selection values were expected for tiger sharks in the network of highly rugose patch reefs north of the island, higher selections values existed on eastern side of BIRNM habitats containing mainly reef, pavement, and sand (**Figure 5D**).

Predator-Prey Relationships

Resource selection functions were also extended to examine potential areas of relative selection overlap between tiger sharks and their potential prey sources. Depending on the species, overlap selections varied in location and intensity. For example, juvenile green turtles (**Figure 6A**) and tiger sharks were most likely to overlap in selection south of the island where seagrass beds along bank habitats were most abundant. For juvenile

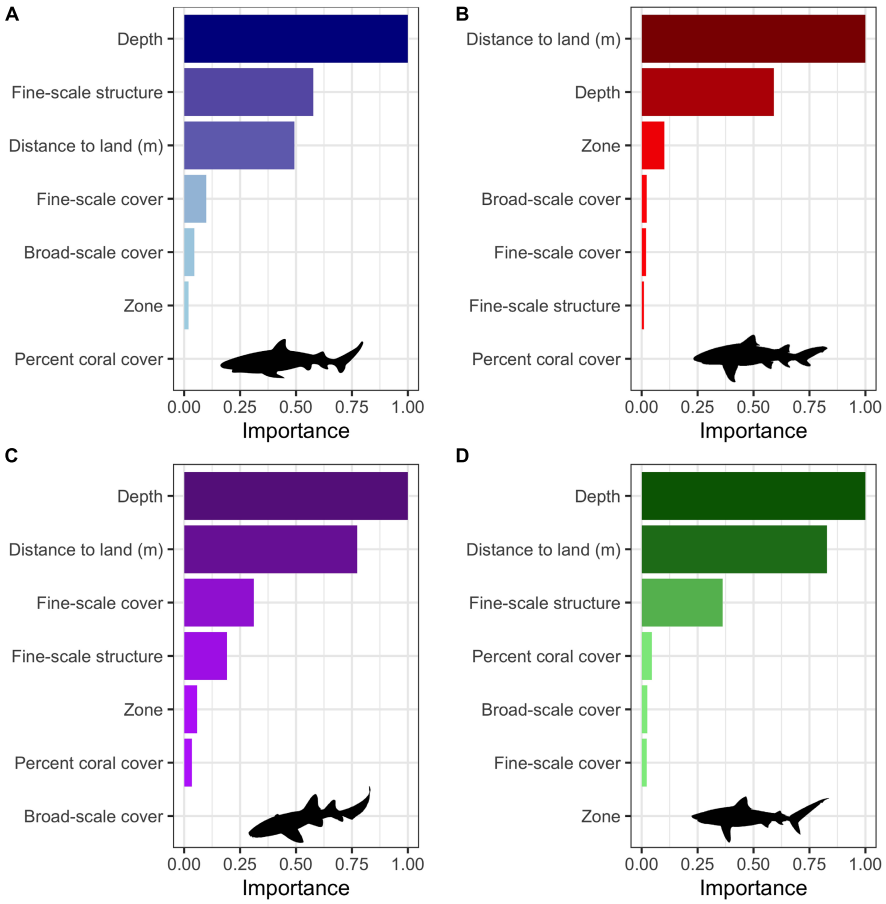

FIGURE 2 | Feature importance, indicating how important each predictor variable is within each shark random forest model for **(A)** Caribbean reef sharks, **(B)** lemon sharks, **(C)** nurse sharks, and **(D)** tiger sharks. Higher values and darker colors indicate greater relative importance. Values were calculated using the mean decrease in accuracy method and, subsequently, all importance values were normalized from 0 to 1 for comparison across species.

Caribbean reef sharks and tiger sharks, averaged overlap selection values were greatest along the western shelf and in the southwest portion of BIRNM (**Figure 6B**). Juvenile Caribbean reef sharks also had high relative selection values north of the island, averaged overlap was comparatively lower in this area to the western shelf due to reduced tiger shark selection values (**Figure 5D**). The averaged overlap selection values between barracuda and tiger sharks followed a similar pattern with higher barracuda relative selection values north of the island but averages reduced due to lower tiger shark relative selection values (**Figure 6C**). Horse-eye jacks, with the most similar relative selection values of tiger sharks, had the greatest averaged overlap values along the western shelf, south of the island, and in the southeastern portion of BIRNM (**Figure 6D**). When tiger shark relative selection values were averaged across the other three species, including juvenile lemon sharks, mutton snapper, and yellowtail snapper, they followed similar patterns with higher averaged relative selection overlap values where the potential prey species had higher selection values unless it was directly north of the island where patch reef systems exist (**Supplementary Figure 4**).

Additional Methodological Considerations

The RF models using COA data of 60-min bins produced similar results to models that used COA data of 90-min bins (**Appendix B**). The top two most important variables remained unchanged for all shark species (**Appendix Figure B1**) and pdps and associated $\hat{y}$ values only changed slightly (**Appendix Figure B2**). Most notably, the new 60-min binned RF models indicated $\hat{y}$ values generally increased (rather than decreased) with depth for Caribbean reef sharks and $\hat{y}$ values for aggregate reef were lower for tiger sharks (**Appendix Figures B2, B3**). However, for Caribbean reef sharks, $\hat{y}$ values related to depth remained similar across the interaction of depth and distance to land.

While accuracy metrics were similar for all species across the 60- and 90-min RF models (∼1–2% differences), some varied substantially (e.g., 5–7%) (**Table 2** and **Appendix Table B1**). The use of 60-min time bins led to a decrease in overall accuracy for nurse sharks (85–80%) but an increase in overall accuracy for barracuda (87–92%), horse-eye jack (80–90%), and mutton snapper (88–96%). Subsequently, model predictions and extrapolations within BIRNM reflected these discrepancies

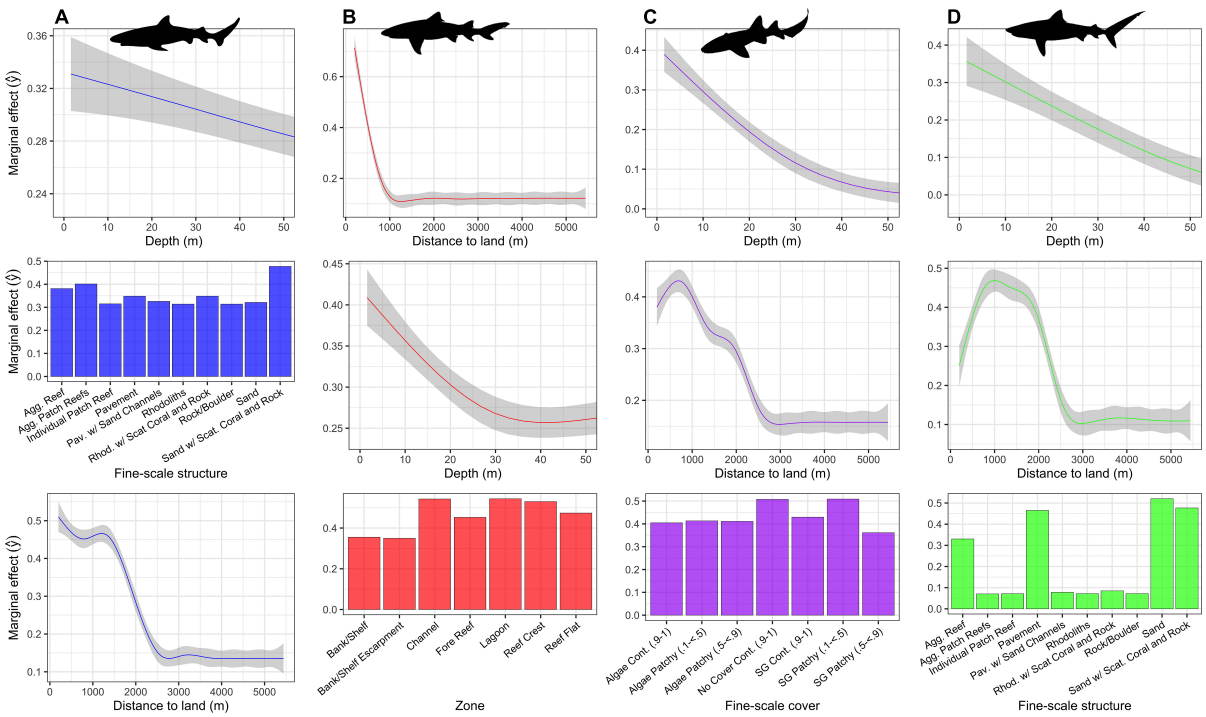

FIGURE 3 | Random forest partial dependency (marginal effects, ŷ) plots for the top three most important (ordered from top to bottom) predictors for **(A)** Caribbean reef sharks, **(B)** lemon sharks, **(C)** nurse sharks, and **(D)** tiger sharks. The marginal effects indicate the relative selection strength for the given variable with greater ŷ values indicating high relative selection probabilities and lesser values indicating low relative selection probabilities. Discrete predictors are displayed with bar plots and continuous predictors are displayed using a generalized additive model smoother with 95% confidence intervals. Partial dependency plots involving depth were restricted to 50 m depth for each species.

with lower accuracy scores producing more homogenous and generalized relative habitat selection patterns than when models produced higher predictive accuracies (**Figures 5**, **6** and **Appendix Figures B4, B5**).

RF models using 200 m buffers for available habitat construction scored lower accuracy measurements and also predicted higher relative selection homogeneously across BIRNM (**Supplementary Figure 5** and **Supplementary Table 1**). Alternatively, models using 600 m buffers for available habitat construction produced similar accuracy measures and predictions across BIRNM as compared to the original models (**Supplementary Figure 5** and **Supplementary Table 1**). Interestingly, relative selection predictions for tiger sharks were higher along the northeastern shelf edge (**Supplementary Figure 5**) than in the original model (**Figure 5**), matching Casselberry et al. (2020) findings.

When assessing extrapolation reliability across BIRNM, extrapolation space became unreliable (univariate extrapolation) in areas off the shelf in deeper and further areas from land (**Appendix A**). However, areas within the MPA that remained in shallower water (<50 m) were analogous to the range of our covariates as measured by depth (m) and distance to land (m), confirming our approach to limiting extrapolations to the maximum observed depth was warranted.

While BBMMs highlighted individual level variation in space use across BIRNM (see examples, **Supplementary Figure 6**), GLMMs and respective marginal and conditional R^2 values indicated variance was largely explained by the fixed effects (marginal R^2) alone (**Supplementary Table 2**). However, the GLMM involving lemon sharks appeared to have substantial variance explained by both the fixed and random effects combined (conditional R^2), suggesting individual variation may be higher within this species dataset. Interestingly, GLMM accuracy for lemon sharks was also nearly as accurate as the RF model (92% versus 94%, respectively). Accuracy metrics for the other GLMMs were substantially lower than respective RF models (**Table 2** and **Supplementary Table 2**).

DISCUSSION

The Approach and Ecological Implications

Using acoustic telemetry data for four shark species, we demonstrate the utility of RSFs with machine learning to accurately predict and understand complex environmental drivers of marine species. Across species, we found variable patterns of relative habitat selection within the MPA, ranging from habitat specialists to generalists. Overall, as depth increased, relative selection decreased for all shark species. While relative selection probabilities for Caribbean reef, lemon, and nurse sharks decreased as distances from land increased, tiger sharks

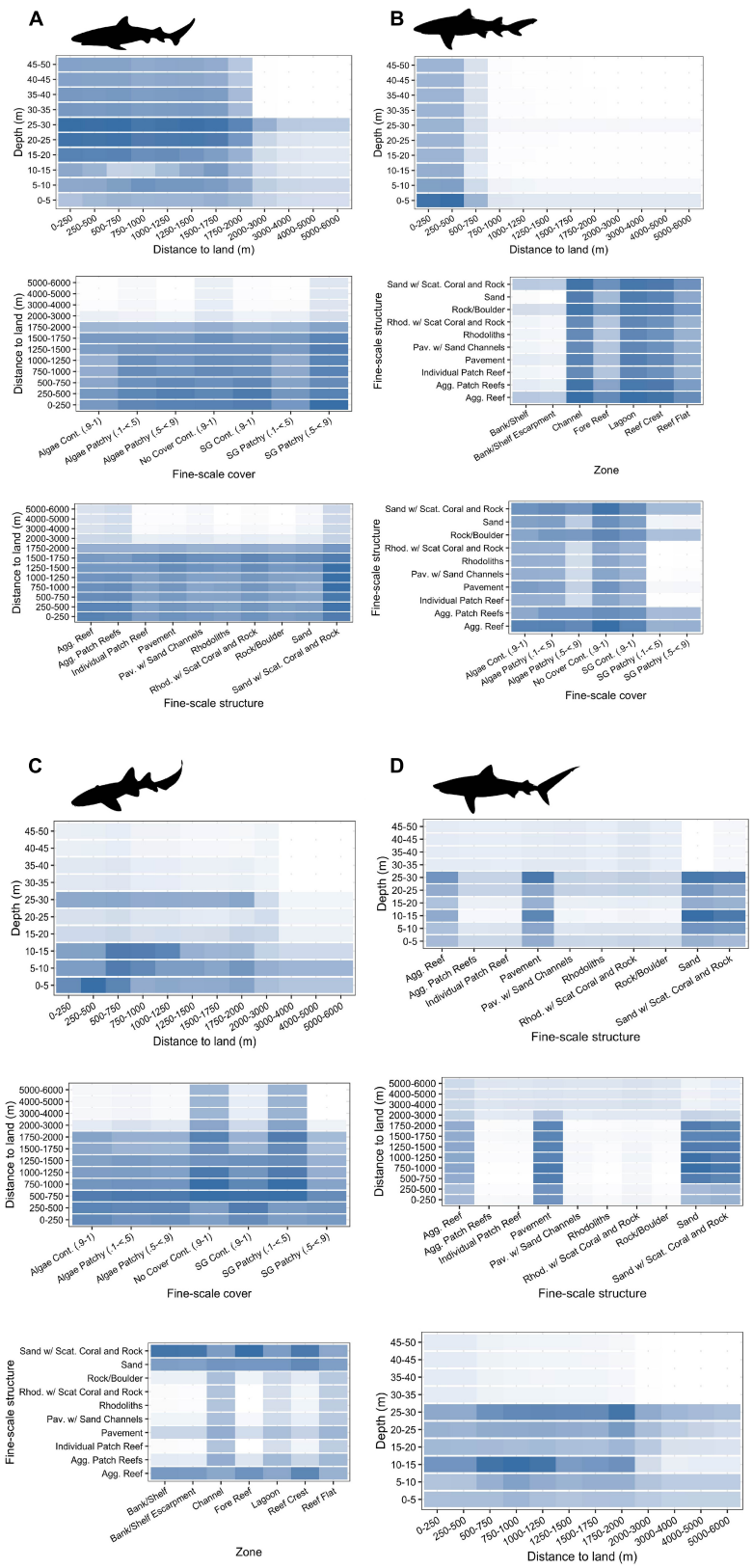

FIGURE 4 | Top three two-way interactions (ordered from top to bottom) displayed and extracted from random forest models for **(A)** Caribbean reef sharks, **(B)** lemon sharks, **(C)** nurse sharks, and **(D)** tiger sharks. Mean marginal effects (ŷ) are shown in each two-way interaction partial dependency plot with colors indicating a continuum from high (blue) to low (white) probabilities of relative selection. Partial dependency plots involving depth were restricted to 50 m depth for each species.

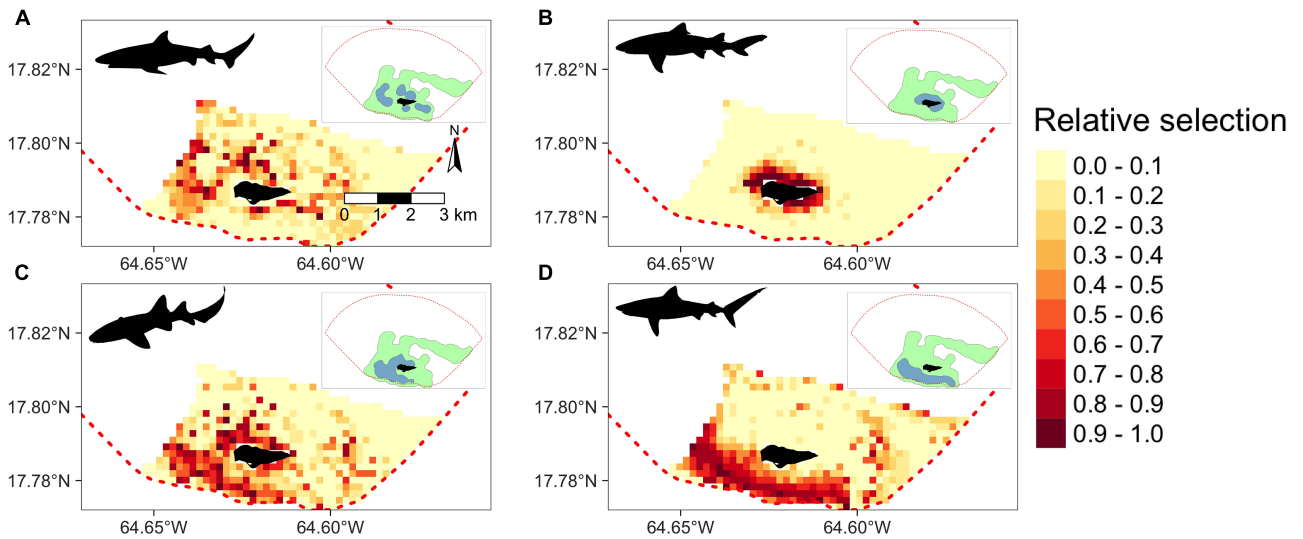

FIGURE 5 | Derived from random forest models, the predicted and extrapolated probability of relative selection across Buck Island National Reef Monument (designated by red dashed line) for **(A)** Caribbean reef sharks, **(B)** lemon sharks, **(C)** nurse sharks, and **(D)** tiger sharks. Higher and darker values indicate higher relative selection values. Model extrapolation across the study areas restricted to the maximum observed depth for each species. 50% (green) and 95% (blue) kernel density estimates shown in the upper right corner of each plot.

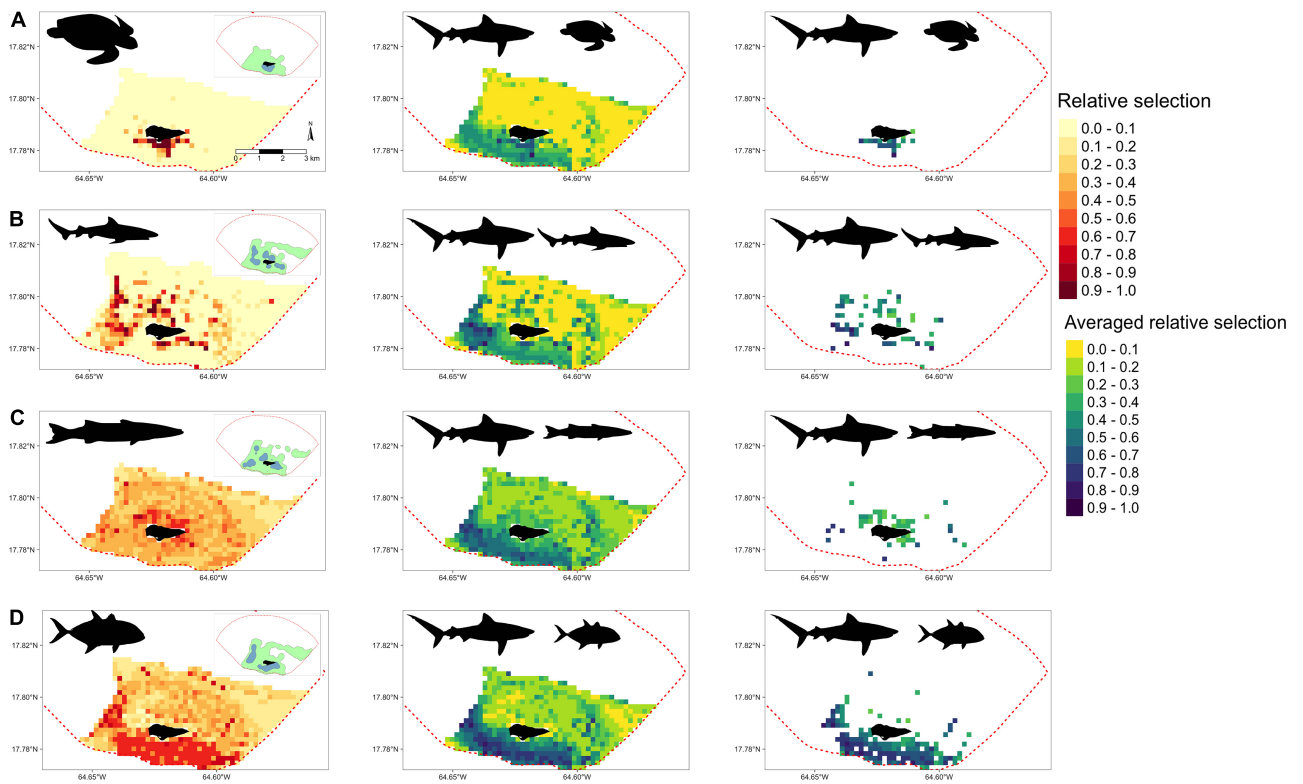

FIGURE 6 | (Left) Derived from random forest models, the predicted and extrapolated probability of relative selection across Buck Island National Reef Monument (designated by red dashed line) for **(A)** juvenile green turtles **(B)** juvenile Caribbean reef sharks, **(C)** great barracuda, and **(D)** horse-eye jacks. 50% (green) and 95% (blue) kernel density estimates shown in the upper right corner of each plot. (Middle) Relative selection values averaged between tiger sharks and each corresponding species on the left. (Right) Relative selection values averaged between tiger sharks and each corresponding species on the left after the removal of cells where potential prey relative selection values were <.5. In all plots, higher and darker values indicate higher relative selection values or their averages.

showed highest affinities for areas between 500 and 2,000 m away from land. Top interactions along with predicted relative selection values highlighted the differences and preferences across species in terms of habitat types, structures, and depths. Finally, using the relative selection values of tiger sharks and their potential prey, we highlight the ability of this framework to generate multiple species selection values that could provide insights into predator-prey relationships when averaged and overlaid with one another.

The results for shark habitat selection presented herein are largely consistent with those presented by Casselberry et al. (2020) that used GLMMs and detection data from fixed receiver locations to model presence in BIRNM habitats, but with improved model accuracy (83–94%) and additional covariates. While GLMMs in Casselberry et al. (2020) were limited by unequal receiver distribution, requiring aggregation across habitat types, the analyses presented here (use/availability framework with COAs and background points) were able to sample across multiple habitats and at finer scales. Further, while GLMMs were limited to a single generalized habitat covariate (factor levels including: unconsolidated sediments, submerged vegetation, and coral, rock, and colonized hardbottom) and depth, RSFs paired with machine learning algorithms were able to easily assess five separate habitat covariates that ranged from two and 10 factor levels each along with two additional continuous predictors. Ultimately, RSFs confirmed use of shallow water habitats near land for lemon sharks and the use of sand and coral associated habitats at mid-depths for nurse sharks, while highlighting tiger sharks' affinity for the continental shelf break and southern sand and seagrass beds. The added analytical flexibility of RSFs and machine learning greatly improved predictions of habitat use for Caribbean reef sharks, whose space use changes dramatically with age across BIRNM's varied landscape (Casselberry et al., 2020). Previous models showed low probability of presence in the acoustic array across habitats and depths compared to the other three shark species, while RSFs highlight specialized use of multiple highly rugose reef habitats at mid-depths. However, GLMMs produced in Casselberry et al. (2020) and the RSF models produced here differed when predicting tiger shark depth preference. Casselberry et al. (2020) showed probability of presence in the acoustic array increasing with depth across habitat types (coral, rock, and colonized hardbottom, submerged vegetation, and unconsolidated sediments), while $\hat{y}$ values consistently decreased with depths beyond 30 m in RSF models.

Examining the tiger shark partial dependency plots reveals high interactions between depth and distance to land at depths between 10–15 and 25–30 m (**Figure 4**). These same depth bins (10–15 and 25–30 m) also had higher $\hat{y}$ values when combined with aggregate reef, pavement, sand, and sand with scattered coral and rock habitats (**Figure 4**). These habitat types and distance to land achieved higher $\hat{y}$ values alone than depth in tiger shark models indicating that these variables have a stronger influence on tiger shark habitat use (**Figure 3**). However, tiger sharks are known to use depths greater than 50 m in and around BIRNM that are beyond the depths of acoustic array coverage (Casselberry unpublished data). This, again, highlighted the need to assess extrapolation reliability (Mesgaran et al., 2014) of RSFs prior to model interpretation since they may have limited ability to extrapolate outside of observed conditions of a given array, for example in areas of BIRNM where depths were greater than 50 m.

When predictions were made within our range of covariates in analogous conditional space (less than 50 m depth), the application of RSFs, as opposed to more traditional use of COAs alone, kernel density estimates, or network analyses, highlighted potentially favorable habitats in BIRNM with limited receiver coverage. The eastern portion of BIRNM has had limited acoustic receiver coverage in part because of the complexity of the coral reef structure in the area. Receiver moorings were not established there in order to avoid damaging the protected reef structure. The RSFs show that favorable habitats exist in this low coverage region for nurse, Caribbean reef, and tiger sharks, particularly at intersections between reef, pavement, and sand habitats. This further highlights the suitability of this MPA for shark conservation and management in St. Croix (**Figure 5**; Casselberry et al., 2020).

Examining overlapping RSFs between tiger sharks and their potential prey highlights regions of potential foraging success for sharks, high predation vulnerability for prey, and areas of ecological importance for managers. Areas of high tiger shark-prey overlap coincide mainly with the seagrass beds south of Buck Island and the western continental shelf break, while many potential prey species also have high selection potential in areas north of Buck Island. This could be a reflection of tiger sharks selectively using areas with higher potential for foraging success (Heithaus et al., 2002). Areas north of the Buck Island are occupied by highly complex coral reef habitats, offering ample areas to refuge or escape from predators (Hixon and Beets, 1993), while habitats south and west of Buck Island are more open at depths of ∼12 m. These waters could be more maneuverable for large juvenile and adult tiger sharks when compared to more structurally complex environments (Fu et al., 2016), perhaps with an increased possibility of foraging success (Heithaus and Dill, 2002; Heithaus et al., 2007; Wirsing et al., 2007). Alternatively, these areas could be a reflection of similar habitat preferences and ecologies among apex and mesopredators in a tropical reef system (Ledee et al., 2016; Heupel et al., 2019). Regardless, areas of high averaged relative selection highlight important regions in BRINM that could be used to inform future habitat monitoring or restoration studies, particularly with the potential for habitat degradation as the climate changes (Graham et al., 2020; Hastings et al., 2020).

As technological tools continue to advance our ability to monitor aquatic animal space use, ecologists are beginning to answer some of the most pressing questions to help direct and prioritize resource management and conservation strategies. Habitat destruction remains on the forefront of decreases in biodiversity, from climate change (Pratchett et al., 2011; Descombes et al., 2015) to destructive landscape use (Rothschild et al., 1994; Coverdale et al., 2013). With calls to protect at least 30% of the ocean by 2030 through establishing MPAs (O'Leary et al., 2016; Sala et al., 2018), an accurate understanding of how marine animals use space and select habitats is increasingly imperative for well informed and effective

marine spatial planning (Foley et al., 2010; Ogburn et al., 2017; Lowerre-Barbieri et al., 2019; Gallagher et al., 2021; Roberts et al., 2021). The RSF modeling framework provided here can produce high accuracy models of relative habitat selection for multiple species of differing ecologies and can be averaged across species to highlight overlapping potential space use or selection. These models can then be used to extrapolate to areas lacking acoustic receiver coverage, as long as within the original measured parameters, accounting for a common issue in acoustic telemetry with incomplete coverage of the study site due to logistic or budgetary limitations. Assuming a sufficient number of individuals are tagged for a given species and age class, the outputs of these models can produce easily interpretable maps for highlighting regions of importance and communicating results to stakeholders, which could result in greater acceptance of study findings given committed stakeholder engagement (Nguyen et al., 2019).

Benefits, Challenges, and Considerations

As technological advancements (e.g., from remote sensing to acoustic telemetry data) allow for high-resolution datasets, machine learning approaches have become increasingly adopted by ecologists because of their ability to handle large datasets and complex non-linear hierarchical relationships and statistical assumptions that are typically violated by conventional parametric approaches, e.g., multiple correlated predictors (Olden et al., 2008; Peters et al., 2014; Durden et al., 2017; Brownscombe et al., 2020). While RSFs have typically been applied within a classical statistical framework (e.g., logistic and linear models) (Johnson et al., 2006; Manly et al., 2007), machine learning does not require non-linear predictor relationships and their interactions to be specified prior to implementing. Thus, allowing for a flexible, realistic, and accessible application when applying RSFs to animal space use in relation to multiple and complex environmental gradients across a landscape (Shoemaker et al., 2018). Further, implementing machine learning with ecological data can also provide highly accurate predictive models (Cutler et al., 2007; Elith et al., 2008; Olden et al., 2008). For instance, Shoemaker et al. (2018) applying RSFs with mule deer (*Odocoileus hemionus*) telemetry data demonstrated machine learning algorithms outperformed the traditional approach of logistic regression with higher prediction accuracy. In another example, although not directly comparable, when implementing a RF using the juvenile green turtle data in this study, we found a higher accuracy compared to that as reported by Griffin et al. (2020) (0.95 versus 0.77, respectively), who also applied RSFs on juvenile green turtle acoustic telemetry data from BIRNM but were fitted with GLMMs and fewer predictor variables.

While machine learning algorithms offer some advantages as an accurate non-parametric technique, the difficulty to account for spatial-temporal autocorrelation and individual level effects presents additional challenges. Whereas RF models are unable to easily incorporate, RSF GLMMs can explicitly include individual ID as a random effect (Gillies et al., 2006). Further, generalized models can incorporate autocorrelation dependency structures (Zuur et al., 2017; Winton et al., 2018a; Griffin et al., 2019; Gutowsky et al., 2020), however, it is worth noting that defining the correct correlation structure still remains challenging within a use/availability (presences/pseudo-absences) sampling design (see Koper and Manseau, 2009; Fieberg et al., 2010). In this study, while BBMMs highlighted individual variation in space use, simplified GLMMs indicated including individual as a random effect contributed relatively less to explaining overall variance than the fixed effects alone. However, this was not the case for lemon sharks, suggesting larger potential differences in relative habitat selection across individuals. Confirmed by individual BBMMs and network analyses from Casselberry et al. (2020), some lemon sharks were consistently close to the island while others used areas farther away and at greater depths. While approaches are being developed to incorporate mixed effects into machine learning algorithms (Hajjem et al., 2014), it is still relatively inaccessible due to its complexity. Future studies using RSFs and machine learning algorithms should attempt to measure or address random variation across individuals and sample size biases either within the approach and/or with complimentary analyses. For example, using test datasets that contain individuals not used in the training dataset may better help to assess model performance and transferability (Buston and Elith, 2011; Raymond et al., 2015). Further, running models for each individual and, subsequently, collectively deriving the 95% confidence interval estimates across the computed marginal effects for all individuals may be a viable approach to assess population level effects. Alternatively, using both mixed effects models and machine learning approaches in tandem may be the most appropriate (see Shoemaker et al., 2018). With consideration to this caveat, machine learning algorithms provide useful and flexible advantages to deal with complex ecological datasets and to obtain accurate results.

Beyond the application of RSFs with machine learning algorithms, by design, passive acoustic telemetry arrays provide an intuitive approach for implementing RSFs since available resource units can easily be defined based upon receiver positioning. Constraining COAs and background points to the available resource units, defined by acoustic receiver location at the year level, allows for the incorporation or removal of additional receivers across a study period. This flexibility is ideal as arrays often change over time due to funding constraints or adapting research questions. However, to safeguard against biased relative selection estimates, it is important to ensure receiver arrays, including their modifications, are designed to capture space use that is representative of the habitat available (Selby et al., 2019; Griffin et al., 2020). For future studies aimed at examining relative selection, we suggest grid array designs (Heupel et al., 2006; Kraus et al., 2018) to achieve proportionally representative coverage of areas rather than deployments guided by *a priori* beliefs of animal space use (Brownscombe et al., 2019b). In addition, detection range and efficiency, should be considered during the array design (Brownscombe et al., 2019b), when constructing COAs (Winton et al., 2018b), when defining available resource units around receivers, or even explicitly in the modeling process (see Brownscombe et al., 2019a). Detection efficiency and range, often limited by physical structure, wind, currents, animal noise, or by human activities

may vary greatly across a given study area (Gjelland and Hedger, 2013; Kessel et al., 2014).

Here, while potentially incorporating biases due to variable detection ranges (see Selby et al., 2016), we chose a 400 m buffer around each receiver to allow for COAs and associated background points to extend beyond observed detection ranges. However, we recommend testing a wide range of parameter inputs from COA time bin selection to available habitat buffer size. Such inputs should be guided by ecological knowledge, acoustic telemetry coverage, and model accuracy metrics. While COA time bins of 60-min provided more accurate measures for some species and refined predictions, we opted for 90-min bins for all species since this would potentially reduce issues with autocorrelation by subsampling further (Swihart and Slade, 1985). Further, COA time bin selection should consider both the programmed tag delay and the speed of tagged animals, with smaller time bins for faster moving species and larger time bins for slow moving animals. In this example, we found a smaller available habitat buffer produced lower accuracy metrics and led to unreliable predictions across BIRNM. Alternatively, applying a larger available habitat buffer provided similar results to the original models that used 400 m buffers and also captured relative selection for tiger sharks in areas (northeast shelf) where we expected higher values. While 400 m buffers were chosen for consistency to Selby et al. (2019) and Griffin et al. (2020), future researchers should explore and evaluate multiple extents for a given species, study area, and array. Along with variable detection range and efficiency, future RSF studies using acoustic telemetry should also investigate the role of spatial and/or temporal scales on selection modeling (McGarigal et al., 2016); this is especially relevant when collapsing habitat and presences/background points for model implementation.

Conclusion

In summary, we highlight the utility of combining acoustic telemetry, RSFs, and machine learning to understand and accurately predict the relative habitat selection of marine animals across both monitored and unmonitored areas. While RSFs have been used extensively within terrestrial environments, we suggest marine ecologists should also adopt these methods to improve resource management actions. Such applications could help to prioritize habitat protection and restoration in the face of continued anthropogenic threats (Millennium Ecosystem Assessment, 2003). This may have particular advantages centered around MPA design. Here, applied to four shark species within an MPA, we found accurate models that could extrapolate to areas where receiver coverage was limited. Further, when these RSF values were extended to examine predator-prey relationships, we found areas that varied in mutual selection, highlighting the potential overlap of predators and their prey.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are not readily available because studies for each individual species are currently ongoing. Data can be made available upon reasonable request through GC at grace.casselberry@gmail.com or directly to the data owners: GS (sharks), AJ (reef fish), and KH (sea turtles). Code for the resource selection function framework with a sample dataset can be found at https://github.com/lucaspgriffin.

ETHICS STATEMENT

The animal study was reviewed and approved by University of Massachusetts Amherst IACUC no. 2013-0031 and 2019-0043. Research conducted within BIRNM was approved by NPS under study no. BUIS-00058 and individual research collection permit nos. BUIS-2013-SCI_0003, BUIS-2014-SCI-0006, and BUIS-2019-0010. For green turtle fieldwork, permitting was under NMFS permits 16146 and 20315, issued to KH, National Park IACUC USGS-SESC2014-02 and USGS IACUC WARC\GNV 2017-04. Additional permits issued to KH BUIS-2011-SCI-0012; BUIS-2014-SCI-0009; BUIS-2016-SCI-0009. Acoustic receiver stations were deployed under the following permit numbers through the United States Army Corp of Engineers: SAJ-2014-01790, SAJ-2015-02061, SAJ-2015-02062, SAJ-2017-00622, and SAJ-2017-00624.

AUTHOR CONTRIBUTIONS

LG and GC conceived and led the study. GC, KH, AJ, SB, AN, BD, CP, IL, ZH-S, AD, and GS conducted the field work. LG analyzed the data. All authors interpreted the findings, wrote the manuscript, and approved the final version.

FUNDING

This work was supported by grants from the following funders: Puerto Rico Sea Grant (R-101-2-14), The New England Aquarium's Marine Conservation Action Fund, The Atlantic White Shark Conservancy, National Geographic Society, The Allen Family Foundation, and the USGS Natural Resource Protection Program and USGS Ecosystems Program.

ACKNOWLEDGMENTS

We acknowledge the USGS Natural Resource Protection Program (NRPP) for funding. AD was supported by the National Institute of Food and Agriculture, United States Department of Agriculture, the Massachusetts Agricultural Experiment Station, and the Department of Environmental Conservation. GC was supported by the NOAA ONMS Dr. Nancy Foster Scholarship. We thank the numerous National Park Service staff and interns for their work maintaining BIRNM receiver array throughout the years, especially Tessa Code and Nathaniel Hanna Holloway. We also thank Jake Brownscombe and Laura D'Acunto for review of an earlier draft of the manuscript. We thank Brace

Thompson for providing his artistic talents to figure one. Finally, we thank the reviewers and associate editor for thorough and constructive feedback. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the United States Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.631262/full#supplementary-material

REFERENCES

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M., and Matthiopoulos, J. (2008). Estimating space−use and habitat preference from wildlife telemetry data. *Ecography* 31, 140–160. doi: 10.1111/j.2007.0906-7590.05236.x

Aines, A. C., Carlson, J. K., Boustany, A., Mathers, A., and Kohler, N. E. (2018). Feeding habits of the tiger shark, *Galeocerdo cuvier*, in the northwest Atlantic Ocean and Gulf of Mexico. *Environ. Biol. Fishes* 101, 403–415. doi: 10.1007/s10641-017-0706-y

Allen, A. M., and Singh, N. J. (2016). Linking movement ecology with wildlife management and conservation. *Front. Ecol. Evol.* 3, 1–13. doi: 10.3389/fevo.2015.00155

Avgar, T., Lele, S. R., Keim, J. L., and Boyce, M. S. (2017). Relative selection strength: quantifying effect size in habitat−and step−selection inference. *Ecol. Evol.* 7, 5322–5330. doi: 10.1002/ece3.3122

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* 3, 327–338. doi: 10.1111/j.2041-210x.2011.00172.x

Barley, S. C., Meekan, M. G., and Meeuwig, J. J. (2017a). Diet and condition of mesopredators on coral reefs in relation to shark abundance. *PLoS One* 12:e0165113. doi: 10.1371/journal.pone.0165113

Barley, S. C., Meekan, M. G., and Meeuwig, J. J. (2017b). Species diversity, abundance, biomass, size and trophic structure of fish on coral reefs in relation to shark abundance. *Mar. Ecol. Prog. Ser.* 565, 163–179. doi: 10.3354/meps11981

Bascompte, J., Melián, C. J., and Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proc. Natl. Acad. Sci.U.S.A.* 102, 5443–5447. doi: 10.1073/pnas.0501562102

Baum, J. K., and Worm, B. (2009). Cascading top-down effects of changing oceanic predator abundances. *J. Anim. Ecol.* 78, 699–714. doi: 10.1111/j.1365-2656.2009.01531.x

Becker, S. L., Finn, J. T., Danylchuk, A. J., Pollock, C. G., Hillis-Starr, Z., Lundgren, I., et al. (2016). Influence of detection history and analytic tools on quantifying spatial ecology of a predatory fish in a marine protected area. *Mar. Ecol. Prog. Ser.* 562, 147–161. doi: 10.3354/meps11962

Becker, S. L., Finn, J. T., Novak, A. J., Danylchuk, A. J., Pollock, C. G., Hillis-Starr, Z., et al. (2020). Coarse-and fine-scale acoustic telemetry elucidates movement patterns and temporal variability in individual territories for a key coastal mesopredator. *Environ. Biol. Fishes* 103, 13–29. doi: 10.1007/s10641-019-00930-2

Beier, P., Majka, D. R., and Spencer, W. D. (2008). Forks in the road: choices in procedures for designing wildland linkages. *Conserv. Biol.* 22, 836–851. doi: 10.1111/j.1523-1739.2008.00942.x

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., et al. (2016). mlr: machine learning in R. *J. Mach. Learn. Res.* 17, 5938–5942.

Bivand, R., and Lewin-Koh, N. (2013). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8 27.

Block, W. M., and Brennan, L. A. (1993). "The habitat concept in ornithology," in *Current Ornithology*, ed. D. M. Power (New York, NY: Springer), 35–91. doi: 10.1007/978-1-4757-9912-5_2

Bouchet, P., Miller, D. L., Roberts, J., Mannocci, L., Harris, C. M., and Thomas, L. (2019). *From Here and Now to There and Then: Practical Recommendations for Extrapolating Cetacean Density Surface Models to Novel Conditions.* (CREEM technical report No. 2019–01). St Andrews: University of St Andrews.

Bouchet, P. J., Miller, D. L., Roberts, J. J., Mannocci, L., Harris, C. M., and Thomas, L. (2020). dsmextra: extrapolation assessment tools for density surface models. *Methods Ecol. Evolution.* 11, 1464–1469. doi: 10.1111/2041-210x.13469

Boyce, M. S. (2006). Scale for resource selection functions. *Divers. Distrib.* 12, 269–276. doi: 10.1111/j.1366-9516.2006.00243.x

Boyce, M. S., and McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. *Trends Ecol. Evol.* 14, 268–272. doi: 10.1016/s0169-5347(99)01593-1

Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecol. Model.* 157, 281–300. doi: 10.1016/s0304-3800(02)00200-4

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Brownscombe, J. W., Griffin, L. P., Chapman, J. M., Morley, D., Acosta, A., Crossin, G. T., et al. (2019a). A practical method to account for variation in detection range in acoustic telemetry arrays to accurately quantify the spatial ecology of aquatic animals. *Methods Ecol. Evol.* 11, 82–94. doi: 10.1111/2041-210X.13322

Brownscombe, J. W., Griffin, L. P., Morley, D., Acosta, A., Hunt, J., Lowerre-Barbieri, S. K., et al. (2020). Application of machine learning algorithms to identify cryptic reproductive habitats using diverse information sources. *Oecologia* 194, 283–298. doi: 10.1007/s00442-020-04753-2

Brownscombe, J. W., Lédée, E. J. I. I., Raby, G. D., Struthers, D. P., Gutowsky, L. F. G. G., Nguyen, V. M., et al. (2019b). Conducting and interpreting fish telemetry studies: considerations for researchers and resource managers. *Rev. Fish Biol. Fish.* 29, 369–400. doi: 10.1007/s11160-019-09560-4

Bryan, D. R., Feeley, M. W., Nemeth, R. S., Pollock, C., and Ault, J. S. (2019). Home range and spawning migration patterns of queen triggerfish Balistes vetula in St. Croix, US Virgin Islands. *Mar. Ecol. Prog. Ser.* 616, 123–139. doi: 10.3354/meps12944

Buston, P. M., and Elith, J. (2011). Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *J. Anim. Ecol.* 80, 528–538. doi: 10.1111/j.1365-2656.2011.01803.x

Calenge, C. (2006). The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. *Ecol. Model.* 197, 516–519. doi: 10.1016/j.ecolmodel.2006.03.017

Campbell, H. A., Watts, M. E., Dwyer, R. G., and Franklin, C. E. (2012). V-Track: software for analysing and visualising animal movement from acoustic telemetry detections. *Mar. Freshw. Res.* 63, 815–820. doi: 10.1071/MF12194

Carlisle, A. B., Tickler, D., Dale, J. J., Ferretti, F., Curnick, D. J., Chapple, T. K., et al. (2019). Estimating space use of mobile fishes in a large marine protected area with methodological considerations in acoustic array design. *Front. Mar. Sci.* 6:256. doi: 10.3389/fmars.2019.00256

Carlson, J., Charvet, P., Blanco-Parra, M. P, Briones Bell-lloch, A., Cardenosa, D., Derrick, D., et al. (2021a). *Carcharhinus perezi. The IUCN Red List of Threatened Species 2021*: e.T60217A3093780. doi: 10.2305/IUCN.UK.2021-1.RLTS.T60217A3093780.en

Carlson, J., Charvet, P., Ba, A., Bizzarro, J., Derrick, D., Espinoza, M., et al. (2021b). *Negaprion Brevirostris. The IUCN Red List of Threatened Species 2021*: e.T39380A2915472. doi: 10.2305/IUCN.UK.2021-1.RLTS.T39380A2915472.en

Carlson, J., Charvet, P., Blanco-Parra, M. P., Briones Bell-lloch, A., Cardenosa, D., Derrick, D., et al. (2021c). *Ginglymostoma Cirratum. The IUCN Red List of Threatened Species 2021*: e.T144141186A3095153. doi: 10.2305/IUCN.UK.2021-1.RLTS.T144141186A3095153.en

Cashion, T., Nguyen, T., ten Brink, T., Mook, A., Palacios-Abrantes, J., and Roberts, S. M. (2020). Shifting seas, shifting boundaries: dynamic marine protected area designs for a changing climate. *PLoS One* 15:e0241771. doi: 10.1371/journal.pone.0241771

Casselberry, G. A., Danylchuk, A. J., Finn, J. T., DeAngelis, B. M., Jordaan, A., Pollock, C. G., et al. (2020). Network analysis reveals multispecies spatial associations in the shark community of a Caribbean marine protected area. *Mar. Ecol. Prog. Ser.* 633, 105–126. doi: 10.3354/meps13158

Chetkiewicz, C. B., and Boyce, M. S. (2009). Use of resource selection functions to identify conservation corridors. *J. Appl. Ecol.* 46, 1036–1047. doi: 10.1111/j.1365-2664.2009.01686.x

Ciarniello, L. M., Boyce, M. S., Heard, D. C., and Seip, D. R. (2007). Components of grizzly bear habitat selection: density, habitats, roads, and mortality risk. *J. Wildl. Manag.* 71, 1446–1457. doi: 10.2193/2006-229

Cooke, S. J. (2008). Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and IUCN Red List threat assessments. *Endanger. Species Res.* 4, 165–185. doi: 10.3354/esr00063

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., et al. (2004). Biotelemetry: a mechanistic approach to ecology. *Trends Ecol. Evol.* 19, 334–343. doi: 10.1016/j.tree.2004.04.003

Costa, B. M., Tormey, S., and Battista, T. A. (2012). *Benthic Habitats of Buck Island Reef National Monument.* NOAA Technical Memorandum NOS NCCOS 142. Silver Spring, MD: NCCOS Center for Coastal Monitoring and Assessment Biogeography Branch, 64.

Coverdale, T. C., Herrmann, N. C., Altieri, A. H., and Bertness, M. D. (2013). Latent impacts: the role of historical human activity in coastal habitat loss. *Front. Ecol. Environ.* 11:69–74. doi: 10.1890/120130

Craig, J. K., and Crowder, L. B. (2002). "Factors influencing habitat selection in fishes with a review of marsh ecosystems," in *Concepts and Controversies in Tidal Marsh Ecology*, eds M. P. Weinstein and D. A. Kreeger (Dordrecht: Springer), 241–266. doi: 10.1007/0-306-47534-0_12

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792. doi: 10.1890/07-0539.1

Daly, R., Smale, M. J., Singh, S., Anders, D., Shivji, M. K., Daly, C. A., et al. (2018). Refuges and risks: evaluating the benefits of an expanded MPA network for mobile apex predators. *Divers. Distrib.* 24, 1217–1230. doi: 10.1111/ddi.12758

Descombes, P., Wisz, M. S., Leprieur, F., Parravicini, V., Heine, C., Olsen, S. M., et al. (2015). Forecasted coral reef decline in marine biodiversity hotspots under climate change. *Glob. Change Biol.* 21, 2479–2487. doi: 10.1111/gcb.12868

Donaldson, M. R., Hinch, S. G., Suski, C. D., Fisk, A. T., Heupel, M. R., and Cooke, S. J. (2014). Making connections in aquatic ecosystems with acoustic telemetry monitoring. *Front. Ecol. Environ.* 12:565–573. doi: 10.1890/130283

Durden, J. M., Luo, J. Y., Alexander, H., Flanagan, A. M., and Grossmann, L. (2017). Integrating "big data" into aquatic ecology: CHALLENGES and opportunities. *Limnol. Oceanogr. Bull.* 26, 101–108. doi: 10.1002/lob.10213

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Espinoza, M., Farrugia, T. J., Webber, D. M., Smith, F., and Lowe, C. G. (2011). Testing a new acoustic telemetry technique to quantify long-term, fine-scale movements of aquatic animals. *Fish. Res.* 108, 364–371. doi: 10.1016/j.fishres.2011.01.011

Feeley, M. W., Morley, D., Acosta, A., Barbera, P., Hunt, J., Switzer, T., et al. (2018). Spawning migration movements of Mutton Snapper in Tortugas, Florida: spatial dynamics within a marine reserve network. *Fish. Res.* 204, 209–223. doi: 10.1016/j.fishres.2018.02.020

Ferreira, L. C., and Simpfendorfer, C. (2019). *Galeocerdo Cuvier.* The IUCN Red List of Threatened Species 2019, e-T39378A2913541.

Ferretti, F., Worm, B., Britten, G. L., Heithaus, M. R., and Lotze, H. K. (2010). Patterns and ecosystem consequences of shark declines in the ocean. *Ecol. Lett.* 13, 1055–1071.

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., and Frair, J. L. (2010). Correlation and studies of habitat selection: problem, red herring or opportunity? *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2233–2244. doi: 10.1098/rstb.2010.0079

Fieberg, J. R., Forester, J. D., Street, G. M., Johnson, D. H., ArchMiller, A. A., and Matthiopoulos, J. (2018). Used−habitat calibration plots: a new procedure for validating species distribution, resource selection, and step−selection models. *Ecography* 41, 737–752. doi: 10.1111/ecog.03123

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., and Calabrese, J. M. (2015). Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology* 96, 1182–1188. doi: 10.1890/14-2010.1

Foley, M. M., Halpern, B. S., Micheli, F., Armsby, M. H., Caldwell, M. R., Crain, C. M., et al. (2010). Guiding ecological principles for marine spatial planning. *Mar. Policy* 34, 955–966. doi: 10.1016/j.marpol.2010.02.001

Fraschetti, S., D'Ambrosio, P., Micheli, F., Pizzolante, F., Bussotti, S., and Terlizzi, A. (2009). Design of marine protected areas in a human-dominated seascape. *Mar. Ecol. Prog. Ser.* 375, 13–24. doi: 10.3354/meps07781

Freitas, C., Olsen, E. M., Knutsen, H., Albretsen, J., and Moland, E. (2016). Temperature−associated habitat selection in a cold−water marine fish. *J. Anim. Ecol.* 85, 628–637. doi: 10.1111/1365-2656.12458

Friedman, J. H., and Popescu, B. E. (2008). Predictive learning via rule ensembles. *Ann. Appl. Stat.* 2, 916–954.

Fu, A. L., Hammerschlag, N., Lauder, G. V., Wilga, C. D., Kuo, C., and Irschick, D. J. (2016). Ontogeny of head and caudal fin shape of an apex marine predator: the tiger shark (*Galeocerdo cuvier*). *J. Morphol.* 277, 556–564. doi: 10.1002/jmor.20515

Gallagher, A. J., Shipley, O. N., van Zinnicq Bergmann, M. P., Brownscombe, J. W., Dahlgren, C. P., Frisk, M. G., et al. (2021). Spatial connectivity and drivers of shark habitat use within a large marine protected area in the caribbean, the bahamas shark sanctuary. *Front. Mar. Sci.* 7:1223.

Gell, F. R., and Roberts, C. M. (2003). Benefits beyond boundaries: the fishery effects of marine reserves. *Trends Ecol. Evol.* 18, 448–455. doi: 10.1016/s0169-5347(03)00189-7

Gillies, C. S., Hebblewhite, M., Nielsen, S. E., Krawchuk, M. A., Aldridge, C. L., Frair, J. L., et al. (2006). Application of random effects to the study of resource selection by animals. *J. Anim. Ecol.* 75, 887–898. doi: 10.1111/j.1365-2656.2006.01106.x

Gjelland, K. Ø, and Hedger, R. D. (2013). Environmental influence on transmitter detection probability in biotelemetry: developing a general model of acoustic transmission. *Methods Ecol. Evol.* 4, 665–674. doi: 10.1111/2041-210x.12057

Gleason, M., McCreary, S., Miller-Henson, M., Ugoretz, J., Fox, E., Merrifield, M., et al. (2010). Science-based and stakeholder-driven marine protected area network planning: a successful case study from north central California. *Ocean Coast. Manag.* 53, 52–68. doi: 10.1016/j.ocecoaman.2009.12.001

Godvik, I. M. R., Loe, L. E., Vik, J. O., Veiberg, V., Langvatn, R., and Mysterud, A. (2009). Temporal scales, trade−offs, and functional responses in red deer habitat selection. *Ecology* 90, 699–710. doi: 10.1890/08-0576.1

Graham, N. A. J., Robinson, J. P. W., Smith, S. E., Govinden, R., Gendron, G., and Wilson, S. K. (2020). Changing role of coral reef marine reserves in a warming climate. *Nat. Commun.* 11, 1–8.

Greenwell, B. M. (2017). pdp: an R package for constructing partial dependence plots. *R J.* 9:421. doi: 10.32614/rj-2017-016

Griffin, L. L. P., Finn, J. T. J., Diez, C., and Danylchuk, A. A. J. (2019). Movements, connectivity, and space use of immature green turtles within coastal habitats of the Culebra Archipelago, Puerto Rico: implications for conservation. *Endanger. Species Res.* 40, 75–90. doi: 10.3354/esr00976

Griffin, L. P., Smith, B. J., Cherkiss, M. S., Crowder, A. G., Pollock, C. G., Starr, Z. H., et al. (2020). Space use and relative habitat selection for immature green turtles within a Caribbean marine protected area. *Anim. Biotelem.* 8, 1–13. doi: 10.1186/s40317-020-00209-9

Gutowsky, L. F. G., Harrison, P. M., Martins, E. G., Leake, A., Patterson, D. A., Zhu, D. Z., et al. (2017). Daily temperature experience and selection by adfluvial bull trout (*Salvelinus confluentus*). *Environ. Biol. Fishes* 100, 1167–1180. doi: 10.1007/s10641-017-0634-x

Gutowsky, L. F. G., Romine, J. G., Heredia, N. A., Bigelow, P. E., Parsley, M. J., Sandstrom, P. T., et al. (2020). Revealing migration and reproductive habitat of invasive fish under an active population suppression program. *Conserv. Sci. Pract.* 2: e119.

Hajjem, A., Bellavance, F., and Larocque, D. (2014). Mixed-effects random forest for clustered data. *J. Stat. Comput. Simul.* 84, 1313–1328. doi: 10.1080/00949655.2012.741599

Hammerschlag, N., Schmitz, O. J., Flecker, A. S., Lafferty, K. D., Sih, A., Atwood, T. B., et al. (2019). Ecosystem function and services of aquatic predators in the anthropocene. *Trends Ecol. Evol.* 34, 369–383. doi: 10.1016/j.tree.2019.01.005

Harrison, P. M., Gutowsky, L. F. G., Martins, E. G., Patterson, D. A., Cooke, S. J., and Power, M. (2016). Temporal plasticity in thermal−habitat selection of burbot Lota lota a diel−migrating winter−specialist. *J. Fish Biol.* 88, 2111–2129. doi: 10.1111/jfb.12990

Hastings, R. A., Rutterford, L. A., Freer, J. J., Collins, R. A., Simpson, S. D., and Genner, M. J. (2020). Climate change drives poleward increases and equatorward declines in marine species. *Curr. Biol.* 30, 1572–1577.e2.

Hays, G. C., Bailey, H., Bograd, S. J., Bowen, W. D., Campagna, C., Carmichael, R. H., et al. (2019). Translating marine animal tracking data into conservation policy and management. *Trends Ecol. Evol.* 34, 459–473.

Hays, G. C., Ferreira, L. C., Sequeira, A. M. M., Meekan, M. G., Duarte, C. M., Bailey, H., et al. (2016). Key questions in marine megafauna movement ecology. *Trends Ecol. Evol.* 31, 463–475.

Hebblewhite, M., and Merrill, E. (2008). Modelling wildlife–human relationships for social species with mixed−effects resource selection models. *J. Appl. Ecol.* 45, 834–844. doi: 10.1111/j.1365-2664.2008.01466.x

Heffelfinger, L. J., Stewart, K. M., Shoemaker, K. T., Darby, N. W., and Bleich, V. C. (2020). Balancing current and future reproductive investment: variation in resource selection during stages of reproduction in a long-lived herbivore. *Front. Ecol. Evol.* 8:163. doi: 10.3389/fevo.2020.00163

Heinrichs, J. A., Aldridge, C. L., O'Donnell, M. S., and Schumaker, N. H. (2017). Using dynamic population simulations to extend resource selection analyses and prioritize habitats for conservation. *Ecol. Model.* 359, 449–459. doi: 10.1016/j.ecolmodel.2017.05.017

Heithaus, M. R., and Dill, L. M. (2002). Food availability and tiger shark predation risk influence bottlenose dolphin habitat use. *Ecology* 83, 480–491. doi: 10.1890/0012-9658(2002)083[0480:faatsp]2.0.co;2

Heithaus, M. R., Dill, L. M., Marshall, G. J., and Buhleier, B. (2002). Habitat use and foraging behavior of tiger sharks (*Galeocerdo cuvier*) in a seagrass ecosystem. *Mar. Biol.* 140, 237–248. doi: 10.1007/s00227-001-0711-7

Heithaus, M. R., Frid, A., Wirsing, A. J., Dill, L. M., Fourqurean, J. W., Burkholder, D., et al. (2007). State−dependent risk−taking by green sea turtles mediates top−down effects of tiger shark intimidation in a marine ecosystem. *J. Anim. Ecol.* 76, 837–844. doi: 10.1111/j.1365-2656.2007.01260.x

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6:e5518. doi: 10.7717/peerj.5518

Heupel, M. R., Knip, D. M., Simpfendorfer, C. A., and Dulvy, N. K. (2014). Sizing up the ecological role of sharks as predators. *Mar. Ecol. Prog. Ser.* 495, 291–298. doi: 10.3354/meps10597

Heupel, M. R., Munroe, S. E. M., Lédée, E. J. I., Chin, A., and Simpfendorfer, C. A. (2019). Interspecific interactions, movement patterns and habitat use in a diverse coastal shark assemblage. *Mar. Biol.* 166:68.

Heupel, M. R., Semmens, J. M., and Hobday, A. J. (2006). Automated acoustic tracking of aquatic animals: scales, design and deployment of listening station arrays. *Mar. Freshw. Res.* 57, 1–13. doi: 10.1071/mf05091

Hijmans, R. J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., et al. (2015). *Package 'raster.'* R package.

Hixon, M. A., and Beets, J. P. (1993). Predation, prey refuges, and the structure of coral−reef fish assemblages. *Ecol. Monogr.* 63, 77–101. doi: 10.2307/2937124

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., and Roberts, C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecol. Lett.* 8, 23–29. doi: 10.1111/j.1461-0248.2004.00686.x

Horne, J. S., Garton, E. O., Krone, S. M., and Lewis, J. S. (2007). Analyzing animal movements using Brownian bridges. *Ecology* 88, 2354–2363. doi: 10.1890/06-0957.1

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., et al. (2015). ECOLOGY. Aquatic animal telemetry: a panoramic window into the underwater world. *Science (New York, N.Y.)* 348:1255642. doi: 10.1126/science.1255642

Jacoby, D. M. P., Ferretti, F., Freeman, R., Carlisle, A. B., Chapple, T. K., Curnick, D. J., et al. (2020). Shark movement strategies influence poaching risk and can guide enforcement decisions in a large, remote marine protected area. *J. Appl. Ecol.* 57, 1782–1792. doi: 10.1111/1365-2664.13654

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York, NY: Springer.

Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDONALD, T. L., and Boyce, M. S. (2006). Resource selection functions based on use−availability data: theoretical motivation and evaluation methods. *J. Wildl. Manag.* 70, 347–357. doi: 10.2193/0022-541x(2006)70[347:rsfbou]2.0.co;2

Johnson, C. J., Seip, D. R., and Boyce, M. S. (2004). A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. *J. Appl. Ecol.* 41, 238–251. doi: 10.1111/j.0021-8901.2004.00899.x

Johnson, D. S., Hooten, M. B., and Kuhn, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *J. Anim. Ecol.* 82, 1155–1164. doi: 10.1111/1365-2656.12087

Keller, J. A., Herbig, J. L., Morley, D., Wile, A., Barbera, P., and Acosta, A. (2020). Grouper tales: use of acoustic telemetry to evaluate grouper movements at Western Dry Rocks in the Florida Keys. *Mar. Coast. Fish.* 12, 290–307. doi: 10.1002/mcf2.10109

Kessel, S. T., Cooke, S. J., Heupel, M. R., Hussey, N. E., Simpfendorfer, C. A., Vagle, S., et al. (2014). A review of detection range testing in aquatic passive acoustic telemetry studies. *Rev. Fish Biol. Fish.* 24, 199–218. doi: 10.1007/s11160-013-9328-4

Knip, D. M., Heupel, M. R., and Simpfendorfer, C. A. (2012). Evaluating marine protected areas for the conservation of tropical coastal sharks. *Biol. Conserv.* 148, 200–209. doi: 10.1016/j.biocon.2012.01.008

Koper, N., and Manseau, M. (2009). Generalized estimating equations and generalized linear mixed−effects models for modelling resource selection. *J. Appl. Ecol.* 46, 590–599. doi: 10.1111/j.1365-2664.2009.01642.x

Kramer, D. L., and Chapman, M. R. (1999). Implications of fish home range size and relocation for marine reserve function. *Environ. Biol. Fishes* 55, 65–79. doi: 10.1023/a:1007481206399

Kramer, D. L., Rangeley, R. W., and Chapman, L. J. (1997). "Habitat selection: patterns of spatial distribution from behavioural decisions," in *Behavioural Ecology of Teleost Fishes*, ed. J.-G. J. Godin (Oxford: Oxford University Press), 37–80.

Kraus, R. T., Holbrook, C. M., Vandergoot, C. S., Stewart, T. R., Faust, M. D., Watkinson, D. A., et al. (2018). Evaluation of acoustic telemetry grids for determining aquatic animal movement and survival. *Methods Ecol. Evol.* 9, 1489–1502. doi: 10.1111/2041-210x.12996

Lea, J. S. E. E., Humphries, N. E., von Brandis, R. G., Clarke, C. R., and Sims, D. W. (2016). Acoustic telemetry and network analysis reveal the space use of multiple reef predators and enhance marine protected area design. *Proc. R. Soc. B Biol. Sci.* 283:20160717. doi: 10.1098/rspb.2016.0717

Ledee, E. J. I., Heupel, M. R., Tobin, A. J., Mapleston, A., and Simpfendorfer, C. A. (2016). Movement patterns of two carangid species in inshore habitats characterised using network analysis. *Mar. Ecol. Prog. Ser.* 553, 219–232. doi: 10.3354/meps11777

Legare, B., Kneebone, J., DeAngelis, B., and Skomal, G. (2015). The spatiotemporal dynamics of habitat use by blacktip (*Carcharhinus limbatus*) and lemon (*Negaprion brevirostris*) sharks in nurseries of St. John, United States Virgin Islands. *Mar. Biol.* 162, 699–716. doi: 10.1007/s00227-015-2616-x

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659–1673. doi: 10.2307/1939924

Lele, S. R., Merrill, E. H., Keim, J., and Boyce, M. S. (2013). Selection, use, choice and occupancy: clarifying concepts in resource selection studies. *J. Anim. Ecol.* 82, 1183–1191. doi: 10.1111/1365-2656.12141

Lichti, N. I., and Swihart, R. K. (2011). Estimating utilization distributions with kernel versus local convex hull methods. *J. Wildl. Manag.* 75, 413–422. doi: 10.1002/jwmg.48

Lovelace, R., Nowosad, J., and Muenchow, J. (2019). *Geocomputation with R*. Boca Raton, FL: CRC Press.

Lowe, C. G., Wetherbee, B. M., Crow, G. L., and Tester, A. L. (1996). Ontogenetic dietary shifts and feeding behavior of the tiger shark, Galeocerdo cuvier, in Hawaiian waters. *Environ. Biol. Fishes* 47, 203–211. doi: 10.1007/bf00005044

Lowerre-Barbieri, S. K., Catalán, I. A., Frugård Opdal, A., and Jørgensen, C. (2019). Preparing for the future: integrating spatial ecology into ecosystem-based management. *ICES J. Mar. Sci.* 76, 467–476. doi: 10.1093/icesjms/fsy209

Lubchenco, J., and Grorud-Colvert, K. (2015). OCEAN. Making waves: the science and politics of ocean protection. *Science (New York, N.Y.)* 350, 382–383. doi: 10.1126/science.aad5443

Lubchenco, J., Palumbi, S. R., Gaines, S. D., and Andelman, S. (2003). Plugging a hole in the ocean: the emerging science of marine reserves 1. *Ecol. Appl.* 13, 3–7. doi: 10.1890/1051-0761(2003)013[0003:pahito]2.0.co;2

Lüdecke, D., Makowski, D., Waggoner, P., and Patil, I. (2019). *performance: Assessment of Regression Models Performance.* R package version 0.4. 0.

Magnusson, A., Skaug, H. J., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., et al. (2017). *glmmTMB: Generalized Linear Mixed Models Using Template Model Builder*.

Manly, B. F. L., McDonald, L., Thomas, D. L., McDonald, T. L., and Erickson, W. P. (2007). *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Berlin: Springer Science & Business Media.

Matley, J. K., Heupel, M. R., Fisk, A. T., Simpfendorfer, C. A., and Tobin, A. J. (2017). Measuring niche overlap between co-occurring *Plectropomus* spp. using acoustic telemetry and stable isotopes. *Mar. Freshw. Res.* 68, 1468–1478. doi: 10.1071/mf16120

McGarigal, K., Wan, H. Y., Zeller, K. A., Timm, B. C., and Cushman, S. A. (2016). Multi-scale habitat selection modeling: a review and outlook. *Landsc. Ecol.* 31, 1161–1175. doi: 10.1007/s10980-016-0374-x

McLoughlin, P. D., Case, R. L., Gau, R. J., Cluff, D. H., Mulders, R., and Messier, F. (2002). Hierarchical habitat selection by barren-ground grizzly bears in the central Canadian Arctic. *Oecologia* 132, 102–108. doi: 10.1007/s00442-002-0941-5

Meager, J. J., Schlacher, T. A., and Nielsen, T. (2012). Humans alter habitat selection of birds on ocean−exposed sandy beaches. *Divers. Distrib.* 18, 294–306. doi: 10.1111/j.1472-4642.2011.00873.x

Mesgaran, M. B., Cousens, R. D., and Webber, B. L. (2014). Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Divers. Distrib.* 20, 1147–1159. doi: 10.1111/ddi.12209

Micheli, F., Amarasekare, P., Bascompte, J., and Gerber, L. R. (2004). Including species interactions in the design and evaluation of marine reserves: some insights from a predator-prey model. *Bull. Mar. Sci.* 74, 653–669.

Millennium Ecosystem Assessment (2003). *Ecosystems and Human Well-being: a Framework for Assessment*. Washington, D.C: Island Press.

Molnar, C., Casalicchio, G., and Bischl, B. (2018). iml: an R package for interpretable machine learning. *J. Open Source Softw.* 3:786. doi: 10.21105/joss.00786

Morris, D. W., and Kingston, S. R. (2002). Predicting future threats to biodiversity from habitat selection by humans. *Evol. Ecol. Res.* 4, 787–810.

Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed−effects models. *Methods Ecol. Evol.* 4, 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Nguyen, V. M., Young, N., Brownscombe, J. W., and Cooke, S. J. (2019). Collaboration and engagement produce more actionable science: quantitatively analyzing uptake of fish tracking studies. *Ecol. Appl.* 29: e01943.

Nielsen, S. E., Boyce, M. S., Stenhouse, G. B., and Munro, R. H. M. (2003). Development and testing of phenologically driven grizzly bear habitat models. *Ecoscience* 10, 1–10. doi: 10.1080/11956860.2003.11682743

Novak, A. J., Becker, S. L., Finn, J. T., Danylchuk, A. J., Pollock, C. G., Hillis-Starr, Z., et al. (2020a). Inferring residency and movement patterns of horse-eye jack Caranx latus in relation to a Caribbean marine protected area acoustic telemetry array. *Anim. Biotelem.* 8, 1–13.

Novak, A. J., Becker, S. L., Finn, J. T., Pollock, C. G., Hillis−Starr, Z., and Jordaan, A. (2020b). Scale of biotelemetry data influences ecological interpretations of space and habitat use in *Yellowtail Snapper*. *Mar. Coast. Fish.* 12, 364–377. doi: 10.1002/mcf2.10119

Ogburn, M. B., Harrison, A.-L., Whoriskey, F. G., Cooke, S. J., Mills Flemming, J. E., and Torres, L. G. (2017). Addressing challenges in the application of animal movement ecology to aquatic conservation and management. *Front. Mar. Sci.* 4:70. doi: 10.3389/fmars.2017.00070

Olden, J. D., Lawler, J. J., and Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83, 171–193. doi: 10.1086/587826

O'Leary, B. C., Winther−Janson, M., Bainbridge, J. M., Aitken, J., Hawkins, J. P., and Roberts, C. M. (2016). Effective coverage targets for ocean protection. *Conserv. Lett.* 9, 398–404. doi: 10.1111/conl.12247

Ordiz, A., Uzal, A., Milleret, C., Sanz-Pérez, A., Zimmermann, B., Wikenros, C., et al. (2020). Wolf habitat selection when sympatric or allopatric with brown bears in Scandinavia. *Sci. Rep.* 10, 1–11.

O'Shea, O. R., Mandelman, J., Talwar, B., and Brooks, E. J. (2015). Novel observations of an opportunistic predation event by four apex predatory sharks. *Mar. Freshw. Behav. Physiol.* 48, 374–380. doi: 10.1080/10236244.2015.1054097

Pearce, J. L., and Boyce, M. S. (2006). Modelling distribution and abundance with presence−only data. *J. Appl. Ecol.* 43, 405–412. doi: 10.1111/j.1365-2664.2005.01112.x

Pebesma, E., Bivand, R., Pebesma, M. E., RColorBrewer, S., and Collate, A. A. A. (2012). *Package 'sp.' The Comprehensive R Archive Network*.

Peel, D., and Lloyd, M. G. (2004). The social reconstruction of the marine environment: towards marine spatial planning? *Town Plan. Rev.* 75, 359–378. doi: 10.3828/tpr.75.3.6

Peters, D. P. C., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., and Villanueva-Rosales, N. (2014). Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere* 5, 1–15.

Pickard, A. E., Vaudo, J. J., Wetherbee, B. M., Nemeth, R. S., Blondeau, J. B., Kadison, E. A., et al. (2016). Comparative use of a Caribbean mesophotic coral ecosystem and association with fish spawning aggregations by three species of shark. *PLoS One* 11:e0151221. doi: 10.1371/journal.pone.0151221

Pikitch, E. K., Chapman, D. D., Babcock, E. A., and Shivji, M. S. (2005). Habitat use and demographic population structure of elasmobranchs at a Caribbean atoll (Glover's Reef, Belize). *Mar. Ecol. Prog. Ser.* 302, 187–197. doi: 10.3354/meps302187
ittman

Pittman, S. J., Hile, S. D., Jeffrey, C. F. G., Caldow, C., Kendall, M. S., Monaco, M. E., et al. (2008). *Fish Assemblages and Benthic Habitats of Buck Island Reef National Monument (St. Croix, U.S. Virgin Islands) and the Surrounding Seascape: A Characterization of Spatial and Temporal Patterns. NOAA Tech Memo NOS NCCOS 71*. Silver Spring, MD: NOAA.

Pratchett, M. S., Hoey, A. S., Wilson, S. K., Messmer, V., and Graham, N. A. J. (2011). Changes in biodiversity and functioning of reef fish assemblages following coral bleaching and coral loss. *Diversity* 3, 424–452. doi: 10.3390/d3030424

Probst, P., Wright, M. N., and Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 9:e1301.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raymond, B., Lea, M., Patterson, T., Andrews−Goff, V., Sharples, R., Charrassin, J., et al. (2015). Important marine habitat off east Antarctica revealed by two decades of multi−species predator tracking. *Ecography* 38, 121–129. doi: 10.1111/ecog.01021

Roberts, K. E., Smith, B. J., Burkholder, D., and Hart, K. M. (2021). Evaluating the use of marine protected areas by endangered species: a habitat selection approach. *Ecol. Solut. Evid.* 2:e12035.

Rosa, R. S., Castro, A. L. F., Furtado, M., Monzini, J., and Grubbs, R. D. (2006a). *Ginglymostoma Cirratum*. IUCN 2016. IUCN Red List of Threatened Species.

Rosa, R. S., Mancini, P., Caldas, J. P., and Graham, R. T. (2006b). *Carcharhinus Perezi*. The IUCN Red List of Threatened Species.

Rosenzweig, M. L. (1974). "On the evolution of habitat selection," in *Proceedings of the First International Congress of Ecology*, (Wageningen: Center for Agricultural Publishing and Documentation), 404.

Rothschild, B. J., Ault, J. S., Goulletquer, P., and Héral, M. (1994). Decline of the Chesapeake Bay oyster population: a century of habitat destruction and overfishing. *Mar. Ecol. Prog. Ser.* 111, 29–39. doi: 10.3354/meps111029

Sala, E., Lubchenco, J., Grorud-Colvert, K., Novelli, C., Roberts, C., and Sumaila, U. R. (2018). Assessing real progress towards effective ocean protection. *Mar. Policy* 91, 11–13. doi: 10.1016/j.marpol.2018.02.004

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A. (2018). Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. *arXiv* [Preprint] arXiv:1803.11266. doi: 10.1016/j.ecolmodel.2019.06.002

Selby, T. H., Hart, K. M., Fujisaki, I., Smith, B. J., Pollock, C. J., Hillis-Starr, Z., et al. (2016). Can you hear me now? Range-testing a submerged passive acoustic receiver array in a Caribbean coral reef habitat. *Ecol. Evol.* 6, 4823–4835. doi: 10.1002/ece3.2228

Selby, T. H., Hart, K. M. K., Smith, B. B. J., Pollock, C. G. C., Hillis-Starr, Z., and Oli, M. M. K. (2019). Juvenile hawksbill residency and habitat use within a Caribbean marine protected area. *Endanger. Species Res.* 40, 53–64. doi: 10.3354/esr00975

Sequeira, A. M. M., Hays, G. C., Sims, D. W., Eguíluz, V. M., Rodríguez, J. P., Heupel, M. R., et al. (2019). Overhauling ocean spatial planning to improve marine megafauna conservation. *Front. Mar. Sci.* 6:639. doi: 10.3389/fmars.2019.00639

Shoemaker, K. T., Heffelfinger, L. J., Jackson, N. J., Blum, M. E., Wasley, T., and Stewart, K. M. (2018). A machine−learning approach for extending classical wildlife resource selection analyses. *Ecol. Evol.* 8, 3556–3569. doi: 10.1002/ece3.3936

Simpfendorfer, C. A., Goodreid, A. B., and McAuley, R. B. (2001). Size, sex and geographic variation in the diet of the tiger shark, *Galeocerdo cuvier*, from Western Australian waters. *Environ. Biol. Fishes* 61, 37–46. doi: 10.1023/a:1011021710183

Simpfendorfer, C. A., Heupel, M. R., and Hueter, R. E. (2002). Estimation of short-term centers of activity from an array of omnidirectional hydrophones and its use in studying animal movements. *Can. J. Fish. Aqua. Sci.* 59, 23–32. doi: 10.1139/f01-191

Simpfendorfer, C. A., Huveneers, C., Steckenreuter, A., Tattersall, K., Hoenner, X., Harcourt, R., et al. (2015). Ghosts in the data: false detections in VEMCO pulse position modulation acoustic telemetry monitoring equipment. *Anim. Biotelem.* 3:55.

Speed, C. W., Meekan, M. G., Field, I. C., McMahon, C. R., Harcourt, R. G., Stevens, J. D., et al. (2016). Reef shark movements relative to a coastal marine protected area. *Reg. Stud. Mar. Sci.* 3, 58–66. doi: 10.1016/j.rsma.2015.05.002

Sundström, L. F. (2015). *Negaprion Brevirostris.* The IUCN Red List of Threatened Species 2015: e. T39380A81769233.

Swihart, R. K., and Slade, N. A. (1985). Testing for independence of observations in animal movements. *Ecology* 66, 1176–1184. doi: 10.2307/1939170

Udyawer, V., Dwyer, R. G., Hoenner, X., Babcock, R. C., Brodie, S., Campbell, H. A., et al. (2018). A standardised framework for analysing animal detections from automated tracking arrays. *Anim. Biotelem.* 6, 1–14. doi: 10.1186/s40317-018-0162-2

Weeks, R., Green, A. L., Joseph, E., Peterson, N., and Terk, E. (2017). Using reef fish movement to inform marine reserve design. *J. Appl. Ecol.* 54, 145–152. doi: 10.1111/1365-2664.12736

White, T. D., Carlisle, A. B., Kroodsma, D. A., Block, B. A., Casagrandi, R., de Leo, G. A., et al. (2017). Assessing the effectiveness of a large marine protected area for reef shark conservation. *Biol. Conserv.* 207, 64–71. doi: 10.1016/j.biocon.2017.01.009

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147

Winton, M. V., Fay, G., Haas, H. L., Arendt, M., Barco, S., James, M. C., et al. (2018a). Estimating the distribution and relative density of satellite-tagged loggerhead sea turtles using geostatistical mixed effects models. *Mar. Ecol. Prog. Ser.* 586, 217–232. doi: 10.3354/meps12396

Winton, M. V., Kneebone, J., Zemeckis, D. R., and Fay, G. (2018b). A spatial point process model to estimate individual centres of activity from passive acoustic telemetry data. *Methods Ecol. Evol.* 9, 2262–2272. doi: 10.1111/2041-210x.13080

Wirsing, A. J., Heithaus, M. R., and Dill, L. M. (2007). Living on the edge: dugongs prefer to forage in microhabitats that allow escape from rather than avoidance of predators. *Anim. Behav.* 74, 93–101. doi: 10.1016/j.anbehav.2006.11.016

Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home−range studies. *Ecology* 70, 164–168. doi: 10.2307/1938423

Wright, M. N., and Ziegler, A. (2015). ranger: a fast implementation of random forests for high dimensional data in C and R. *arXiv* [Preprint] arXiv:1508.04409.

Zeller, K. A., Vickers, T. W., Ernest, H. B., and Boyce, W. M. (2017). Multi-level, multi-scale resource selection functions and resistance surfaces for conservation planning: pumas as a case study. *PLoS One* 12:e0179570. doi: 10.1371/journal.pone.0179570

Zuur, A. F., Ieno, E. N., and Saveliev, A. A. (2017). *Beginner's Guide to Spatial, Temporal, and Spatial-temporal Ecological Data Analysis With R-INLA: Using GLM and GLMM*, Vol. I. Newburgh: Highland Satistics LTd.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 07 July 2021
doi: 10.3389/fmars.2021.680541

Examining Scale Dependent Environmental Effects on American Lobster (*Homarus americanus*) Spatial Distribution in a Changing Gulf of Maine

Jamie Behan[1], Bai Li[2] and Yong Chen[1]*

[1] School of Marine Science, University of Maine, Orono, ME, United States, [2] ECS Federal, LLC in support of NOAA Fisheries Office of Science and Technology, Silver Spring, MD, United States

The Gulf of Maine (GOM) is a highly complex environment and previous studies have suggested the need to account for spatial nonstationarity in species distribution models (SDMs) for the American lobster (*Homarus americanus*). To explore impacts of spatial nonstationarity on species distribution, we compared models with the following three assumptions : (1) large-scale and stationary relationships between species distributions and environmental variables; (2) meso-scale models where estimated relationships differ between eastern and western GOM, and (3) finer-scale models where estimated relationships vary across eastern, central, and western regions of the GOM. The spatial scales used in these models were largely determined by the GOM coastal currents. Lobster data were sourced from the Maine-New Hampshire Inshore Bottom Trawl Survey from years 2000–2019. We considered spatial and environmental variables including latitude and longitude, bottom temperature, bottom salinity, distance from shore, and sediment grain size in the study. We forecasted distributions for the period 2028–2055 using each of these models under the Representative Concentration Pathway (RCP) 8.5 "business as usual" climate warming scenario. We found that the model with the third assumption (i.e., finest scale) performed best. This suggests that accounting for spatial nonstationarity in the GOM leads to improved distribution estimates. Large-scale models revealed a tendency to estimate global relationships that better represented a specific location within the study area, rather than estimating relationships appropriate across all spatial areas. Forecasted distributions revealed that the largest scale models tended to comparatively overestimate most season × sex × size group lobster abundances in western GOM, underestimate in the western portion of central GOM, and overestimate in the eastern portion of central GOM, with slightly less consistent and patchy trends amongst groups in eastern GOM. The differences between model estimates were greatest between the largest and finest scale models, suggesting that fine-scale models may be useful for capturing effects of unique dependencies that may operate at localized scales. We demonstrate how estimates of

OPEN ACCESS

Edited by:
Mark J. Henderson,
U.S. Geological Survey, United States

Reviewed by:
Yan Jiao,
Virginia Tech, United States
John Quinlan,
Southeast Fisheries Science Center (NOAA), United States

****Correspondence:***
Jamie Behan
Jamie.behan@maine.edu

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: *14 March 2021*
Accepted: *08 June 2021*
Published: *07 July 2021*

Citation:
Behan J, Li B and Chen Y (2021) Examining Scale Dependent Environmental Effects on American Lobster (Homarus americanus) Spatial Distribution in a Changing Gulf of Maine. Front. Mar. Sci. 8:680541. doi: 10.3389/fmars.2021.680541

season-, sex-, and size- specific American lobster spatial distribution would vary based on the spatial scale assumption of nonstationarity in the GOM. This information may help develop appropriate local adaptation measures in a region that is susceptible to climate change.

Keywords: nonstationary, spatial distribution, American lobster, Gulf of Maine, scale-dependent, climate change, generalized additive models

INTRODUCTION

American lobster (*Homarus americanus)* is the most valuable fishery in the United States [National Oceanographic and Atmospheric Administration (NOAA), 2018]. The American lobster fishery in the state of Maine was worth 486 million dollars in 2019, which comprised roughly 77.1% of the total worth of the entire lobster fishery on the Atlantic coast in that year (≅$630 million, ACCSP, 2019). The Gulf of Maine (GOM) and Georges Bank (GBK) stock contributes to more than 90% of the American lobster landings in the United States (ASMFC, 2020). Additionally, the GOM has been thought to be warming 99% faster than the global ocean (Pershing et al., 2015). Knowing that the American lobster fishery is the most valuable fishery and that species' distributions commonly shift in pursuit of ideal habitat conditions (Pinsky et al., 2013; Greenan et al., 2019), it is important to understand and accurately estimate the spatial distribution of this species, especially in a rapidly changing environment.

Although the GOM/GBK lobster stock is not overfished and overfishing is not occurring (ASMFC, 2020), lobster abundance throughout the GOM is not uniformly or randomly distributed (Steneck and Wilson, 2001). Environmental factors contribute to the spatial distribution of lobster abundance, and evidence of temperature, salinity, and productivity gradients that range from northeast to southwest GOM have been observed (Lynch et al., 1997; Pettigrew et al., 1998; Chang et al., 2016). These gradients may be attributed in part by the Gulf of Maine Coastal Currents (GMCC), which form cyclonic currents across the GOM (Townsend et al., 2015; Chang et al., 2016). The GMCC can be further distinguished as two sub currents; the Eastern Maine Coastal Current (EMCC) and the Western Maine Coastal Current (WMCC), where the EMCC diverges offshore in the Penobscot Bay area and the WMCC begins along the coast (Xue et al., 2008; Chang et al., 2016). These currents can affect environmental variables as well as processes and interactions such as primary production levels, stock-recruitment relationships, and vertical mixing (Incze et al., 2010; Chang et al., 2016).

Species distribution models are widely used to estimate and predict organisms' spatial and/or temporal distributions across the world (Bakka et al., 2016; Diarra et al., 2018; Becker et al., 2020). Spatial and/or temporal nonstationarity is often present in ecological systems when relationships between response and explanatory variables vary across space and/or time, which means that the association between response and explanatory variables decrease with increasing distance (Brunsdon et al., 1996; Fotheringham et al., 2002). Past literature has demonstrated evidence of spatial nonstationarity in the GOM region (Li et al., 2018; Staples et al., 2019). Accounting for nonstationarity in SDMs allows for the incorporation of spatial and/or temporal dependencies that cannot be explained by environmental variables alone (Bakka et al., 2016). However, past literature often have not compared differences in species distribution estimates between models applied at various spatial scales.

Generalized linear models (GLMs, Nelder and Wedderburn, 1972), generalized additive models (GAMs) (GAMs; Hastie and Tibshirani, 1986), and geographically weighted regression (GWR; Brunsdon et al., 1996) are a few commonly used models for estimating species distributions. Inherently, GLMs and GAMs are stationary models because they estimate global relationships between the response and explanatory variables that are applied to all locations. In contrast, GWR models can estimate unique parameters at each location across space, thus allowing for the assumption of spatial nonstationarity to be met (Charlton and Fotheringham, 2009). However, a limitation of GWR models is that they cannot be used to make estimations outside the study area (extrapolation) or for forecasting to novel periods, as doing so would violate the assumption of nonstationarity one is trying to meet (Osborne et al., 2007; Hothorn et al., 2011; Li et al., 2018). Since extrapolation and forecasted estimations are often desired when modeling species distributions, one recommended approach is to utilize multiple stationary models across a region of interest (Fotheringham et al., 2002; Windle et al., 2009). This approach will not only allow for extrapolation and forecasting procedures, but will also better account for assumptions of nonstationarity as using more than one model will result in multiple unique parameters estimated across localized areas.

Using American lobster in the GOM as a case study, we explore the effects of nonstationary modeling on lobster spatial distributions and compare the results to those of a stationary model. To test the effects of spatial nonstationarity, we develop season-, sex-, and size- specific models that predict the spatial distribution of American lobsters using GAMs of varying spatial scales and extents. Variation in spatial distribution between the models is evaluated and potential management implications are discussed.

MATERIALS AND METHODS

Study Area and Data Sources

American lobster abundance data were sourced from the Maine-New Hampshire Inshore Bottom Trawl Survey. The Maine-New Hampshire Inshore Bottom Trawl Survey will be referenced as

Abbreviations: FLFA, fall female adults; FLMA, fall male adults; FLFJ, fall female juveniles; FLMJ, fall male juveniles; SPFA, spring female adults; SPMA, spring male adults; SPFJ, spring female juveniles; SPMJ, spring male juveniles.

the bottom trawl survey. The bottom trawl survey is conducted by the Maine Department of Marine Resources (DMR) since the fall of 2000. This survey is semiannual, where separate surveys are conducted in the fall and spring seasons of each year. The bottom trawl survey spans 4,665 squared nautical miles (16000.5 km^2) (Sherman et al., 2005) and is subdivided into five regions (**Figure 1**). The five regions include (1) New Hampshire and Southern Maine, (2) Mid-Coast Maine, (3) Penobscot Bay, (4) Mt. Desert Island, and (5) Downeast Maine (**Figure 1**).

The survey area extends 12 nautical miles (22.22 km) offshore and is broken up into 4 different depth strata (**Figure 1**). A target of 115 stations is set for each survey, creating a sampling density of roughly one station for every 40 NM^2 (137.20 km^2). Random stations in this survey are chosen by dividing the survey area into a 1 NM^2 (3.43 km^2) grid, where cells are chosen at random using an Excel random number generator (Sherman et al., 2005). Each survey aims for a target tow of 20 min at a speed of 2.2–2.3 knots (4.1–4.3 km/h), which covers approximately 0.8 NM (1.48 km, Sherman et al., 2005). Data from 486,971 individual lobsters were included in this study. See **Supplementary Figures 1, 2** for mean catch trends in the bottom trawl survey data by region.

This study utilizes data from the 2000–2019 bottom trawl surveys. Biological data taken on each lobster include carapace length (mm), sex, presence of eggs or v-notches, and if any noticeable damage is present. Lobsters are then sorted into baskets by sex and baskets are weighed once filled (Sherman et al., 2005). Data have been standardized to 20-min tows to ensure all catch, weight, and length frequency information is comparable. In addition to biological data, bottom water salinity, bottom water temperature, and depth data were collected during each tow by using a Sea-Bird Electronics^{TM} 19plus SEACAT profiler, which was attached to the starboard door wire, turned on and lowered overboard (Sherman et al., 2005). Bottom trawl survey bottom temperature and bottom salinity data were recorded at a single point along each tow transect and do not represent an average across each tow length, The net used for this survey is a type of modified shrimp net that is used for "near-bottom dwelling species," although not intended for any single species in particular (Sherman et al., 2005). More information about the Maine-New Hampshire Inshore Bottom Trawl survey procedures, protocols, or specifics can be found in Sherman et al. (2005). This survey has been found to yield informative data for studying lobster distributions and habitats in the GOM (Tanaka and Chen, 2016; Tanaka et al., 2019; Hodgdon et al., 2020).

Bottom water temperature, bottom water salinity, average depth, latitude, and longitude information from each tow were used from the bottom trawl survey to inform the models. Distance from shore and median sediment size were also estimated and included in the models. Distance from shore was estimated using the "distances" function from the package "distances" (Savje, 2019) in R, which finds the shortest distance between points, in this case, the distance between the midpoint latitude and longitude of a tow and the closest point on the coast. Sediment data were sourced from the East Coast Sediment Texture Database which is run by the United States Geological Survey (U.S. Geological Survey, 2014). This survey was last updated in 2014 and contains information such as location, description, texture, and size (phi, −log of grain size) taken by different marine sampling programs across various locations around the world. Both mean and median sediment size values are supplied in this dataset, but median sediment size was used over mean sediment size, as the former is more robust to outliers (Tůmová et al., 2019). The median grain size at each survey location was estimated using thin plate splines. These data can be found at https://woodshole.er.usgs.gov/openfile/of2005-1001/htmldocs/datacatalog.htm and more information about the East Coast Sediment Texture Database can be found in U.S. Geological Survey (2014).

Although models were built using bottom trawl survey bottom temperature and bottom salinity data, additional bottom temperature and bottom salinity data were needed to create interpolated distribution plots. These additional data were not used to inform the models, but rather served as data that the models used to be able to estimate lobster density at unsampled locations. Thus, bottom temperature and bottom salinity data throughout the study area were obtained by spatially interpolating Finite-Volume Community Ocean Model (FVCOM) data. The FVCOM is an advanced ocean circulation model that uses an unstructured grid format, making it highly applicable for use in regions with complex coastlines and bathymetry (Chen et al., 2006; Li et al., 2017). The FVCOM was developed by University of Massachusetts Dartmouth and Woods Hole Oceanographic Institution. More information about the FVCOM can be found in Chen et al. (2006).

Forecasted distributions were made for the period 2028–2055. The forecasted bottom temperature and bottom salinity data were sourced from the National Oceanic and Atmospheric Administration (NOAA) and represent an ensemble projection of all models used to create the Intergovernmental Panel on Climate Change's (IPCC) Coupled Model Intercomparison Project Phase 5 (CMIP5) data (NOAA Physical Science Laboratory, n.d.). Data for the Representative Concentration Pathway (RCP) 8.5 "business as usual" scenario were used. These data are forecasted *anomalies* based on the reference time period 1956–2005 and are estimated for the period 2006–2055. These data are anomalies, and thus hindcasted bottom temperature and bottom salinity data must be used in tandem from the same reference period. The anomalies were added to the corresponding reference period FVCOM data. However, the earliest available FVCOM data begins in 1978 rather than 1956, limiting the available reference period in this study to 1978–2005. With the reference period reduced from 50 to 27 years, the CMIP5 forecasting period must also be reduced, respectively, from the initial 2006–2055 to 2028–2055 for this study. The forecasting period 2028–2055 is used because it represents the maximum amount of FVCOM data than can be used while also confidently applying IPCC forecasted anomalies. Delta downscaling methods were also applied so that forecasted anomalies could be applied to the same scale as the FVCOM data. Specifically, bivariate spline interpolation was applied using the package "akima" in R (Akima and Gebhardt, 2016).

Model Development

Lobster densities were standardized per tow and divided into eight groups based on season (fall and spring), sex (female

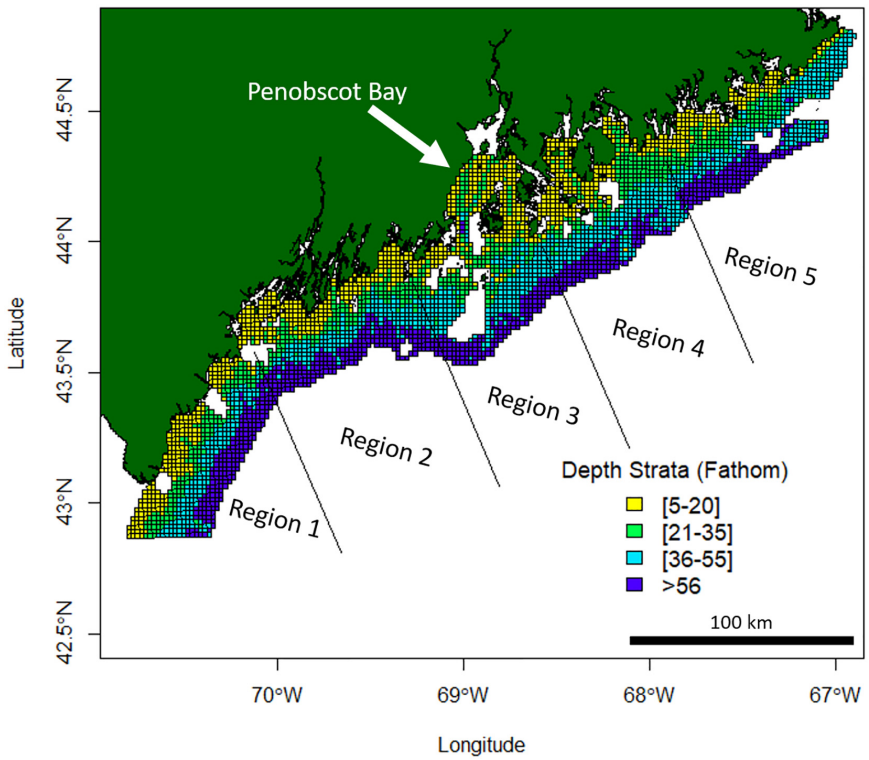

FIGURE 1 | Maine-New Hampshire Inshore Bottom Trawl Survey regions and depth strata. This survey is subdivided into five regions which include (1) New Hampshire and Southern Maine, (2) Mid-Coast Maine, (3) Penobscot Bay, (4) Mt. Desert Island, and (5) Downeast Maine. Missing white areas within the 12-mile survey grid area are non-surveyable locations due to the topography of the ocean floor at those locations.

and male), and size (adult and juvenile; Li et al., 2018; Chang et al., 2010). Juvenile lobsters were distinguished as lobsters with carapace lengths <50 mm due to differences in activity patterns (Lawton and Lavalli, 1995). Each of the eight groups were modeled independently under three different techniques: (1) a GAM that assumes stationary relationships between species distributions and environmental variables (GOM–GAM); (2) a GAM that assumes nonstationary relationships between eastern and western GOM (West-GAM, East1-GAM), and (3) a GAM that assumes nonstationary relationships between eastern, central, and western GOM (West-GAM, Central-GAM, and East2-GAM). Partitioning of data for these models can be visualized in **Figure 2**.

Previous literature in the GOM have estimated species distributions using stationary models at a large spatial scale (Chang et al., 2010; Becker et al., 2020). This technique is represented in this study by the "GOM–GAM" model, which assumes no nonstationarity and is applied at the largest spatial scale. This technique also assumes that nonlinear (but stationary) relationships between lobster density and environmental factors are sufficient to accurately predict a species spatial distribution across an ecologically complex region. Other literature has highlighted differences in environment-abundance relationships between localized regions (Li et al., 2018; Liu et al., 2019). Thus, the bisected (comprized of West-GAM and East1-GAM) and trisected (comprized of West-GAM, Central-GAM, and East2-GAM) models were constructed at smaller spatial scales to capture evidence of these differences. The purpose of this study is to explore how spatial distribution predictions change under models with varying assumptions of nonstationarity (or lack thereof) in hindcasting and forecasting scenarios.

The first set of localized models (West-GAM and East1-GAM) broke up the data into east and west zones. The West-GAM used data that were west of −69.27457 degrees longitude. The East1-GAM was represented by data east of −69.27457 degrees longitude. The decision to split the data up along the −69.27457 degree longitude line was in part because regions one and two of the bottom trawl survey are west of the Penobscot Bay region and −69.27457 is the approximate longitudinal line where region two of the bottom trawl survey intersects the coastline. This decision was also driven by the GOM coastal currents and the supporting literature that states the southern extent of the EMCC *includes* the Penobscot Bay region (Xue et al., 2008; Chang et al., 2016).

Although some literature supports this decision, it is difficult to exactly pinpoint a fine line of where the EMCC diverges and the WMCC begins. Thus, another argument can be made in which the Penobscot Bay area (≅region three in the bottom trawl survey) could act as a potential buffer zone, in which this

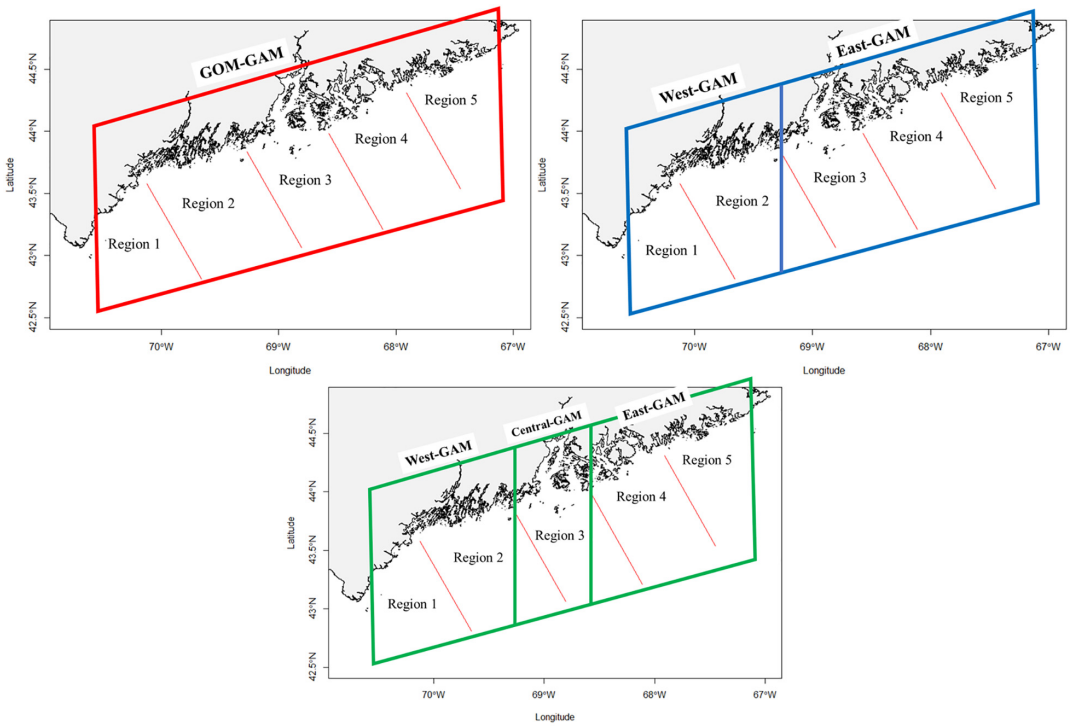

FIGURE 2 | Visual representation of each model utilized in this study. Each colored quadrilateral represents a separate GAM that was ran on the observed data points contained within that area. Red lines designate bottom trawl survey regional boundaries.

area of possible mixing between currents could throw off GAM relationship curves if the this area were to be included into a particular side. One previous study used a similar trisected approach to view relationships between initial intra-annual molts of American lobster and bottom temperatures in the GOM (Staples et al., 2019). Consequently, the second set of localized models (West-GAM, Central-GAM, and East2-GAM) are built in in such a way that the West-GAM is the same in spatial area and extent as previously described, but the Central-GAM is comprised of data between −69.27457 and −68.58246 degrees longitude, and the East2-GAM is comprised of data east of −68.58246 degrees longitude.

Prior to model construction, covariance matrices and variance inflation factor (VIF) tests were run to check for variable independence and multicollinearity. Running multiple covariance metrics showed a high dependence between distance from shore and average depth variables. Distance from shore was kept over average depth because distance from shore had a lower covariance value amongst the rest of the variables than average depth. VIFs quantify the multicollinearity amongst variables. Variables with VIF numbers >3 were excluded from the model (Zuur et al., 2009), supporting the decision to remove average depth as a variable when building the models. The following variables were shown to be significant in every GAM: latitude and longitude combined as an interaction term, and bottom temperature. Bottom salinity, distance from shore, and sediment size were found to be significant in some models, but not all. Significant variables and deviance explained for each group are summarized in **Tables 1**, **2**, respectively.

Generalized Additive Models were used to evaluate the relationships between lobster abundance and environmental variables. A GAM is an extension of a generalized linear model, with a smoothing function added. GAMs follow the assumptions that the functions are additive, and the components of the functions are smooth (Guisan et al., 2002). A separate GAM was created for each group of lobsters that differs in season, sex,

TABLE 1 | Non-Significant variables for each model and group type.

Group	GOM–GAM	East1-GAM	West-GAM	East2-GAM	Central-GAM
FLFJ	Salinity	Salinity	Sediment	Salinity	Salinity, DFS
FLMJ	AS	Salinity	Sediment	Salinity	Salinity
FLFA	Salinity	Salinity	Sediment	AS	Salinity, DFS, sediment
FMLA	Salinity	AS	Sediment	AS	Salinity, DFS
SPFJ	AS	AS	AS	AS	Salinity, sediment
SPMJ	AS	AS	Sediment	AS	Salinity, sediment
SPFA	AS	AS	AS	AS	Salinity, sediment
SPMA	AS	AS	AS	AS	Salinity, DFS

Group acronyms are denoted as follows: FL, fall; SP, spring; FJ, female juvenile; FA, female adult; MJ, male juvenile; MA, male adult. Such that for example FLFJ represents data taken from female juvenile lobsters in the fall season. Possible significant variables in each model include bottom temperature, bottom salinity, latitude and longitude, distance from shore, and sediment grain size. "AS", all significant; meaning all tested variables were significant to that particular model and group. "DFS", distance from shore variable. "Sediment", median sediment size variable; and "Salinity," bottom salinity variable.

TABLE 2 | Deviance explained for each model and group type.

Group	GOM–GAM	Average of West-East1-GAMs	East1-GAM	West-GAM	Average of West-Central-East2-GAMs	East2-GAM	Central-GAM
FLFJ	40.0%	52.5%	52.8%	52.2%	62.1%	62.3%	71.8%
FLMJ	40.7%	52.5%	52.8%	52.2%	63.2%	63.7%	73.6%
FLFA	42.6%	51.8%	47.9%	55.6%	56.3%	57.7%	55.7%
FLMA	41.7%	51.9%	49.2%	54.6%	55.6%	56.6%	55.5%
SPFJ	41.7%	51.9%	47.6%	56.1%	56.8%	48.5%	65.8%
SPMJ	44.0%	52.5%	50.5%	54.5%	58.1%	52.6%	67.1%
SPFA	34.4%	36.2%	35.0%	37.3%	40.9%	37.3%	48.2%
SPMA	38.8%	39.8%	41.8%	37.7%	44.5%	45.7%	50.6%

*See **Table 1** for group acronym explanation.*

and size, based on the assumption that males, females, juveniles, and adults will all respond to environmental variables differently, and that seasons will also impact the relationships with the environment differently. We used a tweedie GAM to estimate lobster abundance (y). GAMs were built using a backward fitting technique based on covariate significance ($p < 0.05$; Chang et al., 2010). A GAM using all potential environmental variables can be written as:

$$\text{Lobster abundance } (y) = s(La, Lo) + s(Bt) + s(Bs) + s(DFS) + s(Ss)$$

where s is a spline smoother, La, Lo is an interaction term between latitude and longitude, Bt is bottom temperature (°C), BS is bottom salinity (PSU), DFS is distance from shore (decimal degrees), and Ss is median sediment size (phi).

Hindcasted distribution plots were created for each lobster season × sex × size group and for each model approach for the years 2000, 2006, 2012, and 2017 for a total of 96 plots. Although there are bottom trawl survey data available from 2000–2019 to inform the models, environmental FVCOM data used for interpolation are only available until 2017, limiting the most recent available hindcasting year that can be spatially interpolated to 2017. Additionally, these years were chosen because they are roughly evenly spaced throughout the hindcast period of interest, albeit these methods could be applied to any year(s) 2000–2017. Forecast distribution plots were also estimated for the 2028–2055 years period, for a total of 24 forecast distribution plots (eight lobster groups ×3 model approaches). Model fitting was accomplished by using all survey data between the years 2000–2019 and predictions were estimated for each tested year (2000, 2006, 2012, 2017, and the forecast period 2028–2055) separately, by using the corresponding annual FVCOM data. Differences between GOM–GAM and localized approaches were determined by calculating relative differences between density distribution estimates. Relative differences were estimated using the equation

$$\text{Relative difference } (i) = \frac{\text{localized estimated density } (i) - \text{“}GOM - GAM\text{” estimated density } (i)}{\text{“}GOM - GAM\text{” estimated density } (i)} \times 100$$

where i is a location within the study area and “localized” represents the estimated lobster density at location i from a localized model (West-GAM, East2-GAM, etc.). Relative difference plots were generated for each lobster season × sex × size group and for the same years as the hindcast and forecast distribution plots. These plots demonstrate the magnitude and location of where the GOM–GAM models tend to over or under predict abundances in relation to the localized approaches. All distribution and relative difference plots were interpolated using bivariate splines using the package “akima” in R in order to achieve high resolution smooth distributions (Akima and Gebhardt, 2016).

Model Fitting and Validation

Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), and Moran’s I were used to access model fit for all models. RMSE measures the differences between predicted and observed values where values closer to zero represent better model fit (Stow et al., 2009). AIC is another method to test goodness of fit and model complexity with a model having smaller returned AIC value being the better model (Zuur et al., 2009). Moran’s I tests for spatial autocorrelation in residuals where a significant Moran’s I of −1 signifies perfect clustering of dissimilar values, a significant Moran’s I value of 0 signifies no autocorrelation, and a significant Moran’s I of +1 signifies perfect clustering of similar values. If values are found to be spatially autocorrelated, this is an issue as it violates the assumption of independence of data (Zuur et al., 2009; Stephanie, 2016). Additionally, two-fold cross validation was performed by separating each of the eight groups’ (2 season × 2 sexes × 2 sizes) data into random training and a testing subset to calibrate the model and validate its predictions (Li et al., 2018). The percentage of data allocated for the testing portion was determined by the equation

$$1/(1 + \sqrt{P - 1})$$

where P is the number of predictor variables (Franklin, 2010; Li et al., 2018). Cross validation allows visualization of model performance to examine if model predictions are on average, over or under predicting abundance compared to observed values. 100 iterations of cross validation were repeated for each model group and average performance was estimated.

RESULTS

Model Performance and Validation

Significant variables differed between model types and between groups. Under the GOM–GAM, only salinity was found to be non-significant in some groups, whereas both salinity and sediment size were found to be non-significant in some West-GAM and East1-GAM groups. Moreover, salinity, sediment, and distance from shore were found to be non-significant in some West-GAM, Central-GAM, and East2-GAM groups. **Table 1** summarizes the non-significant variables which were not included in the final model for each group and spatial scale. The deviance explained for lobster abundance varied between 34.4 and 44.0% for each group of the GOM–GAM, 36.2–52.5% for the average West-GAM and East1-GAM group, and 40.9–63.2% for the average West-GAM, Central-GAM, and East2-GAM group. Full deviance explained for each specific group can be found in **Table 2**. Likewise, the RMSE, AIC and Moran's I tests showed similar trends in model fit, with the GOM–GAM demonstrating the lowest model fit estimates, the West- and East1-GAM approach demonstrating intermediate model fits, and the West-, Central-, and East2-GAM approach demonstrating the greatest model fits (**Table 3**).

The two-fold cross validation results from 100 iterations revealed that the models had reasonable prediction skill, as the average between the 100 iterations was near the 1:1 prediction line for most groups and models. These tests revealed that most models tended to slightly underpredict abundance, with exception of the average spring female adult (SPFA) West- and East1-GAM approach which revealed average slight overpredictions. The West-, Central-, and East2-GAM approach cross validation results demonstrated more precision than West- and East1-GAM or GOM–GAM results. Results from the two-fold cross validation can be found in the **Supplementary Material** section (**Supplementary Figures 1–3**).

Environmental and Spatial Variables

Environmental and spatial variables were also explored via GAM response curves for each significant predictor variable. Latitude and longitude variables were combined as an interaction term in each model to help account for spatial autocorrelation (Siegel and Volk, 2019). Response curves varied greatly depending on independent variable, season, sex, size, and spatial scale of the model. For bottom temperature, highest partial effect on abundance was seen between 6 and 10°C in the spring and around 10–14°C in the fall for GOM–GAMs, and between 4 and 10°C in the spring and 10–14°C in the fall for the localized model approaches. For bottom salinity, highest abundance was seen between 31 and 33 psu for both spring and fall across all models. The relationship spring male adult (SPMA), spring female juvenile (SPFJ), and spring male juvenile (SPMJ) groups had with salinity was unique, compared to other groups. These group's response curves demonstrated a higher partial effect on abundance at salinity levels >32 psu in the west. This may help explain the distinctive relative difference trends generally observed in western GOM for the SPMA group. This difference did not seem to affect the spring juvenile groups, as juvenile lobster tend to stay in more nearshore waters (Lawton and Lavalli, 1995), where FVCOM data has shown salinity levels are generally lower in western GOM. For distance offshore, highest partial effect on abundance was seen generally between 0.00 and 0.1 decimal degrees (≅0–6 nautical miles offshore), and then gradually declined with increasing distance from shore across most models. For sediment size, highest partial effect on abundance was seen between 2 and 6 phi (silt – medium grain sand) across most models. Some season, sex, and size group curves changed more in shape across spatial extents than others, but variation was apparent and supports evidence of spatial nonstationarity in this region. **Figure 3** depicts the response curves between lobster abundance and bottom temperature for SPMAs (**Figures 3A,B**) and fall female juveniles (**Figures 3C,D**). These figures show how the response curves change, depending on the spatial scale and location of the testing data. These figure panels also show where estimated relationship curves overlap, if at all. For example, in **Figure 3B**, one can see high overlap between most model response curves between 5 and 7°C. However, at temperatures greater than 7°C, the relationship curve for the GOM–GAM more closely resembles that of the response curve for the East2-GAM than for the West- or Central-GAM. This suggests that if a large-scale model were used to represent SPMA lobster data, it would better represent eastern GOM data than central or western GOM data in that temperature range, and in a climate warming scenario, would underestimate western GOM abundances. In a region which is expected to continue experiencing warming temperatures, the implications of subordinate model spatial scale selection may increase. Many lobster groupings (season × sex × size) tended to show similar patterns, where the GOM–GAM response curve for a variable, more closely resembled the response curve of one localized region of the GOM more than the other regions.

Model Prediction and Distribution Plots

Fall distribution plots showed greater abundance estimates than spring plots, which correlates with observations in raw trawl survey data. Raw fall trawl survey trends show slight declines in catch in regions three and four since 2015 and in region five since 2016 (**Supplementary Figure 4**), with trends of offshore catch increasing overtime. All three model estimates demonstrated offshore abundance estimates increasing from the 2012–2017 hindcasts, but only the East2-GAM showed indications of a slight decrease in eastern GOM abundance. Model estimates in central GOM were most distinctive between models. A trend emerged in all tested years which demonstrated that as model spatial scale became finer, clear "hot" and "cold" spots emerged within the Penobscot Bay area. The Central-GAM showed this pattern well, with a "hotspot" emerging along the southwest mouth of Penobscot Bay, and a "coldspot" in the northeast Penobscot Bay region (**Figures 4–7**). These patterns correlate well with American lobster settlement patterns found in Steneck and Wilson (2001).

The GOM–GAM tended to overpredict the 2017 hindcast distributions in western GOM, apart from the SPMA group (**Figure 8**). In central GOM, the GOM–GAM models tended to

TABLE 3 | RMSE, Moran's I, and AIC values for each model and group type.

Group	GG RMSE	WEG RMSE	WCEG RMSE	GG Moran's I	WEG Moran's I	WCEG Moran's I	GG AIC	Average WEG AIC	Average WCEG AIC
FLFJ	1.67	1.53	1.44	0.51	0.42	0.16	9,009	4,341	2,757
FLMJ	1.67	1.54	1.43	0.49	0.38	0.14	8,978	4,327	2,736
FLFA	1.29	1.17	1.09	0.45	0.32	0.07	14,398	7,064	4,585
FLMA	1.24	1.12	1.06	0.43	0.30	0.07	14,428	7,069	4,597
SPFJ	1.68	1.57	1.51	0.51	0.41	0.17	10,256	4,950	3,121
SPMJ	1.67	1.58	1.52	0.46	0.37	0.15	10,011	4,884	3,060
SPFA	1.49	1.37	1.32	0.29	0.22	0.09	19,279	9,548	6,087
SPMA	1.41	1.32	1.28	0.28	0.22	0.09	19,124	9,480	6,055

"GG", "GOM–GAM"; "WEG", West- and East1-GAM approach; and "WCEG", West-, Central-, and East2-GAM approach. See ***Table 1*** *for group acronym explanation. RMSE values closer to zero represent better model fit. Moran's I tests for spatial autocorrelation in residuals where significant values closer to 0 signifies no autocorrelation. All reported Moran's I values were significant ($p < 0.05$). Smallest AIC values also indicate a better model.*

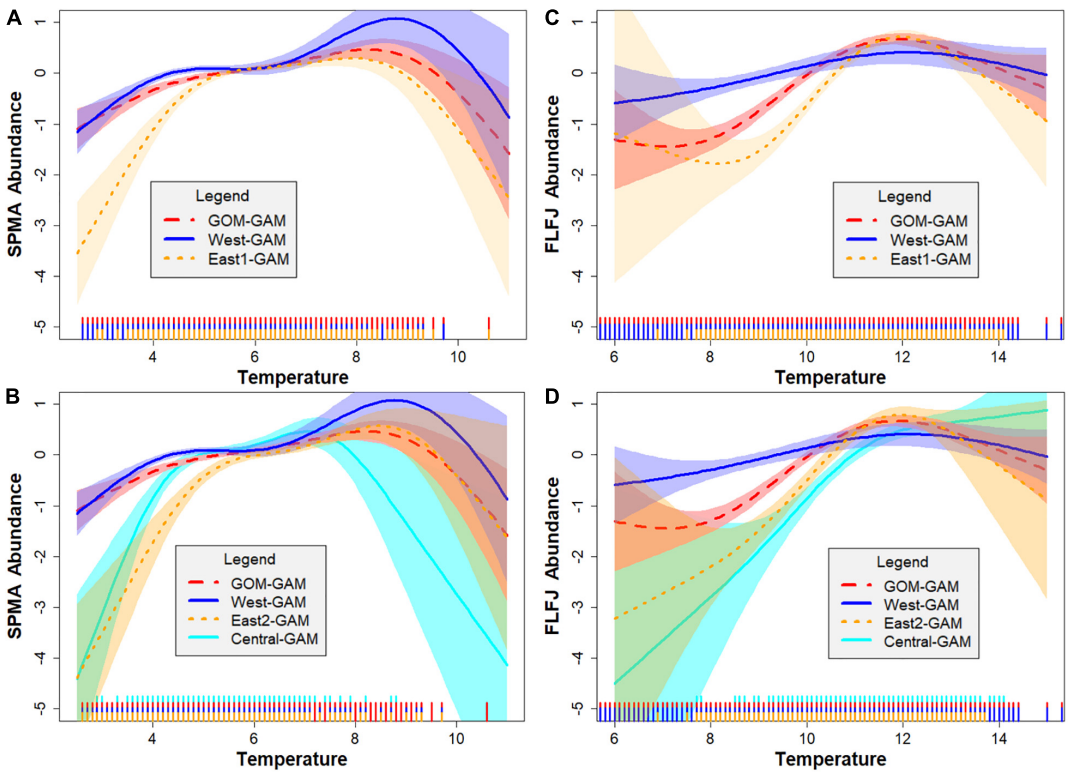

FIGURE 3 | A comparison of spring male adult (SPMA) and fall female juvenile (FLFJ) lobster GAM bottom temperature response curves by spatial location in the GOM. Each plot shows the response curve of bottom temperature (°C) on the *x*-axis, against the partial effect of lobster density on the *y*-axis. Figures **(A,C)** compare response curves estimated for the GOM–GAM, West-GAM, and East1-GAM, while figures **(B,D)** compare response curves estimated for the GOM–GAM, West-GAM, Central-GAM, and East2-GAM. Shaded regions on either side of the response curve line indicate the standard error confidence intervals. Rug plot lines along the *x*-axis of each plot indicate distribution of the bottom temperature data. These response curves were estimated using 2000–2019 ME-NH Inshore Bottom Trawl survey data and are applicable to all years tested.

comparatively underpredict in the western part of Penobscot Bay and overpredict in the eastern part of Penobscot Bay. This was evident across all years in both relative difference comparisons when the GOM–GAM estimates were compared to the West- and East1-GAM approach, as well as in the West-, Central-, and East2-GAM approach (**Figure 8**). In eastern GOM, many GOM–GAMs estimated less abundance approximately between −68.5° and −67.5° W, and higher abundance estimates between −67.5° and −67° W when compared to West- and East1-GAM approaches (**Figures 8**, **9**). These trends were present across all tested years.

Estimates for the 2028–2055 period from localized and large-scale approaches exemplify similar spatial patterns seen in the corresponding distributions from 2000 to 2017. Some season × sex × size groups estimated abundances that extend further offshore than their hindcast counterparts (see

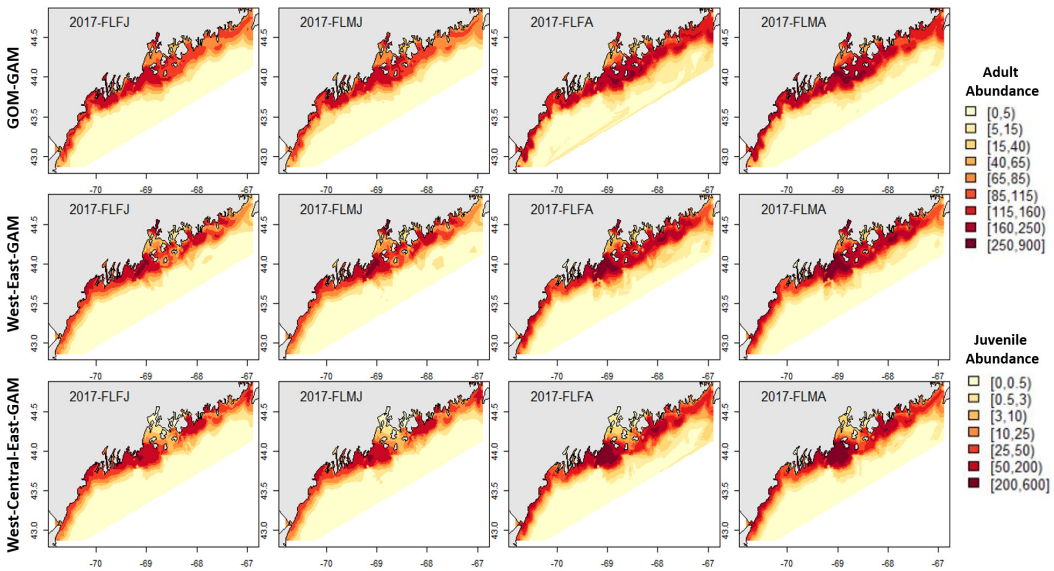

FIGURE 4 | 2017 Fall American lobster estimated spatial distribution. Legend colors increase in abundance estimates from pale yellow to dark red. Each column represents a season × sex × size group. Each row represents the modeling approach used to generate the abundance estimations. Adult abundance legend corresponds with adult lobster group estimates. Juvenile abundance legend corresponds with juvenile lobster group estimates.

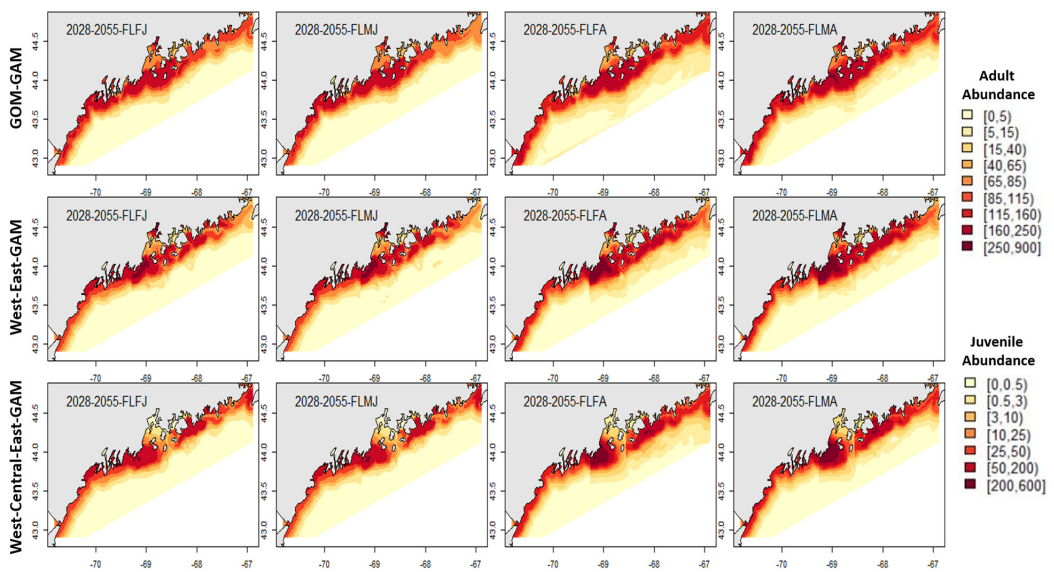

FIGURE 5 | Forecasted fall American lobster estimated spatial distribution for the time period 2028–2055. See **Figure 4** for figure details.

Figures 4–**7**). Spring abundance estimates demonstrate an increase in central and eastern GOM from 2017 to 2028–2055, although this is more notable in the localized models than the GOM–GAMs (see **Figures 6**, **7**). These forecasted estimates correlate with raw spring bottom trawl survey data thus far for regions 3–5, which have all demonstrated general increasing average catch rates (number/tow) from 2000–2019 (**Supplementary Figure 5**).

In general, relative differences between the GOM–GAM and the West-Central-East2-GAM approach resulted in larger differences when compared to the relative differences between the GOM–GAM and the West-East1-GAM approach. This trend was apparent across all tested years. These observations correlate with observations in model fit, as the West-Central-East2-GAM approach showed highest model fits, and the West-East1-GAM approach showed model fits more similar to that of the GOM–GAMs. Fall relative difference plots revealed that the GOM–GAMs were likely to estimate higher abundance in western GOM when compared to the West-GAM (**Figure 10**). In the spring, the GOM–GAM comparatively estimated lower abundance in

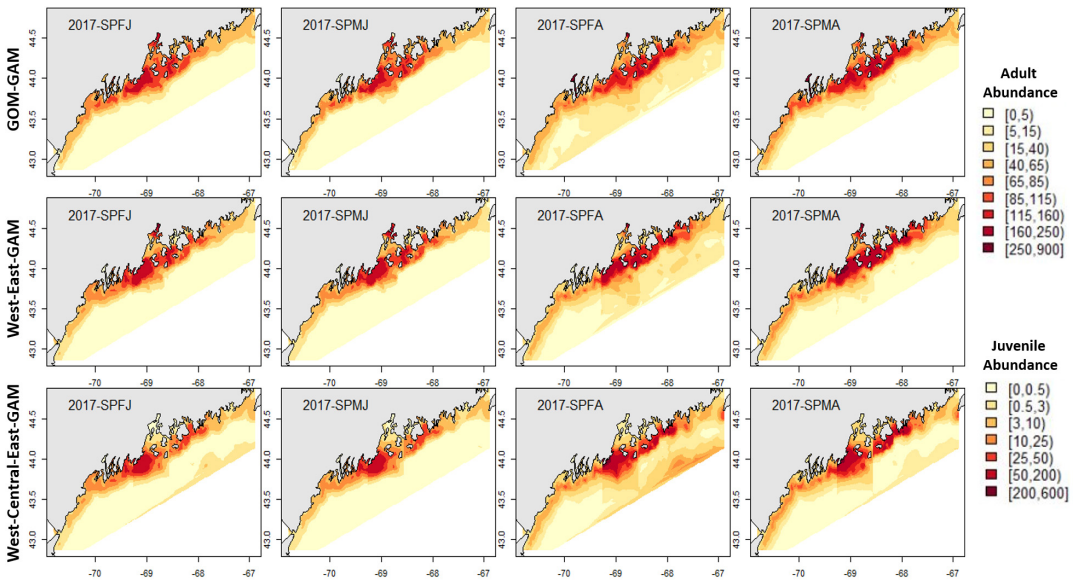

FIGURE 6 | 2017 spring American lobster estimated spatial distribution. See **Figure 4** for figure details.

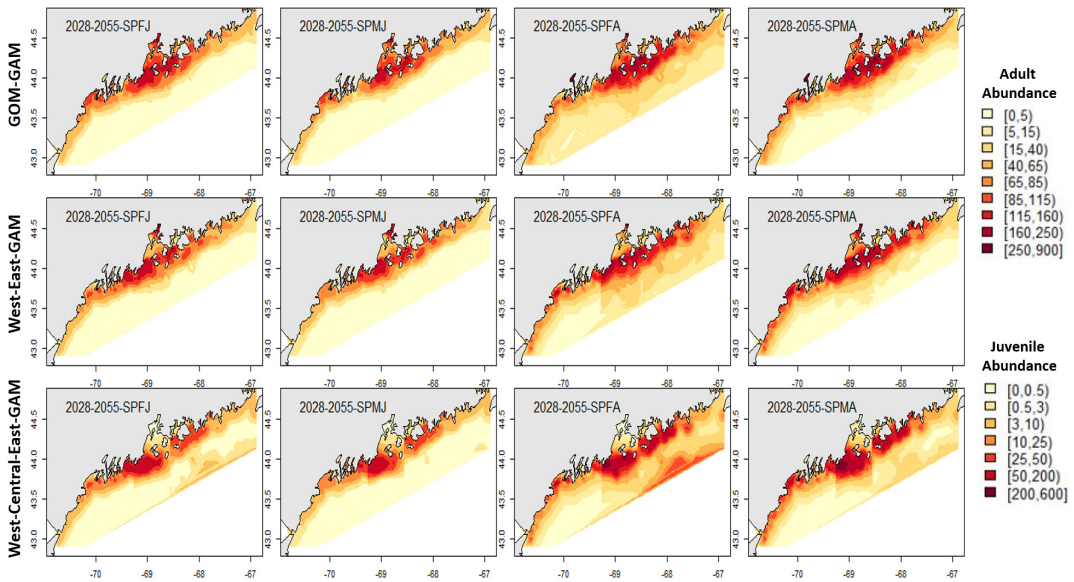

FIGURE 7 | Forecasted spring American lobster estimated spatial distribution for the time period 2028–2055. See **Figure 4** for figure details.

western GOM for spring adult males in the 2028–2055 period (**Figure 10**). For adult females in both fall and spring, however, GOM–GAMs estimated higher abundance in western GOM than the West-GAMs did in that same region for the 2028–2055 period (**Figure 10**). Forecasted GOM–GAM abundance plots estimated lower abundance in the western portion of central GOM (≈−69.3 to −68.9° W) and estimated higher abundance in the eastern portion of central GOM (≈−68.9 to −68.1° W), when compared with distribution estimates derived from the East1-GAM in that same area (**Figure 10**). This trend was also apparent in GOM–GAM and Central-GAM forecasted relative difference plots, but differences were slightly more polarized. There were slightly less consistent and patchy trends in relative differences amongst groups in eastern GOM for the 2028–2055 forecasted period, where both higher and lower estimates were evident (**Figure 10**).

DISCUSSION

We developed a modeling approach to explore and demonstrate how estimates of season-, sex-, and size- specific American lobster spatial distribution and abundance would vary based

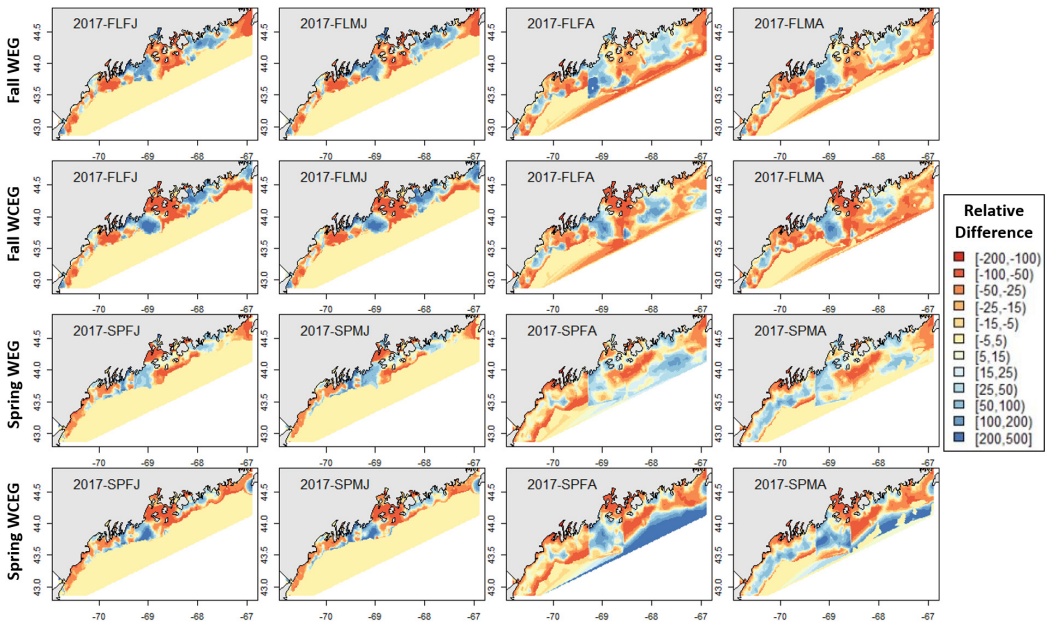

FIGURE 8 | 2017 American lobster relative differences in model abundance estimates. Legend numbers represent relative differences (%) between either the West-East1-GAM (WEG) or the West-Central-East2-GAM (WCEG) approach and the GOM–GAM. Red legend colors indicate areas where the GOM–GAM is predicting higher lobster abundance than the model in comparison. Blue legend colors indicate areas where the GOM–GAM is predicting lower lobster abundance then the localized model approach in comparison. Pale yellow colors indicate similar abundance estimates between the GOM–GAM and localized models. Each column represents a lobster season × sex × size group. Each row represents the season and localized model approach compared to the corresponding GOM–GAM.

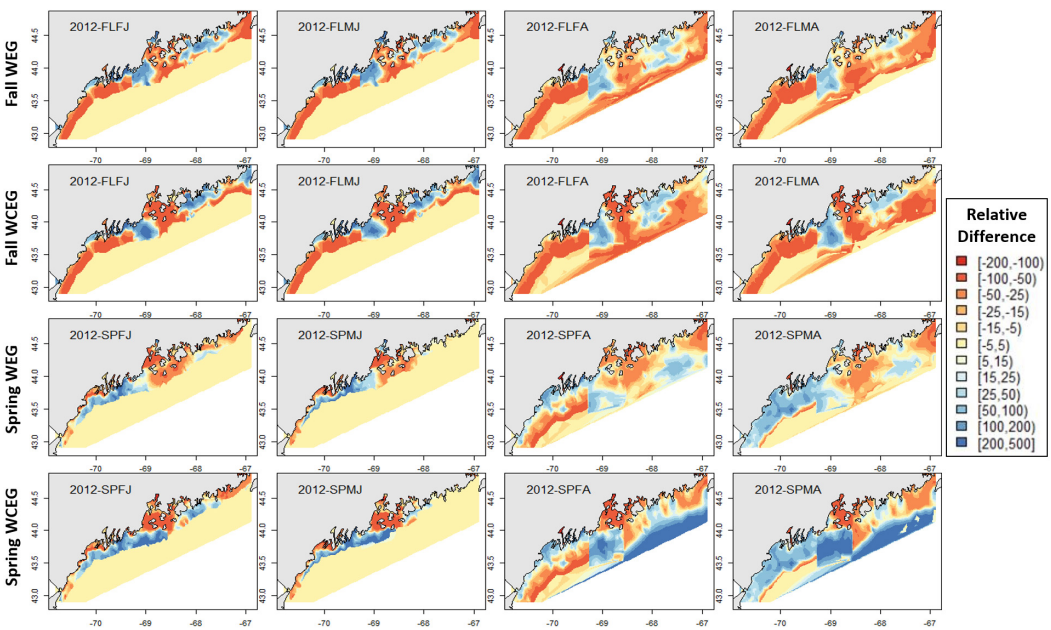

FIGURE 9 | 2012 American lobster relative differences in model abundance estimates. See **Figure 8** for figure details.

on the spatial scale and extent of the area being modeled in the GOM. Validation tests run for each model type and season × sex × size group suggested reasonable predictive ability. Nonsignificant variables varied by model and spatial location. These results correspond with the notion that local patterns may get masked by global statistics, if stationary assumptions are made (Brunsdon et al., 1996; Windle et al., 2012). Stationary assumptions are likely to be violated in the GOM, where

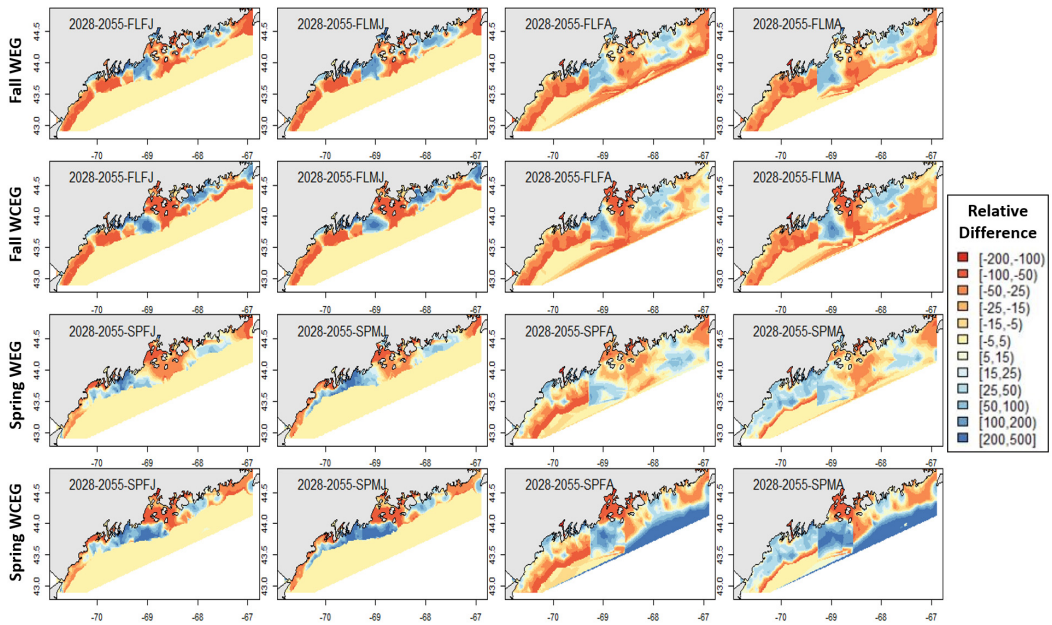

FIGURE 10 | Forecasted American lobster relative differences in model abundance estimates for the period 2028–2055. Legend numbers represent relative differences (%) between the West-East1-GAM (WEG) approach or the West-Central-East2-GAM (WCEG) approach and the GOM–GAM. Red legend colors indicate areas where the GOM–GAM model is predicting higher lobster abundance than the localized model approach in comparison. Blue legend colors indicate areas where the GOM–GAM model is predicting lower lobster abundance then the localized model approach in comparison. Pale yellow colors indicate similar abundance estimates between the large-scale and localized models. Each column represents a lobster season × sex × size group. Each row represents the season and model type compared to the corresponding GOM–GAM.

northeast to southwest gradients of bottom water temperature, salinity, and productivity have been observed (Lynch et al., 1997; Pettigrew et al., 1998; Chang et al., 2016), as well as spatial differences in American lobster stock-recruitment relationships (Chang et al., 2016), and spatially varying patterns in initial molt timing and suddenness (Staples et al., 2019).

A trend in model fit was observed in which as the spatial scale of models became more localized, model fit increased. The West-Central-East2-GAM approach demonstrated the greatest model fit to the bottom trawl survey data and showed the most correlation in abundance estimates with raw bottom trawl survey data, indicating greater distribution estimation capabilities. The West-East1-GAM approach demonstrated the next highest model fit and estimation capabilities, while the GOM–GAM model demonstrated the lowest model fit to the data. We speculate that the West-Central-East2-GAM approach shows the greatest model fit and potential predictive capabilities because of the modeling technique used on these data. By taking into consideration the oceanographic processes in the GOM to determine which localized areas are likely to be the most and least similar in relationships between American lobster abundance and environmental variables, the amount of data used for model estimation can be maximized, while limitations of large-scale models over a biologically complex region can be minimized. Out of the localized scale model approaches, the results of West-Central-East2-GAM approach suggest an improvement upon the West-East1-GAM approach. Although these approaches are similar, the evidence of the West-Central-East2-GAM approach being an improvement upon the West-East1-GAM approach suggests that enough nonstationary exists between central and eastern GOM to make the tripartite model subdivision worthwhile and that this technique may be more biologically reflective. Spatial distribution estimates of the West-Central-East2-GAM approach also seem to correlate well with raw bottom trawl survey data and past literature, especially in central GOM which has shown high increases in average catch over the course of the survey, and where localized "hot" and "cold" spots may be reflective of lobster settlement patterns observed in that region (Steneck and Wilson, 2001).

Most lobster groups demonstrated similar spatial patterns or temporal trends in model results and analysis, with the frequent exception of SPMA groups. We speculate the SPMA lobster groups often did not respond in the same way due to differences in both bottom temperature and bottom salinity response curves. Although each group had more than one significant environmental variable across model techniques, bottom temperature was a significant variable in all models, and spring adult (male and female) bottom temperature response curves were most distinct among groups. Most other season × sex × size groups displayed a relationship with bottom temperature was similar to that of the FLFJ group (**Figure 3D**), where the partial effect of bottom temperature on abundance generally increased then plateaued with increasing temperature. Spring adult lobster often did not follow this pattern, as exemplified in **Figure 3C**, where spring adult curves were typically domed-shaped. This dome-shaped pattern was

present in both female and male spring adult groups, however, so it is likely that other influences, such as salinity, may be a potential factor. The relationship spring adult males had with salinity was unique, compared to spring adult females, which demonstrated a similar pattern to the other season, sex, and size groups. The spring adult male group response curve demonstrated a higher partial effect on abundance at salinity levels >32 psu in the west, which may explain why the GOM–GAMs were more likely to comparatively underestimate lobster abundance in that region.

The West-Central-East2-GAM approach demonstrated the greatest relative differences across all years when comparing its spatial abundance predictions to those of the GOM–GAM. This observation is the result of the multiple unique GAMs run on localized data, and thus assumptions of spatial nonstationarity are better satisfied. However, it is important to recognize that the largest difference from the GOM–GAM does not automatically equate to the best model, as it is difficult to determine the starting biological accuracy of the GOM–GAM. Results from the three modeling techniques at bottom trawl survey locations could be compared to raw bottom trawl data at those same locations to get a better understanding of how biologically accurate each technique is at producing estimates. However, between evidence of model fit and validation, distribution plot results, and correlation with raw survey data, we conclude that applying model techniques that better account for spatial nonstationarity will result in increased model performance.

While the West-Central-East2-GAM approach demonstrated the best model fit out of the tested models, it is important to acknowledge some of the limitations of this model and the techniques used. First, all models tested only included environmental variables. No biological variables were included in the models; thus, these models are working under the assumption that lobster abundance is dependent solely upon environmental variables and spatial scale. Future studies may benefit from including biological variables, such as predator and/or prey abundance, into the models to see how the results would differ. Secondly, the subdivision of data technique used for the localized models (West-East1 and West-Central-East2 GAMs) sometimes resulted in variegated or "patchwork" spatial distribution estimates. Such abrupt changes in abundance estimates along the model extent lines are not likely to be biologically representative of true American lobster spatial distributions in the GOM. Consequently, this piecewise, localized modeling approach should only be used to observe trends in spatial distribution estimates, and not for precise estimations of "true" abundance, especially near the model extent lines. Thirdly, future studies may also benefit from exploring how different ways of subdividing data can impact model results, and if model fit can be further improved with more data partitions. Lastly, it is likely that the relationships explored in this study do not only vary across space, but over time as well. This study only considers spatial nonstationarity in model development, as gradients in environmental conditions throughout the study area have been observed. We did not consider temporal autocorrelation in this study. Based on the results from this study, it is likely that accounting for temporal autocorrelation could impact species distribution results as well and that excluding temporal autocorrelation may have introduced biases into the forecasts. However, this is beyond the scope of this study.

This study indicates that SDM estimations are dependent upon spatial scale and assumptions of nonstationarity. Results from a model that implicitly assumes spatial stationarity would differ from results of a model that better accounts for spatial nonstationary processes. Thus, using results generated by large-scale, stationary models could lead to different, or potentially even ill-informed management decisions which may result in less effective management results. Moreover, accounting for localized processes may be essential when devising localized regulations, as indications of change or unique dependencies of a species may be masked when using global statistics. Management decisions informed by large-scale, stationary models could result in regulations being more effective in one local area and less in others, if the relationship curves that drive the predictions are more representative of a particular area of the study area, rather than well represented throughout. If the West-Central-East2-GAM model distribution estimates are more biologically realistic as the analyses suggest, then comparatively, under an RCP 8.5 "business as usual" climate scenario prediction for the years 2028–2055, large-scale, stationary models could overestimate lobster abundances in western GOM, with the exception of spring adult males. In such case, it is important local heterogeneity is considered in American lobster management in the GOM because false overestimations of abundance could lead to relaxed regulations or ill-informed biological reference point calculations, which could potentially lead to overfishing in western GOM.

Using large-scale, stationary modeling techniques to forecast American lobster spatial distribution could result in subordinate perceptions of where lobster populations will be spatially, and to what extent. More accurate predictions of American lobster spatial distributions will help stakeholders prepare and employ best practice measures to ensure the sustainability and longevity of the industry.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Multiple datasets were utilized in this study. Most are publicly available, but the ME-NH Inshore Bottom Trawl Survey Data are not. Publicly available data have been compiled and can be found at https://github.com/Jamie-Behan/Nonstationary-Effects-on-Spatial-Distribution.git. Requests to access these datasets should be directed to Jamie Behan, jamie.behan@maine.edu.

AUTHOR CONTRIBUTIONS

YC drafted the project conception. BL supplied some foundational code. JB substantially contributed to the finalized code which was used. JB accomplished interpretation of data, with support from BL and YC. JB completed the manuscript drafting, with manuscript revisions by BL and YC. All authors drafted the project design, and approve the submitted manuscript.

FUNDING

This work was funded by the Maine Department of Marine Resources.

ACKNOWLEDGMENTS

We thank the Maine Department of Marine Resources for funding this work and Rebecca Peters from the Maine Department of Marine Resources for providing Maine-New Hampshire Inshore Bottom Trawl Survey data and valuable comments. We would also like to thank Cameron Hodgdon and Mackenzie Mazur for their insight and assistance in problem troubleshooting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.680541/full#supplementary-material

REFERENCES

ACCSP (2019). *ACCSP Data Warehouse Non-Confidential Commercial Landings. Data Warehouse, Comprehensive Landings Data for the Atlantic Coast*. Arlington, VA: Atlantic Coastal Cooperative Statistics Program.

Akima, H., and Gebhardt, A. (2016). *akima: Interpolation of Irregularly and Regularly Spaced Data. (R package version 0.6-2) [Computer software]*.

ASMFC (2020). *2020 American Lobster Benchmark Stock Assessment ad Peer Review Report*. Arlington, VA: Atlantic States Marine Fisheries Commission, 548.

Bakka, H., Vanhatalo, J., Illian, J., Simpson, D., and Rue, H. (2016). Accounting for physical barriers in species distribution modeling with non-stationary spatial random effects. *arXiv: Appl.* 1–21.

Becker, E. A., Carretta, J. V., Forney, K. A., Barlow, J., Brodie, S., Hoopes, R., et al. (2020). Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. *Ecol. Evol.* 10, 5759–5784. doi: 10.1002/ece3.6316

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geograph. Anal.* 28, 281–298. doi: 10.1111/j.1538-4632.1996.tb00936.x

Chang, J.-H., Chen, Y., Halteman, W., and Wilson, C. (2016). Roles of spatial scale in quantifying stock-recruitment relationships for American lobsters in the inshore Gulf of Maine. *Can. J. Fish. Aquat. Sci.* 73, 885–909.

Chang, J.-H., Chen, Y., Holland, D., and Grabowski, J. (2010). Estimating spatial distribution of American lobster *Homarus americanus* using habitat variables. *Mar. Ecol. Progress Ser.* 420, 145–156. doi: 10.3354/meps08849

Charlton, M., and Fotheringham, A. S. (2009). *Geographically Weighted Regression: White Paper*. Dublin: Science Foundation Ireland.

Chen, C., Cowles, G., and Beardsley, R. (2006). An unstructured grid, finite-volume coastal ocean model (FVCOM) system. *Oceanography* 19, 78–89. doi: 10.5670/oceanog.2006.92

Diarra, M., Fall, M., Fall, A. G., Diop, A., Lancelot, R., Seck, M. T., et al. (2018). Spatial distribution modelling of *Culicoides* (Diptera: Ceratopogonidae) biting midges, potential vectors of African horse sickness and bluetongue viruses in Senegal. *Parasit. Vectors* 11:341.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The analysis of Spatially Varying Relationships*. New York, NY: John Wiley & Sons, Ltd.

Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511810602

Greenan, B. J. W., Shackell, N. L., Ferguson, K., Greyson, P., Cogswell, A., Brickman, D., et al. (2019). Climate change vulnerability of american lobster fishing communities in Atlantic Canada. *Front. Mar. Sci.* 6:579. doi: 10.3389/fmars.2019.00579

Guisan, A., Edwards, T. C., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.* 157, 89–100. doi: 10.1016/S0304-3800(02)00204-1

Hastie, T., and Tibshirani, R. (1986). Generalized additive models. *Stat. Sci.* 1, 297–310.

Hodgdon, C. T., Tanaka, K. R., Runnebaum, J., Cao, J., and Chen, Y. (2020). A framework to incorporate environmental effects into stock assessments informed by fishery-independent surveys: a case study with American lobster (*Homarus americanus*). *Can. J. Fish. Aquat. Sci.* 77, 1700–1710. doi: 10.1139/cjfas-2020-0076

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

Incze, L., Xue, H., Wolff, N., Xu, D., Wilson, C., Steneck, R., et al. (2010). Connectivity of lobster (*Homarus americanus*) populations in the coastal Gulf of Maine: part II. coupled biophysical dynamics. *Fish. Oceanogr.* 19, 1–20. doi: 10.1111/j.1365-2419.2009.00522.x

Lawton, P., and Lavalli, K. L. (1995). "Postlarval, juvenile, adolescent, and adult ecology," in *Biology of the Lobster*, ed. J. R. Factor (San Diego, CA: Academic Press), 47–88. doi: 10.1016/B978-012247570-2/50026-8

Li, B., Cao, J., Guan, L., Mazur, M., Chen, Y., and Wahle, R. A. (2018). Estimating spatial non-stationary environmental effects on the distribution of species: a case study from American lobster in the Gulf of Maine. *ICES J. Mar. Sci.* 75, 1473–1482. doi: 10.1093/icesjms/fsy024

Li, B., Tanaka, K. R., Chen, Y., Brady, D. C., and Thomas, A. C. (2017). Assessing the quality of bottom water temperatures from the finite-volume community ocean model (FVCOM) in the Northwest Atlantic shelf region. *J. Mar. Syst.* 173, 21–30. doi: 10.1016/j.jmarsys.2017.04.001

Liu, C., Liu, J., Jiao, Y., Tang, Y., and Reid, K. B. (2019). Exploring spatial nonstationary environmental effects on Yellow Perch distribution in Lake Erie. *PeerJ* 7:e7350. doi: 10.7717/peerj.7350

Lynch, D. R., Holboke, M. J., and Naimie, C. E. (1997). The maine coastal current: spring climatological circulation. *Cont. Shelf Res.* 17, 605–634. doi: 10.1016/S0278-4343(96)00055-6

National Oceanographic and Atmospheric Administration (NOAA) (2018). *Fisheries of the United States*. Silver Spring, MD: National Oceanographic and Atmospheric Administration.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A (General)* 135:370. doi: 10.2307/2344614

NOAA Physical Science Laboratory (n.d.). *Climate Change Web Portal _ Maps MM: NOAA Physical Sciences Laboratory*. Available online at: https://psl.noaa.gov/ipcc/ocn/ (accessed January 13, 2021).

Osborne, P. E., Foody, G. M., and Suárez-Seoane, S. (2007). Non-stationarity and local approaches to modelling the distributions of wildlife. *Divers. Distrib.* 13, 313–323. doi: 10.1111/j.1472-4642.2007.00344.x

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science* 350, 809–812. doi: 10.1126/science.aac9819

Pettigrew, N., Townsend, D., Xue, H., Wallinga, J. P., Brickley, P. J., and Hetland, R. D. (1998). Observations of the eastern maine coastal current and its offshore extensions in 1994. *J. Geophys. Res. Oceans* 103:30623. doi: 10.1029/98JC01625

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science (American Association for the Advancement of Science)* 341, 1239–1242. doi: 10.1126/science.1239352

Savje, F. (2019). *Tools for Distance Metrics, Package: Distances (0.1.8) [R]*.

Sherman, S., Stepanek, K., and Sowles, J. (2005). Maine-New Hampshire Inshore Groundfish Trawl Survey: Procedures and Protocols. Maine Department of Marine Resources. Available online at: https://www.maine.gov/dmr/science-research/projects/trawlsurvey/reports/documents/proceduresandprotocols.pdf (accessed June 24, 2020).

Siegel, J. E., and Volk, C. J. (2019). Accurate spatiotemporal predictions of daily stream temperature from statistical models accounting for interactions between climate and landscape. *PeerJ* 7:e7892. doi: 10.7717/peerj.7892

Staples, K. W., Chen, Y., Townsend, D. W., and Brady, D. C. (2019). Spatiotemporal variability in the phenology of the initial intra-annual molt of American lobster (*Homarus americanus* Milne Edwards, 1837) and its relationship with bottom temperatures in a changing Gulf of Maine. *Fish. Oceanogr.* 28, 468–485. doi: 10.1111/fog.12425

Steneck, R. S., and Wilson, C. J. (2001). Large-scale and long-term, spatial and temporal patterns in demography and landings of the American lobster, *Homarus americanus*, in Maine. *Mar. Freshw. Res.* 52, 1303–1319. doi: 10.1071/mf01173

Stephanie. (2016). *Moran's I: Definition, Examples. Statistics How To*. Available online at: https://www.statisticshowto.com/morans-i/ (accessed August 25, 2016)

Stow, C., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, I., Friedrichs, M., et al. (2009). *Skill Assessment For Coupled Biological/Physical Models of Marine Systems. OpenAIRE - Explore*. Available online at: https://explore.openaire.eu//search/publication?articleId=od_____267::5c31b0c05a25082338dee4e62a910e70 (accessed October 20, 2020).

Tanaka, K., and Chen, Y. (2016). Modeling spatiotemporal variability of the bioclimate envelope of Homarus americanus in the coastal waters of Maine and New Hampshire. *Fish. Res.* 177, 137–152. doi: 10.1016/j.fishres.2016.01.010

Tanaka, K. R., Chang, J.-H., Xue, Y., Li, Z., Jacobson, L., and Chen, Y. (2019). Mesoscale climatic impacts on the distribution of *Homarus americanus* in the US inshore Gulf of Maine. *Can. J. Fish. Aquat. Sci.* 76, 608–625. doi: 10.1139/cjfas-2018-0075

Townsend, D. W., Pettigrew, N. R., Thomas, M. A., Neary, M. G., McGillicuddy, D. J., and O'Donnell, J. (2015). Water masses and nutrient sources to the gulf of maine. *J. Mar. Res.* 73, 93–122. doi: 10.1357/002224015815848811

Tůmová, Š, Hrubešová, D., Vorm, P., Hošek, M., and Grygar, T. M. (2019). Common flaws in the analysis of river sediments polluted by risk elements and how to avoid them: Case study in the Ploučnice River system, Czech Republic. *J. Soils Sediments* 19, 2020–2033. doi: 10.1007/s11368-018-2215-9

U.S. Geological Survey (2014). *ECSTDB2014.SHP: U.S. Geological Survey East Coast Sediment Texture Database (2014) (No. 2005–1001)*. Reston, VA: U.S Geological Survey.

Windle, M. J. S., Rose, G. A., Devillers, R., and Fortin, M.-J. (2009). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES J. Mar. Sci.* 67, 145–154. doi: 10.1093/icesjms/fsp224

Windle, M. J. S., Rose, G. A., Devillers, R., and Fortin, M.-J. (2012). Spatio-temporal variations in invertebrate–cod–environment relationships on the Newfoundland–Labrador Shelf, 1995–2009. *Mar. Ecol. Progress Ser.* 469, 263–278. doi: 10.3354/meps10026

Xue, H., Incze, L., Xu, D., Wolff, N., and Pettigrew, N. (2008). Connectivity of lobster populations in the coastal Gulf of Maine: part I: circulation and larval transport potential. *Ecol. Modell.* 210, 193–211. doi: 10.1016/j.ecolmodel.2007.07.024

Zuur, A. F., Leno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

Conflict of Interest: BL was employed by ECS Federal, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 31 August 2021
doi: 10.3389/fmars.2021.717469

A Biophysical Model and Network Analysis of Invertebrate Community Dispersal Reveals Regional Patterns of Seagrass Habitat Connectivity

John Cristiani[1], Emily Rubidge[2,3], Coreen Forbes[1], Ben Moore-Maley[4] and Mary I. O'Connor[1]*

[1] Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada, [2] Institute of Ocean Sciences, Fisheries and Oceans Canada, Sidney, BC, Canada, [3] Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC, Canada, [4] Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada

The dispersal of marine organisms is a critical process for the maintenance of biodiversity and ecosystem functioning across a seascape. Understanding the patterns of habitat connectivity that arise from the movement of multiple species can highlight the role of regional processes in maintaining local community structure. However, quantifying the probability and scale of dispersal for marine organisms remains a challenge. Here, we use a biophysical model to simulate dispersal, and we conduct a network analysis to predict connectivity patterns across scales for the community of invertebrates associated with seagrass habitat in British Columbia, Canada. We found many possible connections and few isolated habitat meadows, but the probability of most connections was low. Most habitat connections occurred within 3 days of dispersal time over short distances, indicating potential limits to long distance dispersal and little effect of species-specific dispersal abilities on the potential spatial extent of habitat connectivity. We then highlight the different roles that individual seagrass meadows can play in maintaining network connectivity. We also identify clusters of connected meadows and use these clusters to estimate the spatial scale of community dynamics. The connectivity patterns generated by our dispersal simulations highlight the importance of considering marine communities in their broad seascape context, with applications for the prioritization and conservation of habitat that maintains connectivity.

Keywords: community connectivity, biophysical modeling, Salish Sea, eelgrass (*Zostera marina*), Lagrangian particle tracking, network analysis, community detection, marine spatial planning

OPEN ACCESS

Edited by:
Mary C. Fabrizio, College of William & Mary, United States

Reviewed by:
Jonathan Lefcheck, Smithsonian Institution, United States
Maria Gabriela Palomo, Independent Researcher, Ciudad de Buenos Aires, Argentina

****Correspondence:***
John Cristiani
jcristia@zoology.ubc.ca

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: *31 May 2021*
Accepted: *05 August 2021*
Published: *31 August 2021*

Citation:
Cristiani J, Rubidge E, Forbes C, Moore-Maley B and O'Connor MI (2021) A Biophysical Model and Network Analysis of Invertebrate Community Dispersal Reveals Regional Patterns of Seagrass Habitat Connectivity. Front. Mar. Sci. 8:717469. doi: 10.3389/fmars.2021.717469

INTRODUCTION

A key challenge in marine ecology and conservation science is to identify the spatial scale of biodiversity patterns and the relative role of the complex oceanographic processes that may influence these patterns. This requires moving beyond the study of biodiversity in single habitat patches to instead considering the seascape as a mosaic of habitat patches connected by dispersal (Boström et al., 2011; Pittman et al., 2011). Dispersal, a foundational process in metacommunity

theory, has been shown to be a key driver determining diversity patterns at local and regional scales (Kneitel and Miller, 2003; Loreau et al., 2003; Mouquet and Loreau, 2003; Massol et al., 2017; Thompson et al., 2020). In coastal systems, the diversity of a region or any habitat within the region can depend on the spatial arrangement of habitats and the variation among organisms' abilities to move between them (Cowen and Sponaugle, 2009; Boström et al., 2010). Therefore, understanding the ecological consequences of movement requires spatially explicit knowledge of functional connectivity – how dispersal behavior and habitat configuration combine to influence the movement of an individual across the seascape (Kindlmann and Burel, 2008; Kool et al., 2013).

Spatially explicit movement information facilitates the analysis of dispersal patterns as a network in the context of meta-population/community theory (Hanski, 2001; Leibold and Chase, 2018), which can reveal emergent spatial properties of the seascape and focal communities that would otherwise not be evident without a network perspective (Urban et al., 2009). A disconnected network would indicate isolated communities that do not interact, whereas a highly connected network of habitat patches may operate as a single regional community. Furthermore, identifying patches of habitat that are central, in terms of how they link populations in other patches through dispersal and colonization, indicates areas of habitat that may provide stepping stones that are important for maintaining regional connectivity and thus maintaining biodiversity patterns (Saura et al., 2014; Albert et al., 2017). A network analysis can also identify clusters of patches based on higher levels of ecological exchange within than outside the cluster (Thomas et al., 2014; Gilarranz et al., 2017). This can identify dispersal barriers and subsequently reveal the spatial scale of metapopulations or provide a first prediction of the scale of metacommunities in the absence of local biological and environmental influences on persistence. These and other seascape-scale patterns that have consequences for biodiversity only emerge from a network approach that can consider movement rates among many patches simultaneously.

Marine connectivity research has primarily focused on the connectivity of coral reefs or pelagic species, but the connectivity of other patchy nearshore habitat types and the species that disperse among them remains poorly understood (Bryan-Brown et al., 2017). Seagrass, like coral, provides foundational biogenic habitat for a high level of biodiversity (Orth et al., 1984), including communities of epifaunal invertebrates (Heck and Thoman, 1984; Duffy et al., 2015). The patchy distribution of seagrass meadows across the seascape may create the structure for a metacommunity of seagrass-associated species that are connected by animal movement (Bell, 2006; Boström et al., 2006; Whippo et al., 2018). There are a variety of life histories present in this community which results in a range of dispersal abilities. These species may move as larvae, juveniles or adults through mostly passive transportation in ocean currents and settle on distant meadows of seagrass (Boström et al., 2010). Dispersal and connectivity have been suggested as one of the important drivers of local and regional biodiversity patterns in eelgrass-associated communities (France and Duffy, 2006; Yamada et al., 2014; Whippo et al., 2018; Stark et al., 2020). These inferences were based on spatial biodiversity patterns and in the absence of oceanographically informed estimates of potential connectivity. Consequently, spatially explicit dispersal information is essential to gain further insight into how these communities are structured across spatially heterogeneous seascapes.

In the nearshore seascape, connectivity patterns are determined by physical oceanographic processes (e.g., tidal, wind, and freshwater forcing), dispersal ecology (dispersal and post settlement survival), and the physical arrangement of habitat (Werner et al., 2007). A biophysical model that incorporates hydrodynamic models and biological properties can be an effective tool for predicting movement pathways and quantifying connectivity between habitats in the form of probabilities linking a matrix of habitat (Siegel et al., 2003; Treml et al., 2012; Sunday et al., 2014; Schill et al., 2015; Wren et al., 2016). However, modeling nearshore dynamics is difficult compared to pelagic studies. While we know the broad-scale movement of ocean currents adjacent to the coast of a continent (e.g., California current, Alaska current), predicting connectivity for spatially complex coastal areas and for species with low dispersal abilities requires high resolution hydrodynamic models and spatial habitat data (Werner et al., 2007).

A biophysical model for nearshore habitat would allow for the quantification of connectivity at the smaller spatial and temporal scales relevant to ecological processes. While genetic studies can support inferences of connectivity from parentage analysis, logistical constraints limit the spatial scope of these studies (D'Aloia et al., 2015; Bode et al., 2019), they may only be applicable at very large scales (Riginos et al., 2019), and regional genetic structure may not reflect the levels of connectivity that influence the population dynamics that maintain biodiversity at the community level. For example, in areas influenced by past glaciation, genetic structure may still be detectable from historical gene flow patterns and may not reflect current pathways of exchange (Hedgecock et al., 2003; Sunday et al., 2014; Selkoe et al., 2016). However, biophysical modeling can still predict what may be considered weak connections but genetically significant dispersal events (e.g., one migrant per generation), as even minimal levels of gene flow can homogenize populations (Waples, 1998; Jenkins and Stevens, 2018). Therefore, "ecological connectivity" is the level of exchange of individuals that can influence population and community dynamics at non-evolutionary time scales (Treml et al., 2012). In addition, an ecological connectivity analysis based on a biophysical modeling approach can supply the information most relevant to conservation and the design of a network of connected protected areas, such as population rescue effects, source sink dynamics, and trophic dynamics (Burgess et al., 2014; Guzman et al., 2019).

The objective of this study is to quantify potential connectivity among communities inhabiting seagrass (*Zostera marina*) habitat on the British Columbia (BC) coast of Canada in the form of the functional connectivity that is generated by the varying dispersal abilities of different species of the seagrass associated epifaunal invertebrate community. We expect that asymmetric

ocean currents, dispersal ability, and the spatial arrangement of habitat create patterns of connectivity that vary spatially across the region. We use a biophysical modeling and network analysis approach to answer the following questions: (1) To what degree is an invertebrate community, consisting of multiple life-history strategies, likely connected by dispersal among seagrass meadows? (2) Do increases in dispersal potential result in higher habitat connectivity? (3) Which seagrass meadows are important for maintaining network connectivity, and does habitat location and/or size determine this importance? (4) Do coastal topography and water currents create distinct clusters of meadows likely to be ecologically connected? Together, these questions allow us to assess patterns of connectivity across spatial scales and for multiple taxa.

MATERIALS AND METHODS

We used a biophysical modeling approach to simulate dispersal and estimate potential connectivity of the community of seagrass associated invertebrate fauna (i.e., "community connectivity"; **Table 1**). The model consists of five components: (1) seascape physical structure information, (2) invertebrate community trait data that influences dispersal ability, (3) hydrodynamic model results that contain ocean current velocities, (4) an individual-based particle tracking model (IBM) to estimate dispersal trajectories with the influence of dispersal traits, (5) and a network and cluster detection analysis to interpret the connectivity of the dispersal trajectories in an ecological context (**Figure 1**).

Study System

We focused on the Salish Sea region of the BC and Washington coast. The Salish Sea is a semi-enclosed system bounded by Vancouver Island, and connected to the Pacific Ocean through the Juan de Fuca Strait in the south and narrow channels to the north (**Figure 2**). The topographic complexity and glacial history of BC's coastline, that likely influence species distributions (Pielou, 1991), creates unique challenges for modeling and as a result nearshore connectivity is poorly understood. In this region, there are also two important climatic changes that drive strong seasonal differences in hydrodynamics and have ecological relevance: (1) the spring transition between Aleutian Low and North Pacific High pressure dominance over the northeast Pacific that suppresses winter storm activity resulting in phytoplankton blooms (Kathleen Collins et al., 2009; Bakri and Jackson, 2019), and (2) the summer freshet, dominated by the undammed Fraser River that increases the stratification and reduces the residence time of the surface layer (Pawlowicz et al., 2007).

The dominant habitat-forming seagrass is eelgrass (*Z. marina*) which is patchily distributed along the entire coast of BC in sheltered intertidal and subtidal areas. Eelgrass occurs to a maximum depth of 10 m depending on turbidity (Christiaen et al., 2015) and can form meadows that range in size from a few seasonally intermittent shoots to more permanent meadows greater than 30 km^2 (Murphy G. E. P. et al., 2021). As a primary producer and coastal habitat forming species, eelgrass provides habitat and a productive algal food source for multiple trophic levels (Heck et al., 2008; Amundrud et al., 2015; Duffy et al., 2015; Huang et al., 2015). Eelgrass has also been identified as an ecological conservation priority for current marine conservation planning efforts in BC (Gale et al., 2019; Rubidge et al., 2020).

Eelgrass provides biogenic habitat for a variety of epifaunal and infaunal invertebrates. Common taxa include Amphipoda, Isopoda, Decapoda, Polychaeta, and Bivalvia, and may include the full life cycle of a species or just the juvenile or adult stages. While not all species within these groups are eelgrass habitat specialists, together they constitute a unique assemblage distinct from communities in the surrounding substrates, and we are focusing on their probabilities of connecting spatially distinct habitat as a possible route of dispersal. In general, most species in this community are direct developers or planktotrophic/lecithotrophic, and they are semi-mobile or sessile as adults (Boström et al., 2010). In addition to larval drifting, pelagic dispersal may also occur by rafting on seagrass and epiphytic algae which has been shown as a viable mode of transport for small invertebrates and can enable kilometer-scale dispersal for sessile species (Worcester, 1994; Brooks and Bell, 2001). Reproductive eelgrass shoots can remain buoyant for up to 26 days (Harwell and Orth, 2002; Källström et al., 2008). While species with a multi-day pelagic larval phase may have the greatest dispersal distance potential, species that are direct developers may still make short movements through rafting that can be more influential to population dynamics than longer distance low probability dispersal (Johannesson, 1988). Therefore, we included all sampled species in the community regardless of development type and mobility.

Seascape Structure Spatial Data

The first component of the biophysical model is the structural data of the seascape, consisting of coastline and seagrass spatial data. A coastline vector dataset was derived from a 1:20,000 scale provincial government dataset which provided the sufficient detail to represent nearshore features. Small islands (<0.01 km^2) were removed to reduce the complexity of the dataset. Eelgrass spatial data was obtained from multiple government and non-governmental sources, which used a variety of survey methods. While the dataset achieves near coastwide coverage, only presences were consistently documented and there are likely areas of incomplete sampling effort which means there are likely meadows that exist that are not included in our analysis. However, much of the data have been ground truthed, and we are confident that most major meadows are included. The seagrass dataset was simplified to more closely match the resolution of the oceanographic model (0.5 km). Primarily, this involved aggregating seagrass polygons that were within 100 m of each other.

Parameterize Model With Dispersal Trait Values

To simulate the dispersal of the seagrass invertebrate community we compiled dispersal related traits from a literature search for 63

TABLE 1 | Glossary of key terms.

General connectivity	
Community connectivity	The movement of multiple species (that co-occur and interact) among a spatially distinct habitat type (Carr et al., 2017).
Ecologically relevant connectivity	The level of exchange of individuals that can influence population and community dynamics at non-evolutionary time scales (Treml et al., 2012).
Dispersal	A once in a lifetime movement to a new habitat patch, which may occur in a gamete or larval phase, or by chance as a juvenile or adult on marine debris, and is distinct from migration and foraging movements (Guzman et al., 2019).
Biophysical model	
Pelagic propagule duration (PD)	The maximum time that a species can disperse by drifting in the pelagic environment. This term is generalized to include dispersal at all life stages of an organism and is inclusive of planktonic larval duration (PLD) and rafting on seagrass and algal debris (Shanks, 2009). We simulated PD times of 1, 3, 7, 21, and 60 days. For species that do not have a planktonic larval phase, we considered them capable of rafting on seagrass debris, which can remain buoyant for up to 3 weeks.
Mortality rate	The daily probability of mortality during dispersal. This was set at 15% per day.
Settlement behavior	If an organism drifts over another seagrass meadow, it will stop drifting and settle on that meadow.
Stranding behavior	If an organism drifts into the coastline during dispersal, it will remain at that position for the remainder of the simulation.
Time period	The experimental range when we ran a dispersal simulation with date-specific hydrodynamic data. The selected times are intended to capture variation in connectivity by year, season, and tidal cycle. We ran simulations over 3 years (2011, 2014, and 2017), and for 3 seasons within each year (winter: Jan-Mar, spring: May-Jul, and summer: Aug-Oct). In each simulation, we released particles every 4 h for 2 weeks.
Network analysis	
Node	A point in a graph that represents a habitat patch (i.e., seagrass meadow).
Connection probability	A link between two nodes in a graph that results from the dispersal between seagrass meadows. The probability of a connection is directional and is calculated as the number of particles that settle on a distant meadow divided by the total amount of particles released from the origin meadow (×100).
Community averaged connectivity	A connectivity network that is averaged across all PD levels and time periods to represent multi-species connectivity (Melià et al., 2016; D'Aloia et al., 2017). For each connection, we averaged the connectivity established by each PD level within a time period, and then averaged across all time periods.
Probability of Connectivity (PC)	A graph-wide metric that quantifies the total amount of habitat connected by dispersal. It combines the habitat area available for movement within a patch (intraconnectivity) with the area made available by connections between patches (interconnectivity). The area of habitat connected by inter-patch movement is weighted by connection probability (Saura and Rubio, 2010).
*d*PC (intra, flux, and connect)	The contribution of each node to the overall PC metric. It is calculated by removing a node and calculating the percent change in PC. It is comprised of three fractions: intra, flux, and connect. *Intra* represents the intra-connectivity of a patch (i.e., the area available for within patch movement). *Flux* quantifies all the area-weighted connections in and out of that patch. *Connector* measures how much a patch is included in the paths between other patches and therefore acting as a stepping stone in the system (Saura and Rubio, 2010).
Cluster detection	
Cluster	Groups of nodes that are strongly connected to each other and weakly connected to other nodes
Weighted connectivity length scale	The average connectivity length scaled by connection probability within a cluster, used to compare the variation in connectivity among clusters (Thomas et al., 2014). Calculated as: $$\frac{\sum_{all\ connections} connection\ prob. \times length}{\sum_{all\ connections} connection\ prob.}$$

species that were identified in biodiversity surveys of meadows along the coast of BC (Whippo et al., 2018; Stark et al., 2020). We considered the potential pelagic propagule duration (PD) and a daily mortality rate in the biophysical model. PD is inclusive of larval drift, adult movement, and rafting (Shanks, 2009). In addition, all species were assumed to have a settlement behavior trait, which simply means that they will settle if they drift over suitable seagrass habitat (see **Figure 3** for a diagram of processes that influence connectivity). We binned PD values into five levels, and to achieve equal width bins we used values of e^n days, where $n = 0, 1, 2, 3, 4$, which we rounded to 1, 3, 7, 21, and 60 days on a linear scale (**Supplementary Figure 1**). Although some species have a PD longer than 60 days, early testing showed that the coastal boundary constraints of the Salish Sea prevent most larvae from drifting longer than 60 days before stranding. For species that do not have a planktonic larval phase, we considered them capable of rafting on seagrass debris, which can remain buoyant for up to 3 weeks (Harwell and Orth, 2002; Källström et al., 2008).

We applied a single instantaneous mortality rate for all species due to a lack of information for individual species. Frequently used rates of invertebrate larvae mortality range between 0.15 and 0.23 day^{-1} depending on methodology (Rumrill, 1990). White et al. (2014) revisited the Rumrill (1990) data and estimated mortality rates of <0.15 day^{-1} using a different methodology. Therefore, we used 0.15 day^{-1} to ensure that an adequate mortality rate was still represented for all species but that it was not set unrealistically high.

We did not include active or directed swimming behavior in the biophysical model. There is limited information on swimming speed of invertebrate larvae for many of these species. In addition, the sustained swimming speeds of small larvae are usually much less than current speeds (Orth, 1992; Daigle et al.,

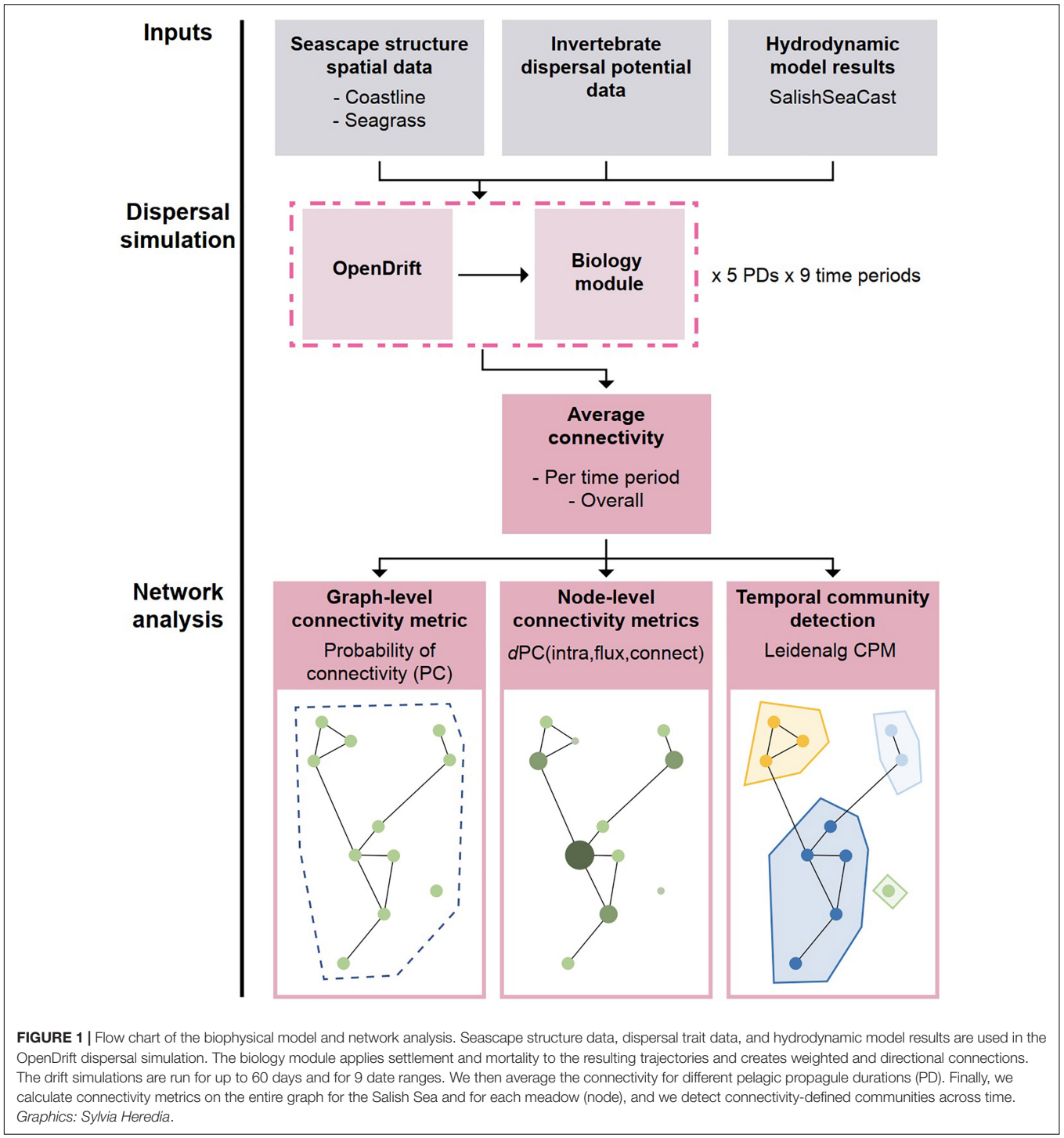

FIGURE 1 | Flow chart of the biophysical model and network analysis. Seascape structure data, dispersal trait data, and hydrodynamic model results are used in the OpenDrift dispersal simulation. The biology module applies settlement and mortality to the resulting trajectories and creates weighted and directional connections. The drift simulations are run for up to 60 days and for 9 date ranges. We then average the connectivity for different pelagic propagule durations (PD). Finally, we calculate connectivity metrics on the entire graph for the Salish Sea and for each meadow (node), and we detect connectivity-defined communities across time. *Graphics: Sylvia Heredia*.

2016). Therefore, we make the assumption that modeling passive dispersal as influenced by advection and diffusion is adequate when considering large-scale movements.

Hydrodynamic Model

The hydrodynamic fields used to force the dispersal simulations were obtained from the SalishSeaCast configuration of the Nucleus for European Modelling of the Ocean, a finite-difference, hydrostatic, community ocean model (Gurvan et al., 2017). SalishSeaCast is described in detail by Soontiens et al. (2016), Soontiens and Allen (2017), and Olson et al. (2020). Briefly, the configuration uses approximately 0.5 km horizontal resolution and 40 z-coordinate layers ranging in thickness from 1 m near the surface to 27 m at depth. Hourly surface wind and meteorological forcing fields are sourced from the 2.5 km High Resolution Deterministic Prediction System maintained by

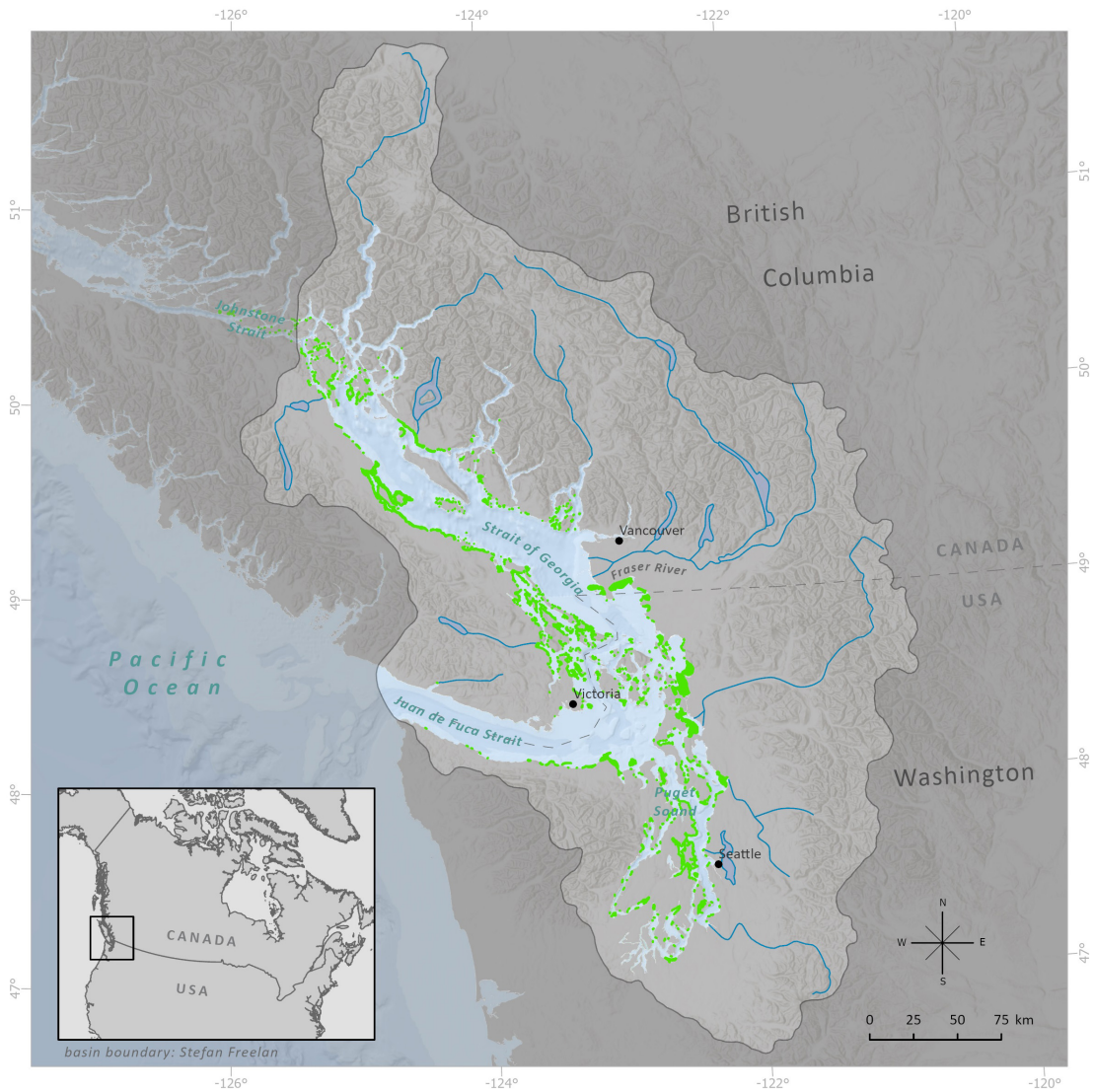

FIGURE 2 | The Salish Sea and seagrass meadows. The basin boundary and major rivers are highlighted to show the significant freshwater inputs in this estuarine system. The SalishSeaCast oceanographic model extends from Puget Sound in the south to Johnstone Strait in the north, with open boundaries to the Pacific Ocean at Johnstone Strait and Juan de Fuca Strait. The model has a near-uniform grid spacing of 440 m × 500 m and a 1-h temporal resolution. Seagrass meadows range in size from <0.001 to 50 km^2, with a total area of 519 km^2. Seagrass polygon size is exaggerated in this map for visualization.

Environment and Climate Change Canada (ECCC; Milbrandt et al., 2016). Runoff at 150 rivers is prescribed using monthly watershed climatologies (Morrison et al., 2012) along with daily observations from the ECCC Fraser River flow gage at Hope, BC. Oceanic forcing of temperature, salinity, and eight tidal constituents is implemented at open boundaries in Juan de Fuca Strait and Johnstone Strait.

SalishSeaCast is optimized for the Strait of Georgia and reproduces extensive observations of water level (Soontiens et al., 2016) and temperature and salinity (Olson et al., 2020) in that portion of the domain with competitive skill relative to similar models of the region (e.g., Khangaonkar et al., 2018). This skill was achieved through careful tuning of tides, bathymetry, and sub-grid scale physics to accurately resolve several important features of the circulation, including mixing over sills and annual flushing of the deep Strait of Georgia (Soontiens and Allen, 2017). While the SalishSeaCast velocity fields have not been directly evaluated against observations, the lack of significant temperature and salinity bias in the presence of strong spatial gradients suggests that near-surface currents are statistically accurate. Aside from model tuning, this accuracy is primarily owed to the high-resolution wind forcing as wind is a dominant driver of surface currents along with rivers and tides (Halverson and Pawlowicz, 2016).

We used hourly current velocities from the SalishSeaCast model for 3 years (2011, 2014, and 2017). Within each year we considered three distinct time periods that may have ecological significance: January–March (winter), May–July (spring), and

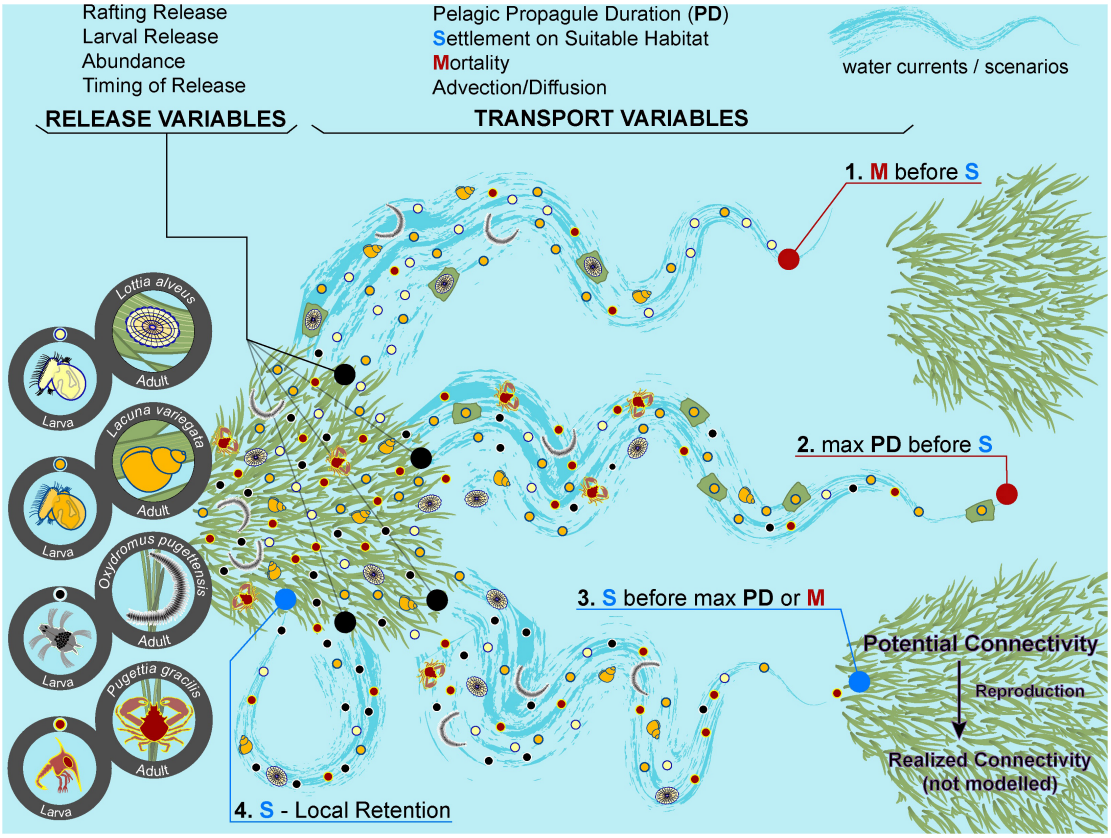

FIGURE 3 | Examples of dispersal scenarios and the processes that contribute to connectivity. Each numbered scenario represents a potential fate of a particle. Four representative species are shown, but a scenario is not specific to any one species. In a simulation, particles are released from a seagrass meadow. RELEASE VARIABLES: The *timing of release* varies by year, season, and hour within a tidal cycle. The *abundance* released is proportional to meadow size. Particles can drift as pelagic *larvae* or by *rafting* on seagrass as a juvenile or adult (if rafting, pelagic duration = 21 days). TRANSPORT VARIABLES: During transport, particles are *advected* and *diffused* by the hydrodynamic model and they experience a 15% daily *mortality* rate (randomly applied), thus reducing density and abundance through time. When a particle drifts over another seagrass meadow, it *settles* and is removed from the simulation. If a particle encounters the coastline, it strands and is unable to drift further. A particle can drift for as long as its *pelagic propagule duration* (PD). We do not model any swimming behavior, but for our species it is negligible compared to the advection speeds. SCENARIOS: In scenario 1, the particles experience mortality before reaching suitable habitat. In scenario 2, a particle reaches its maximum PD before drifting over any suitable habitat. In scenario 3, a particle reaches suitable habitat before experiencing mortality or reaching its max PD. Reaching this stage is considered "potential connectivity." Full "realized connectivity" requires the individual to reproduce to establish a genetic connection. In our simulations, we only model up to potential connectivity. In scenario 4, a particle is advected back to its meadow of origin. *Graphics: Sylvia Heredia*.

August–October (summer). Pawlowicz et al. (2019) noted distinct differences in circulation between winter and summer as a result of differences in freshwater inputs and upwelling over the outer shelf. In addition, peak seagrass reproduction occurs during August, followed by senescence in the fall when shoots are most likely to break (Källström et al., 2008), which may be a time when more rafting occurs. Only the surface layer of the model was used for our dispersal simulations. We justify this assumption because seagrass is a shallow subtidal habitat and we are interested in successful connections between meadows and not the fate of particles that sink.

Despite the high resolution of the model, there were still areas of the Salish Sea too narrow to be resolved (e.g., inlets, passages). The hydrodynamic model criteria requires a modeled area to be at least two grid points wide, and narrow areas were either widened or not considered (Soontiens et al., 2016). Therefore, we removed seagrass meadows that overlap with any narrow area not considered in the hydrodynamic model. This resulted in the removal of only 24 out of 994 meadows.

Dispersal Simulation

The dispersal of eelgrass-associated invertebrates was simulated using the Python-based framework OpenDrift – an IBM for Lagrangian particle tracking (Dagestad et al., 2018). In addition to the Forward–Euler numerical integration scheme provided by OpenDrift, we wrote a custom module to incorporate PD, mortality, and settlement. The basic description of a simulation is as follows: for each of the nine experimental time periods (3 seasons × 3 years), particles are released simultaneously from all seagrass meadows every 4 h for the first 2 weeks to account for tidal variation, they are tracked as they are advected and diffused across the seascape, daily mortality is applied by randomly selecting particles and removing them from the simulation, if a particle drifts over another seagrass meadow

or returns to the same meadow it is considered settled and removed from the simulation, the simulation is run until the end of the PD or until all particles have settled or stranded on the coast (**Figure 3**). These simulations model *potential connectivity* (transport and settlement only), whereas *realized connectivity* requires the individual to reproduce and establish a genetic connection.

The number of particles released per meadow scales with meadow area and release locations are spaced evenly within a meadow. In total, 3.8 million particles were released per period, which was sufficient to capture the variation of particle destinations while scaling within the computing resources available. The position of particles was updated every 30 s in the simulation. A 1.5 m^2/s diffusion rate (K) was used to represent the effect of subgrid-scale turbulent motions on particle displacement. We implemented this diffusion using a statistical relationship between K and the particle velocity variance V^2

$$\overline{V^2} = \frac{2K}{dt}$$

where dt is the time step (LaCasce, 2008). A random walk was then applied to the particle displacements using a Gaussian distribution defined by the velocity variance.

Network Analysis

We conducted a network analysis to answer our first three questions on quantifying and characterizing potential connectivity. Network methods analyze connections resulting from the dispersal simulation in graphical form to study their topological relationships and uncover spatial patterns of connectivity. With a graphical approach, seagrass meadows are nodes and dispersal connections are edges in a graph, which are directional and weighted by connection probability (Minor and Urban, 2007). Connection probabilities were calculated as the percentage of particles released from the origin meadow that settle on a destination meadow.

To answer our first question on the connection probabilities of whole seagrass invertebrate communities, we averaged connectivity across PD scenarios and across time to move from population-level to community-level estimates of connectivity. This approach is useful for characterizing the functional role of habitat to multi-species patterns of movement (Melià et al., 2016; D'Aloia et al., 2017). In the averaging scheme, we weighted connections by how common they were across all PD scenarios and time periods. For example, if a connection between two meadows was made in just the 60-day PD scenario in only 1 of the 9 time periods, then it would be considered less important to overall community connectivity than a connection made at multiple PD levels and in every time period.

To answer our second question on the relationship between dispersal ability and overall habitat connectivity, we calculated the Probability of Connectivity (PC) metrics from the Conefor software package for each PD level (Saura and Torné, 2009; Saura and Rubio, 2010). PC incorporates dispersal probabilities and weights them by an additional patch attribute, typically area, to calculate a measurement of "habitat availability," indicating how well connected (i.e., available for movement) the entire system is. By incorporating patch area, we start with the assumption that a patch itself provides area for movement, which may be important for seagrass-associated invertebrates with limited dispersal abilities. Then, any connections made between patches add to the area available for an organism to move between. For instance, a connection probability of 10% between two large patches connects more habitat than the same strength connection between two smaller patches. Thus, the *intra*connectivity of a network provides a baseline measurement of connectivity to compare to the additional area made accessible by *inter*connectivity. This allows us to move beyond simply knowing a quantity of nodes connected which may not be as informative for understanding the importance of a patch to the overall network. An additional benefit of considering intraconnectivity is to avoid characterizing isolated meadows as having no functional role in supporting animal movement, and therefore in supporting community diversity.

By weighting connections by area as the patch attribute in the PC calculation, we are using area as a proxy for intrapatch movement. However, we also intend area to be a general proxy for other patch importance metrics that may scale with area, but non-linearly, such as habitat quality, local retention, and species diversity (Minor and Urban, 2007; Saura and Rubio, 2010; Pereira et al., 2011; Engelhard et al., 2017). Since our patch areas spanned 7 orders of magnitude with a right-skewed distribution, we log-transformed areas to achieve a normal distribution of patch areas, so as not to overweight the importance of large patches or deem small patches as completely insignificant to the multiple functional roles that they may play in influencing connectivity patterns.

To answer our third question on characterizing the contribution of individual seagrass meadows to the overall connectivity, we calculated the change in PC (dPC) when that meadow is removed, indicating the importance of that node to contributing to and maintaining connectivity. We calculated dPC for each dispersal scenario and averaged the results. dPC is comprised of three component parts: intra, flux, and connector. These components represent the different ways a node can contribute to connectivity. They are non-overlapping properties of the network and provide a more comprehensive assessment of connectivity than just considering traditional connectivity metrics separately (e.g., betweenness centrality, node degree). Intra represents the intra-connectivity of a patch (i.e., the area available for within patch movement and local retention). Flux represents how much a patch is connected to other patches by considering all the area-weighted connections in and out of that patch. Connector measures how much a patch is included in the multi-step paths between other patches and therefore acting as a stepping-stone to link the system.

Together, these metrics show the different ways that a seagrass patch can contribute to the overall connectivity of the network (Saura and Rubio, 2010). Given that we are interested in dispersal as a fundamental ecological process, it was important to use ecologically relevant metrics that have both a structural and functional basis and relate pattern and process (Pittman, 2018). The PC metric and its component parts allow us to interpret the

functional role of seagrass habitat connectivity patterns in the context of the invertebrate dispersal process.

Cluster Detection

To answer our fourth question on identifying distinct clusters of connected seagrass habitat that may arise from topography and ocean currents, we used "community detection" methods. Community detection algorithms identify clusters of nodes that are strongly connected to each other and weakly connected to other nodes in the network (a "community" refers to a graphical property and not an ecological community of species, to avoid confusion with an ecological community, we refer to a graph theoretical "community" as a cluster). This allows us to quantify the spatial scale of dispersal between interacting clusters and potentially identifies ecologically distinct regions.

We used the CPM function in the Leidenalg Python package to identify meadow clusters of varying clustering strength (Traag et al., 2019). CPM gives the user control over a resolution parameter that sets a threshold of connectivity for community membership. Maintaining control over the resolution parameter allows for different ecological interpretations of the network clustering, as opposed to just identifying the one mathematically optimal partitioning which may not be ecologically interpretable. For instance, setting a low threshold value will select for large clusters which will identify where the strongest barriers to dispersal are in the system as only very rare connections would connect clusters. Alternatively, setting a high threshold value will select fewer nodes per cluster and identify the strongest connected clusters of nodes, but the boundaries of a cluster are more permeable (Thomas et al., 2014).

We used a temporal cluster detection method to identify meadows potentially clustered across time periods (Mucha et al., 2010; Traag et al., 2019). Using this multidimensional method, nodes could take on membership in multiple clusters which allowed us to identify how variable seagrass meadow clusters are through time. To implement this method, the user provides "interslice" weightings to indicate how similar the overall connectivity results between time periods should be considered. Knowing that hydrodynamics vary seasonally in the Salish Sea with less interannual variation, and with evidence that community composition and abundance for meadows can vary seasonally (Lefcheck et al., 2016a; Whippo et al., 2018), we chose to focus on seasonal variation. Therefore, we weighted our interslices so that between-season membership could vary more compared to year-to-year variation. This allows the seasonal dynamics to be more prevalent.

We followed similar methodology to Thomas et al. (2014) and calculated a range of temporal cluster configurations by varying the connectivity probability threshold. To identify potentially unique configurations from this range, we plotted the amount of connectivity occurring between clusters against the connectivity threshold. At threshold values where the connectivity between clusters plateaus or scales inconsistently, this indicates a stable configuration where a barrier allows the connectivity within the cluster to increase but not the connectivity between clusters (**Supplementary Figure 2**). For the configurations at the plateaus, we then calculated the weighted connectivity length scale for each cluster (see **Table 1**) and compared these values between clusters. This comparison assesses if connectivity probability scales with distance consistently across the region. Configurations with highly varying weighted connection lengths among detected clusters indicate unique dispersal patterns that may be the result of spatially distinct hydrodynamic/topographic features that are only evident at that resolution.

Ultimately, our approach analyzes connectivity at three graphical levels: a graph-wide level (PC metric), a node level (*d*PC metric), and a regional cluster level (temporal community detection) (**Figure 1**). The multi-level approach allows us to assess multi-species dispersal as it relates to the Salish Sea, individual seagrass meadows, and to sub-regional dynamics (i.e., sub-sections of the Salish Sea).

RESULTS

Community-Level Connectivity

An overall average of community connectivity probabilities is presented when averaged by PD and through time (**Figure 4**), which highlights the relative importance of a connection to all species in the community. The biophysical model predicted many possible connections and few isolated meadows, but the probability of most connections was low. Connection probabilities ranged from 0.0001 to 84% (median: 0.03%, mean: 3.9%), and connection probability got weaker as distance between meadows increased (**Figure 5**). Dispersal was not limited to immediately adjacent meadows. While the strongest connections were made among meadows nearby on the same section of coastline, there was significant cross-basin movement (**Figure 4**). Only 35 of 970 meadows were completely isolated throughout all simulations and these were primarily located in sheltered channels or bays. Most of these isolated meadows were in Johnstone Strait toward the northern end of the model domain and therefore may not be isolated if the model boundary was extended. The meadows in the north are technically not part of the Salish Sea.

In all iterations of the simulation, >99% of particles either (1) settled on another seagrass meadow, (2) were retained by the source meadow, (3) stranded on the coastline, (4) or were selected for mortality. The remainder of active particles after 60 days were at the model boundary at the exit of Juan de Fuca Strait. This indicates that the Salish Sea operates as a mostly closed system when considering regular ecological exchange for nearshore habitat.

Dispersal Potential and Habitat Connectivity

We used the graph-wide metric, PC, to answer the question of how overall connectivity of the network changes with dispersal potential. We compared the percentages of connectivity that are attributable to *inter*connectivity, as *intra*connectivity (i.e., the total seagrass area of the network) is the same for all dispersal abilities and provides a baseline of connected area (**Figure 6**). The relatively small amount of area attributable

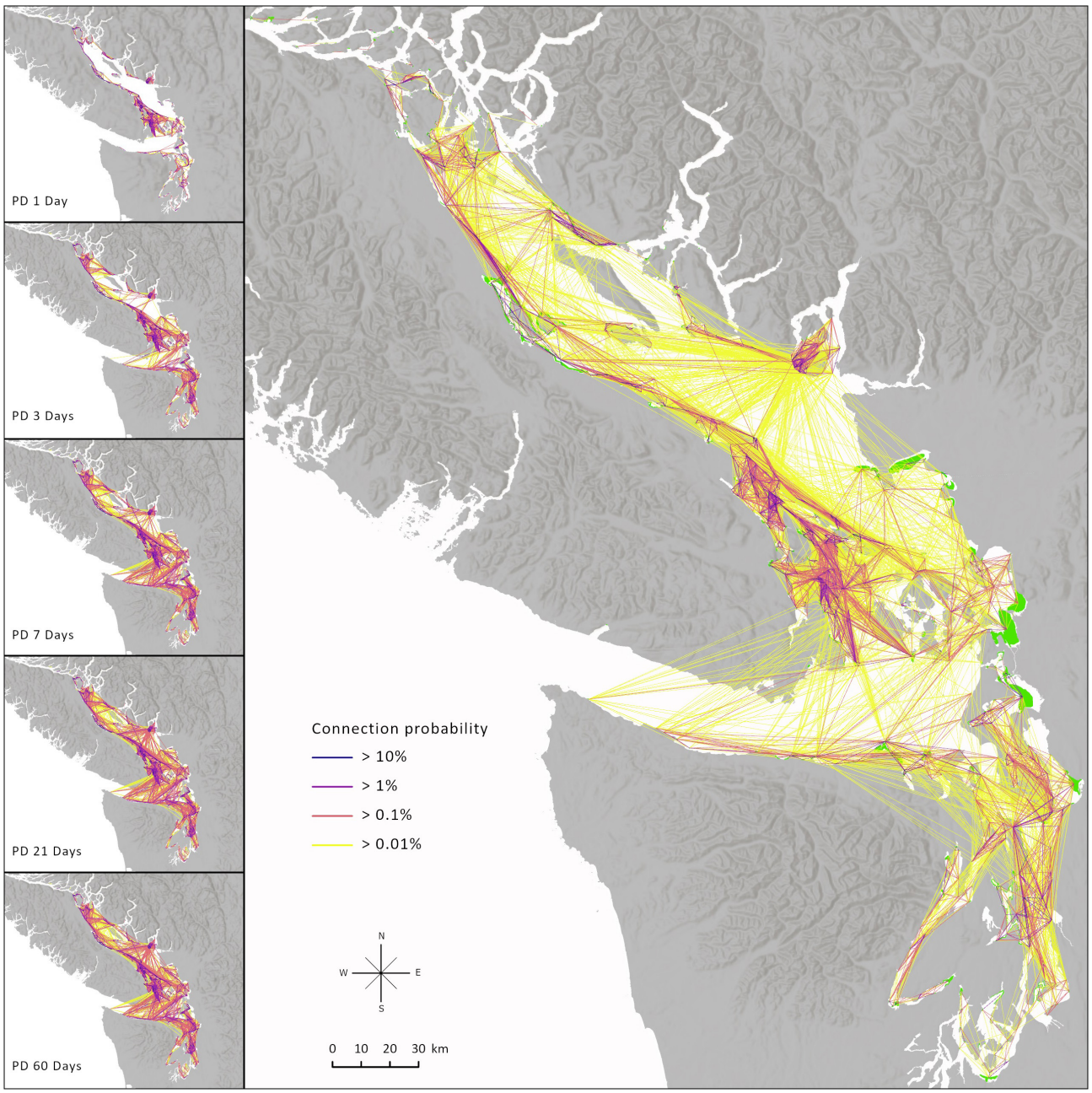

FIGURE 4 | Averaged community connectivity and individual pelagic propagule duration (PD) level connectivity. Connection probability is the percentage of particles released from the origin meadow that successfully settled on another meadow. The individual PD scenarios are for one period to show differences of connectivity for one run of the model. The overall connectivity results from averaging the PD scenarios within each period and then averaging across all time periods. Most connections will decrease in probability because not all connections were common among PDs and time periods. This prevents weighting the network toward higher dispersing species and it represents community connectivity.

to inter-meadow movement (∼4.0–4.7%) is due to the low dispersal probabilities connecting most meadows. The total area made available from inter-meadow movement increased with PD, as species that were able to drift longer were able to travel further to reach more meadows, but a limit was reached at higher values. 1-day of dispersal resulted in the largest increase in interconnectivity (4.0%), and most habitat connectivity was achieved by 3 days of dispersal time (4.6%). After 1 day, increases in PD resulted in only small increases in habitat availability because the new connections established were relatively weak (**Figure 5**). This indicates that there is a limitation to longer distance dispersal most likely caused by the constraints of the Salish Sea topography. Lastly, although differences in connectivity between seasons were minimal

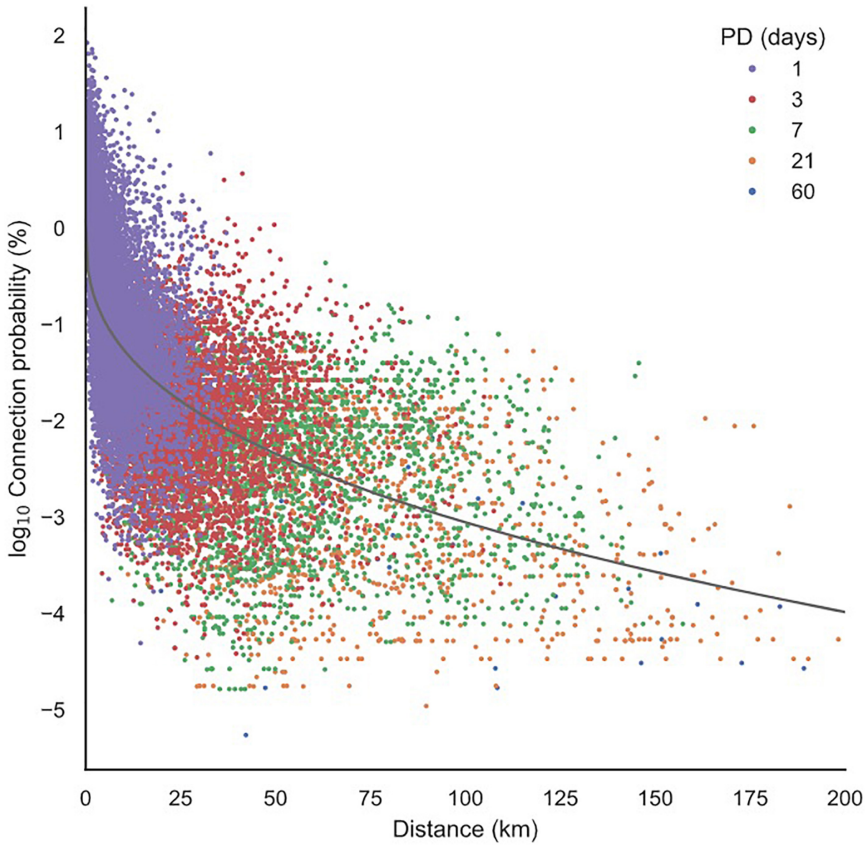

FIGURE 5 | The relationship of connection probability and distance. Probability decreased with increasing distance. Each point represents a directional connection between seagrass meadows. Connection probability is the percentage of particles released from the origin meadow that successfully settled on another meadow. An exponential curve was fitted to the data, $y = -0.52 \times 0.38$, $R^2 = 0.46$. Connections are symbolized by the pelagic propagule duration (PD) interval in which the connection was made. Generally, longer distance connections are made by species with longer PDs.

(∼0.2–0.4%), connectivity was consistently lower in the winter than in the spring/summer.

Meadow-Level Connectivity and Importance

To answer our third question about which meadows contribute most to connectivity patterns, we used the *d*PC metric and its component parts (intra, flux, and connect) to estimate each meadow's contribution as well as its functional role in the type of connectivity it maintains. In the Salish Sea, the meadows with the highest *d*PC values were primarily large meadows. This is expected when the system is connected by mostly low dispersal probabilities and most of the connectivity is represented by the intraconnectivity of meadows. However, the flux and connect values still combined to influence the overall *d*PC score, indicating that a meadow's position in a network can influence its importance (**Figure 7**).

With so many meadows in the network and with dispersal among them on relatively short (days) time scales, our analysis suggests that the overall connectivity patterns are robust to excluding any one meadow from the system. A meadow's connect value plays the smallest role in a meadow's importance due to no one meadow standing out as a sole stepping-stone connecting groups of meadows (**Figure 7D**). In addition, the meadows with the most flux are not necessarily the largest (which would be expected because they release the most particles), indicating that dispersal is restricted to some degree (**Figure 7C**). If there were no barriers to movement (e.g., asymmetric currents, land barriers), then flux would scale with area.

Meadow Clustering

Using temporal cluster detection and the relationship between intra and inter cluster connectivity (**Supplementary Figure 2**), we uncovered unique and stable configurations, of which we present two configurations that may provide ecologically relevant information (**Figure 8**). In **Figure 8A**, the connection probability required for cluster membership is extremely low ($\sim 10^{-4}$). This results in large clusters with minimal movement between them (0.4% of total connectivity), indicating where strong barriers to dispersal may exist.

In **Figure 8B**, the connection probability required for cluster membership is high ($\sim 10^{-1}$), which creates smaller clusters. The connectivity of these clusters is more certain, but the boundaries are more permeable (11.3% of total connectivity). This resolution

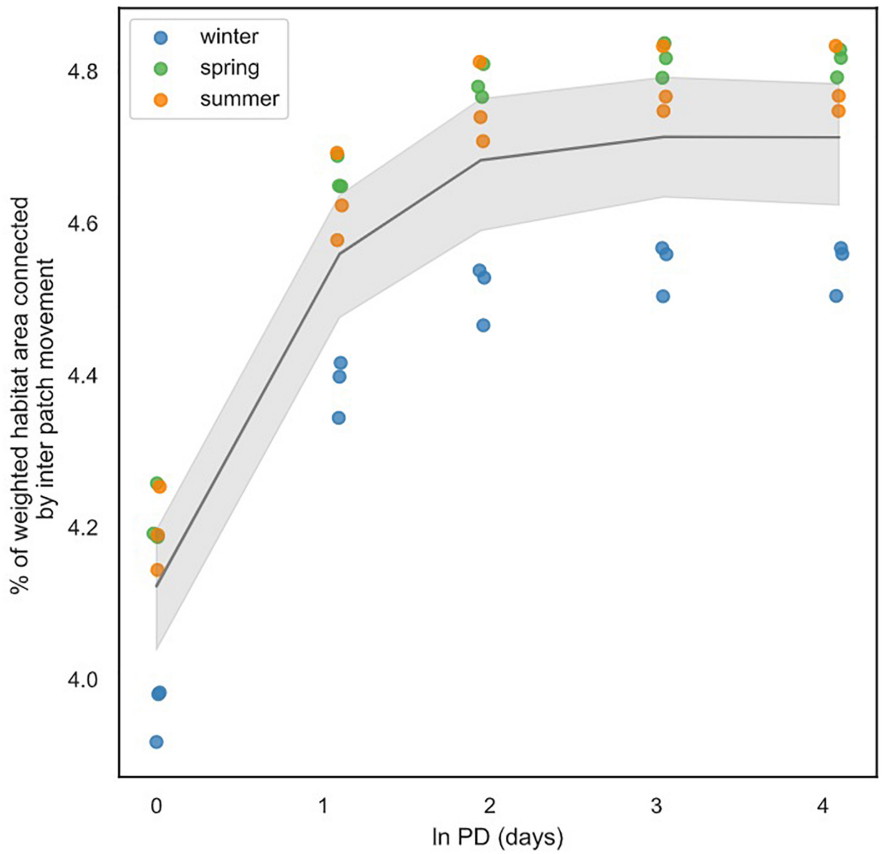

FIGURE 6 | The percent of weighted habitat area (as derived from the PC metric) connected by inter-meadow movement for each PD. For reference, 100% interconnectivity would occur if the particles from every meadow settled on another meadow. The natural log of the pelagic duration (PD) levels (i.e., 1, 3, 7, 21, and 60 days) is used to display the data in equal intervals. The black line is the mean and the gray shading indicates the 95% confidence interval.

level showed the most variation in average weighted connectivity lengths between clusters indicating unique dispersal patterns that may be the result of spatially distinct hydrodynamic/topographic features that are only evident at this resolution.

In both configurations, a node can be part of up to three different clusters because the temporal community detection analysis was set up so that dynamics could vary between three seasons. In both configurations, many nodes on the boundaries of their clusters will vary in membership. Even with high barriers to dispersal (**Figure 8A**), nodes have membership in more than 1 cluster due to the strong seasonal changes in connectivity. In **Figure 8B**, while many of the nodes are part of more than one cluster, this is mostly due to a cluster being subset into different parts and not as much from large-scale boundary overlap.

DISCUSSION

Scales of Connectivity Analysis

We posed four questions to characterize seascape patterns of connectivity from the local to regional scale. This involved analyzing connectivity at three spatial scales: the entire Salish Sea (PC metric), individual meadows (*d*PC metric), and clusters of meadows (temporal community detection). By using ecologically relevant metrics, we can interpret the functional role of an individual seagrass meadow or characterize the entire network in the context of dispersal ability, thus relating pattern to process across scales.

The individual and averaged PD results (**Figures 4–6**) quantify community-level connectivity and describe how connectivity increases with PD. The establishment of many connections after just 1 day of dispersal indicates that the relevant spatial scale for understanding seagrass community dynamics extends beyond an individual meadow scale since most species in any given meadow can likely reach other meadows. In an open system with more symmetric movement and without mortality, we would expect PC to continue to increase linearly with PD. However, we find that most connectivity is established by 3 days. Beyond this level, the topography of the Salish Sea and mortality restrict longer distance movement and most of the connectivity established is through weak connections. Pawlowicz et al. (2019) also found that the mean time to drifter stranding was 3.5 days and was not very sensitive to source location. Lastly, at each PD level, connectivity was consistently lower in the winter than in the spring/summer. Pawlowicz et al. (2019) estimated that water traveling from the Fraser River to the Pacific takes 23 days in

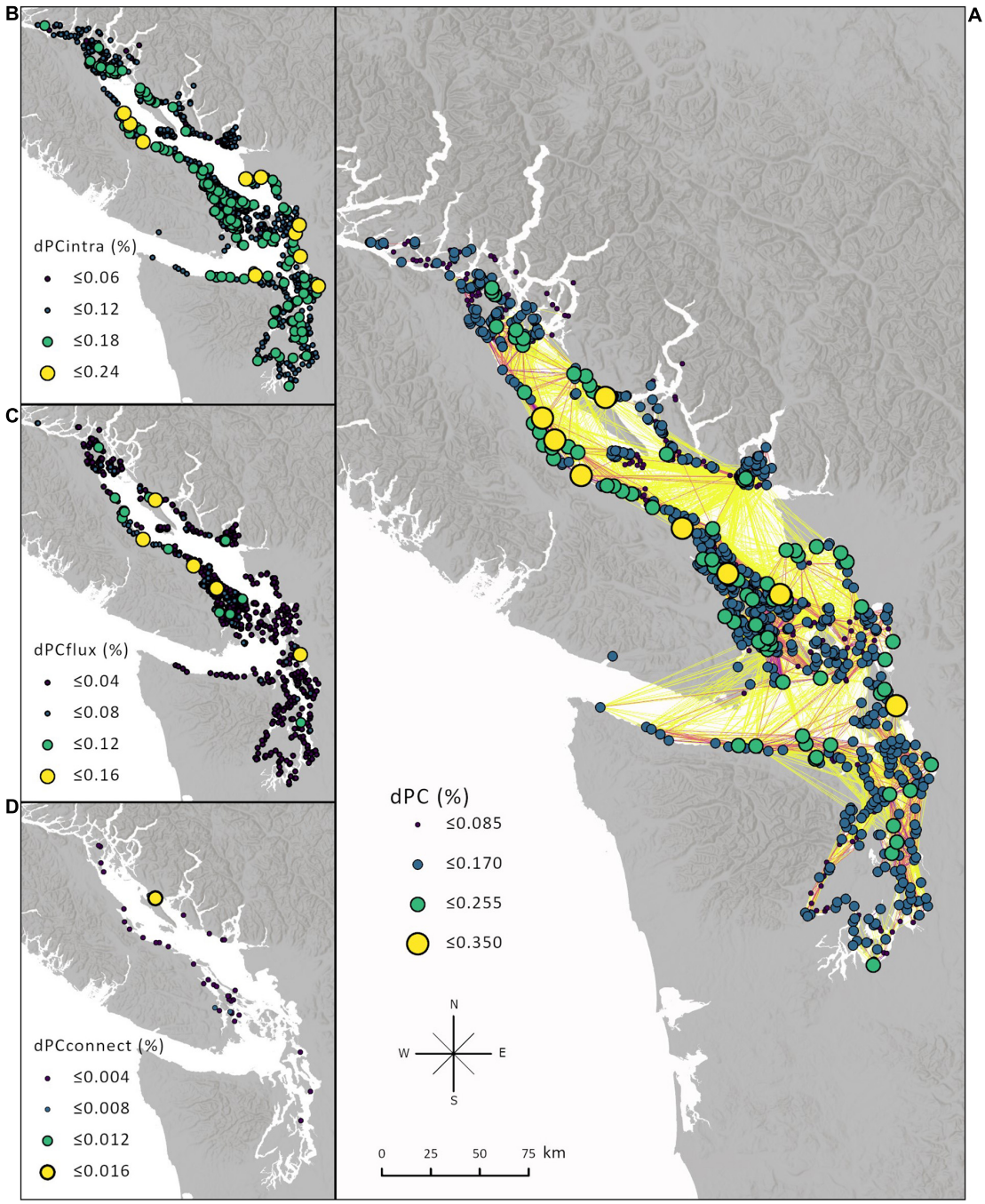

FIGURE 7 | (A) *d*PC of each node. The change in the PC metric that results from removing a node indicates the importance of that node to maintaining the connectivity of the network. *d*PC considers intra meadow movement (*d*PC*intra*) and inter meadow movement (*d*PC*flux* and *d*PC*connect*). Generally, **(B)** *intra* scales with the area of the meadow, **(C)** *flux* indicates how well the meadow is connected to other meadows, **(D)** and *connect* places more emphasis on the topological position of the meadow and its use as a stepping stone. Only meadows greater than 0.001 are shown for *flux* and *connect*. The range of values is different for each component and equal intervals are used to symbolize which nodes stand out for that component.

the summer and 53 days in winter. This is driven by changes in coastal upwelling and freshwater inflow.

The node level analysis reveals the different roles that individual seagrass meadows play in maintaining network connectivity and allows us to rank meadows by their contribution to connectivity (**Figure 7**). We can understand our results by comparing them to two predictions for species with varying dispersal abilities. (1) A species with a very low dispersal ability would primarily rely on intrapatch movement and local retention, whereas a species with a very high dispersal ability

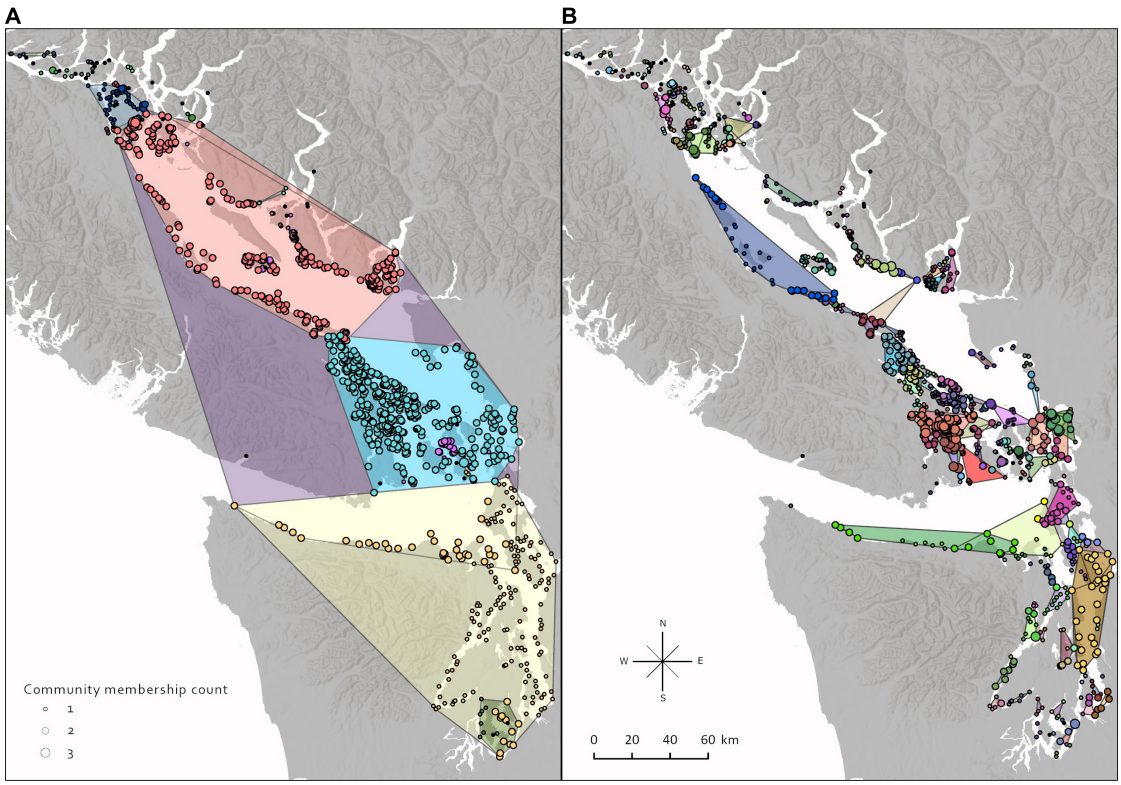

FIGURE 8 | Two seagrass network community configurations obtained from the temporal community detection analysis. A node can be a part of multiple communities through time as connectivity changes seasonally, and therefore some of the polygon boundaries overlap. Configurations were obtained by varying the threshold of connectivity probability required for community membership, and then calculating among community connectivity characteristics to select two configurations that potentially have ecological significance. **(A)** A low connectivity threshold results in large communities and indicates where barriers to dispersal occur between groups of meadows within the Salish Sea. **(B)** A higher threshold results in smaller communities but a higher certainty of where regular exchange occurs.

could make a direct connection to every patch and the network would essentially act as one large patch. Thus, in both scenarios the largest area patches would be selected as the most important for low and high dispersing species. (2) At intermediate levels of dispersal where a network is not uniformly connected, the topological position of a node (e.g., stepping stone meadows) becomes more important to maintaining connectivity (Saura and Rubio, 2010). We found that the Salish Sea is not lacking in connections, although most connections have a low probability. Therefore, large meadows are mostly selected as important, but the spatial position of a meadow is also important, as the flux and connector values combined to influence importance (**Figure 7**). This is to be expected since we are considering a community of species with a range of dispersal abilities, and stepping-stones and flux quantity may matter to overall connectivity at the low-intermediate level of dispersal ability.

Lastly, we identified clusters of nodes where hydrodynamics and topography create distinct clusters of connectivity. This allowed us to characterize connectivity at a sub-regional level. There are limitations to interpreting connectivity at the regional and local scales depending on the relative scale of analysis. Therefore, it is useful to also know the sub-regional clustering of meadows to narrow research and management considerations. For instance, a field study may need to know where regular exchange occurs to compare populations inside or outside of these subregions. In addition, to underpin marine spatial management, one may be interested in where potential barriers to dispersal exist to designate planning subregions.

The Biophysical Modeling Approach for Understanding Ecological Connectivity

We used a multi-species individual-based biophysical modeling approach to quantify the connectivity of seagrass habitat in the Salish Sea. By using an IBM, we were able to obtain spatially explicit movement information, which allowed us to uncover the actual pathways of movement. By basing our study on a spatially distinct habitat type, we were able to build understanding of the seascape-scale dynamics of an invertebrate community that would otherwise be difficult to track and characterize if not linked to habitat. In addition, by classifying the seascape as a series of transfer probabilities, we not only uncovered direct individual movement, but from this we can also predict the probabilities of multi-generational stepping-stone movement among multiple patches (Crandall et al., 2012; Hock and Mumby, 2015).

We found the extent of the Salish Sea to be an appropriate spatial scale for assessing ecologically relevant connectivity with a biophysical model. The fates of most particles indicate that the Salish Sea is mostly operating as a closed system. However, a small percentage of particles would have exited the system at Juan de Fuca Strait (∼<0.005%). This amount is trivial when considering population dynamics, but it may be relevant to understanding connectivity at evolutionary time scales (Treml et al., 2012). A drift card study found similar patterns (Pawlowicz et al., 2019). When cards were released from various locations the majority stranded within the Salish Sea. However, a few cards were found on the west coast of Vancouver Island and four cards were found 6 months later in Alaska, although mortality and drift time makes this an extremely unlikely scenario for a real organism.

Seagrass Metacommunity Dynamics

Our study is a necessary first step toward uncovering the spatially explicit scale that seagrass communities are functionally connected. Our results complement recent studies of seagrass-associated invertebrate communities in BC that have investigated the drivers of observed spatial biodiversity patterns. In the absence of direct estimates of connectivity that can include oceanographic currents, these studies have inferred dispersal limitation from comparisons of community composition over Euclidean distances. Stark et al. (2020) found low turnover in community composition over 1,000 km of coastline which suggests that dispersal is not limiting the presence of species in seagrass meadows in this region and that most meadows are likely connected by dispersal at least often enough to rescue populations from stochastic extinction. While our results show low probabilities for long distance connections, we found a large number of possible short-distance connections and few completely isolated meadows. Therefore, much of the Salish Sea could be connected through multi-generational stepping-stone dispersal to connect distant meadows, which could explain the community composition patterns found in Stark et al. (2020). Additionally, Stark et al. (2020) found that the subset of taxa present in all sampled meadows represented multiple dispersal strategies, and no one strategy dominated cosmopolitan taxa. Our results are consistent with this finding in that any species that can disperse as larvae or rafting on debris for at least 1+ days could connect most of the Salish Sea through stepping-stone dispersal, and generalists utilizing other habitat types could potentially accomplish this even more efficiently. Another study in BC found that a salinity gradient correlated with abundance patterns of common species (Whippo et al., 2018). However, this gradient did not explain patterns for all groups of species, and some meadows in close proximity and with similar environments had significantly different community compositions. Whippo et al. (2018) speculated that varying dispersal rates influenced by directional hydrodynamics could be structuring these patterns. While our study area did not extend to Barkley Sound, we did find that connection probability can vary substantially (0.0001–99%) over short distances (0–30 km; the maximum distance between meadows in Whippo et al., 2018), suggesting that not all meadows in close proximity are equally connected. Our study provides the first estimates of dispersal for these eelgrass associated organisms that reflect the water currents of the region, going beyond previous, less direct inferences about the possible scales of dispersal limitation in this system (Whippo et al., 2018; Stark et al., 2020). Together, these studies imply that regional biodiversity processes, in addition to local habitat conditions, likely play a role in eelgrass biodiversity.

Our results can be used to generate predictions concerning how the spatiotemporal variation in species' dispersal patterns structure local and regional diversity. These predictions can then be tested with empirical data from field sampling and genetic analysis to indicate if other processes besides those included in the biophysical model are influencing successful dispersal and settlement. While transport is an important part of connectivity (Cowen and Sponaugle, 2009), it is still largely unknown how exchange at the spatial and temporal scale in our model relate to observed biodiversity patterns. Upon arrival in a patch, local environmental conditions and biotic interactions may determine if an individual can actually settle and persist. For instance, biodiversity patterns in seagrass have been shown to be structured by salinity (Whippo et al., 2018), water temperature, seagrass cover, algal biomass (Murphy C. E. et al., 2021), and metrics of fragmentation (Yeager et al., 2019). In addition, microsite selection may occur among nearby meadows (e.g., within a bay) suggesting that transport alone does not determine the final location of an individual (Orth, 1992). A necessary future research direction will be to link transport quantities, abiotic conditions and biotic interactions using metacommunity modeling (e.g., Thompson et al., 2020) to understand realized patterns of seagrass biodiversity.

Lastly, our temporal clustering data can be used when considering the complexities of the shifting spatial scale of metacommunity dynamics through time. Due to environmental variation through time, a static analysis of a metacommunity may not adequately link processes to observed patterns (Stier et al., 2019; Jabot et al., 2020). Seasonal changes to abiotic conditions may alter biotic interactions and reproduction rates which may influence dispersal rates for select species, thus changing community composition through time. While the variation in our clusters is only the result of changing physical ocean dynamics, they once again provide a first prediction for how aspects of the abiotic environment may influence the spatial scale of a metacommunity.

Conservation and Management Applications

A multi-scale, multi-species approach is necessary for the effective management of natural resources across a seascape (Guichard et al., 2004; Pittman, 2018). Dispersal in the context of seagrass habitat requires us to think beyond a single-patch approach to conservation and consider that the true objective may be managing ecological communities through space and time. Functional connectivity is one of the primary design principles for marine protected areas (Aichi Biodiversity Target 11). Reserves that are connected through functional linkages create redundancy and resilience

for important ecosystems, communities, and populations. This ensures that the flow of materials, individuals, and genes is considered across scales (Carr et al., 2017). In BC, eelgrass is a conservation priority that has been targeted for inclusion in the MPA network planning process (Rubidge et al., 2020; Martone et al., 2021), and our connectivity results can be included in future planning efforts for designing a nearshore reserve.

Additional management-related applications of our biophysical model include predicting the consequences of disturbance (e.g., *climate change*, *pollution*, *habitat loss*, and *invasive species*) on connectivity patterns. Through changes in temperature, *climate change* can alter larval development and mortality rates (O'Connor et al., 2007; Lawlor and Arellano, 2020), thus reducing functional connectivity and potentially preventing a species from tracking their environment (Gerber et al., 2014). The parameters in our model can be altered to predict the potential outcome. *Pollution* can also create barriers to dispersal through increased mortality during dispersal (Puritz and Toonen, 2011), and identifying reductions in connectivity will be crucial for understanding the broader regional effects to what may initially seem like a localized problem (Jonsson et al., 2020). In addition, seagrass worldwide is being lost at an alarming rate (Waycott et al., 2009; Dunic et al., 2021), and the consequences of *habitat loss* for biodiversity will depend on the specific connectivity characteristics of the remaining habitat (Thompson et al., 2016). By quantifying the contribution of each meadow or set of meadows to network connectivity, we can predict the consequences of losing a seagrass meadow. Lastly, the degree of direct connectivity and modularity of a network can determine how fast an *invasive species* with passive dispersal can spread. Variance in connectivity and a high degree of clustering can slow the spread of an invasive (Morel-Journel et al., 2018). Currently in the Salish Sea, the spread of the invasive European green crab among seagrass meadows will provide an interesting case study.

Limitations

The interpretive power of the biophysical model could be improved by addressing some of its key assumptions and limitations. (1) We assume passive surface dispersal, and although we believe surface movement captures successful transport between coastal shallow areas, the absolute quantities could be improved with 3D hydrodynamic data and vertical swimming behavior. This would also address the influence that vertical migration may have on connectivity (Metaxas and Saunders, 2009; Snauffer et al., 2014), because diel vertical movement may alter the distance traveled during pelagic dispersal (Paris et al., 2007; Daigle et al., 2016). (2) We assumed that the quality of all seagrass habitat was equal, and that abundance of individuals scaled with area. Future iterations of the model could recognize differences in meadow characteristics and the implications for invertebrate abundance and reproduction. (3) We used a constant daily mortality rate due to a lack of data, but other rates and distributions may more accurately model mortality (e.g., Weibull distribution; Treml et al., 2015). (4) We did not include larval precompetency values (the minimum required time of development before larvae can settle), although the functionality to do so was included in the model. Data on the precompetency period for most species do not exist. While this increases the uncertainty of the timing of settlement on suitable habitat, we found that most settlement occurred in conjunction with coastal stranding. Therefore, the spatial position would remain accurate and only the exact timing of successful settlement would be uncertain. (5) Due to a lack of species-specific data, we assumed that rafting on seagrass and algal debris is a primary mode of passive dispersal for species without a pelagic life-stage. However, for many of these species the exact mode of transport between seagrass meadows is largely unknown (Lefcheck et al., 2016b). (6) The spatial scale of the hydrodynamic model does not match the scale of the smallest meadows ($\sim$30 $\times$ 30 m). Although velocities are interpolated between points, a higher resolution model would be beneficial for reducing uncertainty in these areas. (7) We only modeled potential transport and settlement. A full measure of connectivity requires an individual surviving and reproducing in its destination meadow. Our dispersal results could be combined with population modeling to predict successful connectivity. (8) Lastly, we modeled connectivity between just seagrass habitat. Not all the species we considered are seagrass specialists and some may utilize other habitat types for movement. We view our results as providing a baseline of minimum movement required for seagrass habitat to be connected, but the model could be improved by including other habitat types in the simulation (e.g., kelp).

The model is also limited by the difficulty in validating the results. Physical oceanography studies can partially validate the accuracy of movement by comparing drifter tracks in the Salish Sea to simulated data (Pawlowicz et al., 2019). However, validation of ecological connectivity is more difficult. Genetic similarity data have been used to validate differences in connectivity across large scales (Sunday et al., 2014), but at the spatial scale of the Salish Sea, differences in allele frequencies may not be sufficient to detect differences in regular ecological exchange (Waples, 1998; Riginos et al., 2019). Despite the difficulty in validating our results, we feel we have adequately accounted for the variation that may be present in the system by releasing a large number of particles, incorporating a diffusion value, and averaging across seasons and years.

DATA AVAILABILITY STATEMENT

The code and datasets generated for this study are available at https://doi.org/10.5281/zenodo.5177779. The velocity fields from the SalishSeaCast model are available at https://salishsea.eos.ubc.ca/erddap/index.html.

AUTHOR CONTRIBUTIONS

JC, ER, and MIO conceived of and designed the study. JC wrote the manuscript with input, direction, and review from ER, MIO,

BM, and CF. BM provided code for the analysis and technical direction for the simulations. CF compiled species trait data and provided direction on the theoretical context of the study. JC ran the simulations and analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

This research is sponsored by the NSERC Canadian Healthy Oceans Network and its Partners: Department of Fisheries and Oceans Canada and INREST (representing the Port of Sept-Îles and City of Sept-Îles). It is also funded through the Fisheries and Oceans Canada Strategic Program for Ecosystem-based Research and Assessment (SPERA).

ACKNOWLEDGMENTS

We thank Knut-Frode Dagestad for his technical support with Opendrift, Vincent Traag for his support with the Leidenalg package, and Susan Allen and Mike Foreman for providing early direction on the use of hydrodynamic models. We also thank Patrick Thompson and Matt Whalen for their review of the manuscript, and Sylvia Heredia for the graphic design.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.717469/full#supplementary-material

REFERENCES

Albert, C. H., Rayfield, B., Dumitru, M., and Gonzalez, A. (2017). Applying network theory to prioritize multispecies habitat networks that are robust to climate and land-use change. *Conserv. Biol.* 31, 1383–1396. doi: 10.1111/cobi.12943

Amundrud, S. L., Srivastava, D. S., and O'Connor, M. I. (2015). Indirect effects of predators control herbivore richness and abundance in a benthic eelgrass (*Zostera marina*) mesograzer community. *J. Anim. Ecol.* 84, 1092–1102. doi: 10.1111/1365-2656.12350

Bakri, T., and Jackson, P. (2019). Statistical and synoptic analyses of offshore wind variations. *Int. J. Climatol.* 39, 3201–3217. doi: 10.1002/joc.6012

Bell, S. S. (2006). "Seagrasses and the Metapopulation Concept: Developing a Regional Approach to the Study of Extinction, Colonization, and Dispersal," in *Marine Metapopulations*, eds J. P. Kritzer and P. F. Sale (Elsevier Inc.), 387–408. doi: 10.1016/B978-0-12-088781-1.50014-5

Bode, M., Leis, J. M., Mason, L. B., Williamson, D. H., Harrison, H. B., Choukroun, S., et al. (2019). Successful validation of a larval dispersal model using genetic parentage data. *PLoS Biol.* 17:e3000380. doi: 10.1371/journal.pbio.3000380

Boström, C., Jackson, E. L., and Simenstad, C. A. (2006). Seagrass landscapes and their effects on associated fauna: A review. *Estuar. Coast. Shelf Sci.* 68, 383–403. doi: 10.1016/j.ecss.2006.01.026

Boström, C., Pittman, S. J., Simenstad, C., and Kneib, R. T. (2011). Seascape ecology of coastal biogenic habitats: Advances, gaps, and challenges. *Mar. Ecol. Prog. Ser.* 427, 191–217. doi: 10.3354/meps09051

Boström, C., Törnroos, A., and Bonsdorff, E. (2010). Invertebrate dispersal and habitat heterogeneity: Expression of biological traits in a seagrass landscape. *J. Exp. Mar. Bio. Ecol.* 390, 106–117. doi: 10.1016/j.jembe.2010.05.008

Brooks, R. A., and Bell, S. S. (2001). Mobile corridors in marine landscapes: Enhancement of faunal exchange at seagrass/sand ecotones. *J. Exp. Mar. Bio. Ecol.* 264, 67–84. doi: 10.1016/S0022-0981(01)00310-0

Bryan-Brown, D. N., Brown, C. J., Hughes, J. M., and Connolly, R. M. (2017). Patterns and trends in marine population connectivity research. *Mar. Ecol. Prog. Ser.* 585, 243–256. doi: 10.3354/meps12418

Burgess, S., Nickols, K., Griesemer, C., Barnett, L. A., Dedrick, A., Satterthwaite, E., et al. (2014). Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected area design: supplementary Information. *Ecol. Soc. Am.* 24, 8. doi: 10.1890/13-0710.1

Carr, M. H., Robinson, S. P., Wahle, C., Davis, G., Kroll, S., Murray, S., et al. (2017). The central importance of ecological spatial connectivity to effective coastal marine protected areas and to meeting the challenges of climate change in the marine environment. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 27, 6–29. doi: 10.1002/aqc.2800

Christiaen, B., Ferrier, L., Dowty, P., Gaeckle, J., and Berry, H. (2015). *Puget Sound Seagrass Monitoring Report*. Available online at: https://sites.google.com/a/psemp.org/psemp/home (accessed August 14, 2018).

Cowen, R. K., and Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. *Ann. Rev. Mar. Sci.* 1, 443–466. doi: 10.1146/annurev.marine.010908.163757

Crandall, E. D., Treml, E. A., and Barber, P. H. (2012). Coalescent and biophysical models of stepping-stone gene flow in neritid snails. *Mol. Ecol.* 21, 5579–5598. doi: 10.1111/mec.12031

Dagestad, K. F., Röhrs, J., Breivik, O., and Ådlandsvik, B. (2018). OpenDrift v1.0: A generic framework for trajectory modelling. *Geosci. Model Dev.* 11, 1405–1420. doi: 10.5194/gmd-11-1405-2018

Daigle, R. M., Chasse, J., and Metaxas, A. (2016). The relative effect of behaviour in larval dispersal in a low energy embayment. *Prog. Oceanogr.* 144, 93–117. doi: 10.1016/j.pocean.2016.04.001

D'Aloia, C. C., Bogdanowicz, S. M., Francis, R. K., Majoris, J. E., Harrison, R. G., and Buston, P. M. (2015). Patterns, causes, and consequences of marine larval dispersal. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13940–13945. doi: 10.1073/pnas.1513754112

D'Aloia, C. C., Daigle, R. M., Côté, I. M., Curtis, J. M. R., Guichard, F., and Fortin, M. J. (2017). A multiple-species framework for integrating movement processes across life stages into the design of marine protected areas. *Biol. Conserv.* 216, 93–100. doi: 10.1016/j.biocon.2017.10.012

Duffy, J. E., Reynolds, P. L., Boström, C., Coyer, J. A., Cusson, M., Donadi, S., et al. (2015). Biodiversity mediates top-down control in eelgrass ecosystems: A global comparative-experimental approach. *Ecol. Lett.* 18, 696–705. doi: 10.1111/ele.12448

Dunic, J. C., Brown, C. J., Connolly, R. M., Turschwell, M. P., and Côté, I. M. (2021). Long-term declines and recovery of meadow area across the world's seagrass bioregions. *Glob. Chang. Biol.* 15, 15684. doi: 10.1111/gcb.15684

Engelhard, S. L., Huijbers, C. M., Stewart-Koster, B., Olds, A. D., Schlacher, T. A., and Connolly, R. M. (2017). Prioritising seascape connectivity in conservation using network analysis. *J. Appl. Ecol.* 54, 1130–1141. doi: 10.1111/1365-2664.12824

France, K. E., and Duffy, E. (2006). Diversity and dispersal interactively affect predictability of ecosystem function. *Nature* 441, 1139–1143. doi: 10.1038/nature04729

Gale, K. S., Frid, A., Lee, L., McCarthy, J.-B., Robb, C., Rubidge, E., et al. (2019). A framework for identification of ecological conservation priorities for marine protected area (MPA) network design and its application in the Northern Shelf Bioregion. DFO Can. *Sci. Advis. Secr.* 2018, 186.

Gerber, L. R., Mancha-Cisneros, M. D. M., O'Connor, M. I., and Selig, E. R. (2014). Climate change impacts on connectivity in the ocean: Implications for conservation. *Ecosphere* 5, art33. doi: 10.1890/ES13-00336.1

Gilarranz, L. J., Rayfield, B., Liñán-Cembrano, G., Bascompte, J., and Gonzalez, A. (2017). Effects of network modularity on the spread of perturbation impact in experimental metapopulations. *Science (80-.).* 357, 199–201. doi: 10.1126/science.aal4122

Guichard, F., Levin, S., Hastings, A., and Siegel, D. A. (2004). Toward a dynamic metacommunity approach to marine reserve theory. *Bioscience* 54, 1003–1011.

Gurvan, M., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., et al. (2017). *NEMO ocean engine*. *city pub.

Guzman, L. M., Germain, R. M., Forbes, C., Straus, S., O'Connor, M. I., Gravel, D., et al. (2019). Towards a multi-trophic extension of metacommunity ecology. *Ecol. Lett.* 22, 19–33. doi: 10.1111/ele.13162

Halverson, M., and Pawlowicz, R. (2016). Tide, wind, and river forcing of the surface currents in the fraser river plume. *Atmos. Ocean* 54, 131–152. doi: 10.1080/07055900.2016.1138927

Hanski, I. (2001). Spatially realistic theory of metapopulation ecology. *Nature* 2001, 372–381. doi: 10.1007/s001140100246

Harwell, M. C., and Orth, R. J. (2002). Long-distance dispersal potential in a marine macrophyte. *Ecology* 83, 3319–3330.

Heck, K. L., Carruthers, T. J. B., Duarte, C. M., Randall Hughes, A., Kendrick, G., Orth, R. J., et al. (2008). Trophic transfers from seagrass meadows subsidize diverse marine and terrestrial consumers. *Ecosystems* 11, 1198–1210. doi: 10.1007/s10021-008-9155-y

Heck, K. L., and Thoman, T. A. (1984). The nursery role of seagrass meadows in the upper and lower reaches of the Chesapeake Bay. *Estuaries* 7, 70–92. doi: 10.2307/1351958

Hedgecock, D., Barber, P. H., and Edmands, S. (2003). Genetic Approaches to Measuring Connectivity. *Oceanography* 20, 70–79. doi: 10.1073/pnas.0401921101

Hock, K., and Mumby, P. J. (2015). Quantifying the reliability of dispersal paths in connectivity networks. *J. R. Soc. Interface* 12, 20150013. doi: 10.1098/rsif.2015.0013

Huang, A. C., Essak, M., and O'Connor, M. I. (2015). Top-down control by great blue herons Ardea herodias regulates seagrass-associated epifauna. *Oikos* 124, 1492–1501. doi: 10.1111/oik.01988

Jabot, F., Laroche, F., Massol, F., Arthaud, F., Crabot, J., Dubart, M., et al. (2020). Assessing metacommunity processes through signatures in spatiotemporal turnover of community composition. *Ecol. Lett.* 23, 1330–1339. doi: 10.1111/ele.13523

Jenkins, T. L., and Stevens, J. R. (2018). Assessing connectivity between MPAs: Selecting taxa and translating genetic data to inform policy. *Mar. Policy* 94, 165–173. doi: 10.1016/j.marpol.2018.04.022

Johannesson, K. (1988). The paradox of Rockall: why is a brooding gastropod (Littorina saxatilis) more widespread than one having a planktonic larval dispersal stage (L. littorea)? *Mar. Biol.* 99, 507–513. doi: 10.1007/BF00392558

Jonsson, P. R., Hammar, L., Wåhlström, I., Pålsson, J., Hume, D., Almroth-Rosell, E., et al. (2020). Combining seascape connectivity with cumulative impact assessment in support of ecosystem-based marine spatial planning. *J. Appl. Ecol.* 1–11. doi: 10.1111/1365-2664.13813 *vol,

Källström, B., Nyqvist, A., Åberg, P., Bodin, M., and André, C. (2008). Seed rafting as a dispersal strategy for eelgrass (*Zostera marina*). *Aquat. Bot.* 88, 148–153. doi: 10.1016/j.aquabot.2007.09.005

Kathleen Collins, A., Allen, S. E., and Pawlowicz, R. (2009). The role of wind in determining the timing of the spring bloom in the Strait of Georgia. *Can. J. Fish. Aquat. Sci.* 66, 1597–1616. doi: 10.1139/F09-071

Khangaonkar, T., Nugraha, A., Xu, W., Long, W., Bianucci, L., Ahmed, A., et al. (2018). Analysis of Hypoxia and Sensitivity to Nutrient Pollution in Salish Sea. *J. Geophys. Res. Ocean.* 123, 4735–4761. doi: 10.1029/2017JC013650

Kindlmann, P., and Burel, F. (2008). Connectivity measures: A review. *Landsc. Ecol.* 23, 879–890. doi: 10.1007/s10980-008-9245-4

Kneitel, J. M., and Miller, T. E. (2003). Dispersal Rates Affect Species Composition in Metacommunities of Sarracenia purpurea Inquilines. *Am. Nat.* 162, 165–171.

Kool, J. T., Moilanen, A., and Treml, E. A. (2013). Population connectivity: Recent advances and new perspectives. *Landsc. Ecol.* 28, 165–185. doi: 10.1007/s10980-012-9819-z

LaCasce, J. H. (2008). Statistics from Lagrangian observations. *Prog. Oceanogr.* 77, 1–29. doi: 10.1016/j.pocean.2008.02.002

Lawlor, J. A., and Arellano, S. M. (2020). Temperature and salinity, not acidification, predict near-future larval growth and larval habitat suitability of Olympia oysters in the Salish Sea. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-69568-w

Lefcheck, J. S., Marion, S. R., Lombana, A. V., and Orth, R. J. (2016a). Faunal communities are invariant to fragmentation in experimental seagrass landscapes. *PLoS One* 11:1–24. doi: 10.1371/journal.pone.0156550

Lefcheck, J. S., Marion, S. R., and Orth, R. J. (2016b). Restored Eelgrass (*Zostera marina* L.) as a Refuge for Epifaunal Biodiversity in Mid-Western Atlantic Coastal Bays. *Estuaries and Coasts* 40, 1–13. doi: 10.1007/s12237-016-0141-x

Leibold, M. A., and Chase, J. M. (2018). *Metacommunity Ecology*. Princeton, NJ: Princeton University Press, 59.

Loreau, M., Mouquet, N., and Gonzalez, A. (2003). Biodiversity as spatial insurance in heterogeneous landscapes. *Proc. Natl. Acad. Sci.* 100, 12765–12770. doi: 10.1073/pnas.2235465100

Martone, R. G., Gale, K., Martone, R. G., Robb, C. K., Gale, K. S. P., Frid, A., et al. (2021). *Design Strategies for the Northern Shelf Bioregional Marine Protected Area Network Canadian Science Advisory Secretariat, science advisory Report, 1919-5087; 2019/026*.

Massol, F., Altermatt, F., Gounand, I., Gravel, D., Leibold, M. A., and Mouquet, N. (2017). How life-history traits affect ecosystem properties: effects of dispersal in meta-ecosystems. *Oikos* 126, 532–546. doi: 10.1111/oik.03893

Melià, P., Schiavina, M., Rossetto, M., Gatto, M., Fraschetti, S., and Casagrandi, R. (2016). Looking for hotspots of marine metacommunity connectivity: a methodological framework. *Sci. Rep.* 6, 23705. doi: 10.1038/srep23705

Metaxas, A., and Saunders, M. (2009). Quantifying the " Bio- " Components in Biophysical Models of Larval Transport in Marine Benthic Invertebrates: Advances and Pitfalls. *Biol. Bull.* 216, 257–272. doi: 10.2307/25548159

Milbrandt, J. A., Bélair, S., Faucher, M., Vallée, M., Carrera, M. L., and Glazer, A. (2016). The pan-canadian high resolution (2.5 km) deterministic prediction system. *Weather Forecast.* 31, 1791–1816. doi: 10.1175/WAF-D-16-0035.1

Minor, E. S., and Urban, D. L. (2007). Graph theory as a proxy for spatially explicit populations models in conservation planning. *Ecol. Appl.* 17, 1771–1782.

Morel-Journel, T., Assa, C. R., Mailleret, L., and Vercken, E. (2018). Its all about connections: hubs and invasion in habitat networks. *Ecol. Lett.* 22, 13192. doi: 10.1111/ele.13192

Morrison, J., Foreman, M. G. G., and Masson, D. (2012). A method for estimating monthly freshwater discharge affecting British Columbia coastal waters. *Atmos. Ocean* 50, 1–8. doi: 10.1080/07055900.2011.637667

Mouquet, N., and Loreau, M. (2003). Community Patterns in Source-Sink Metacommunities. *Am. Nat.* 162, 544–557. doi: 10.1086/378857

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science (80-.).* 328, 876–878. doi: 10.1126/science.1184819

Murphy, C. E., Orth, R. J., and Lefcheck, J. S. (2021). Habitat Primarily Structures Seagrass Epifaunal Communities: a Regional-Scale Assessment in the Chesapeake Bay. *Estuaries and Coasts* 44, 442–452. doi: 10.1007/s12237-020-00864-4

Murphy, G. E. P., Dunic, J. C., Adamczyk, E. M., Bittick, S. J., Côté, I. M., Cristiani, J., et al. (2021). From coast to coast to coast: ecology and management of seagrass ecosystems across Canada. *Facets* 6, 1–41. doi: 10.1139/facets-2020-0020

O'Connor, M. I., Bruno, J. F., Gaines, S. D., Halpern, B. S., Lester, S. E., Kinlan, B. P., et al. (2007). Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1266–1271. doi: 10.1073/pnas.0603422104

Olson, E. M., Allen, S. E., Do, V., Dunphy, M., and Ianson, D. (2020). Assessment of Nutrient Supply by a Tidal Jet in the Northern Strait of Georgia Based on a Biogeochemical Model. *J. Geophys. Res. Ocean.* 125, 1–25. doi: 10.1029/2019JC015766

Orth, R. J. (1992). "A Perspective on Plant-Animal Interactions in Seagrasses: Physical and Biological Determinats influencing Plant and Animal Abundance," in *Plant-Animal Interactions in the Marine Benthos*, eds D. M. John, S. J. Hawkins, and J. H. Price (Oxford : Clarendon Press), 147–164.

Orth, R. J., Heck, K. L., and van Montfrans, J. (1984). Faunal communities in seagrass beds: A review of the influence of plant structure and prey characteristics on predator-prey relationships. *Estuaries* 7, 339–350. doi: 10.2307/1351618

Paris, C. B., Chérubin, L. M., and Cowen, R. K. (2007). Surfing, spinning, or diving from reef to reef: Effects on population connectivity. *Mar. Ecol. Prog. Ser.* 347, 285–300. doi: 10.3354/meps06985

Pawlowicz, R., Hannah, C., and Rosenberger, A. (2019). Lagrangian observations of estuarine residence times, dispersion, and trapping in the Salish Sea. *Estuar. Coast. Shelf Sci.* 225, 106246. doi: 10.1016/j.ecss.2019.106246

Pawlowicz, R., Riche, O., and Halverson, M. (2007). The circulation and residence time of the Strait of Georgia using a simple mixing-box approach. *Atmos. Ocean* 45, 173–193. doi: 10.3137/ao.450401

Pereira, M., Segurado, P., and Neves, N. (2011). Using spatial network structure in landscape management and planning: A case study with pond turtles. *Landsc. Urban Plan.* 100, 67–76. doi: 10.1016/j.landurbplan.2010.11.009

Pielou, E. (1991). *After the ice age: The return of life to glaciated North America 1991*. Chicago, LI: University of Chicago Press.

Pittman, S. J. (2018). *Seascape Ecology*. Hoboken, NJ: Wiley-Blackwell.

Pittman, S. J., Kneib, R. T., and Simenstad, C. A. (2011). Practicing coastal seascape ecology. *Mar. Ecol. Prog. Ser.* 427, 187–190. doi: 10.3354/meps09139

Puritz, J. B., and Toonen, R. J. (2011). Coastal pollution limits pelagic larval dispersal. *Nat. Commun.* 2, 226. doi: 10.1038/ncomms1238

Riginos, C., Hock, K., Matias, A. M., Mumby, P. J., van Oppen, M. J. H., and Lukoschek, V. (2019). Asymmetric dispersal is a critical element of concordance between biophysical dispersal models and spatial genetic structure in Great Barrier Reef corals. *Divers. Distrib.* 25, 1684–1696. doi: 10.1111/ddi.12969

Rubidge, E., Jeffery, S., Gregr, E. J., Gale, K. S. P., and Frid, A. (2020). *Assessment of nearshore features in the Northern Shelf Bioregion against criteria for determining Ecologically and Biologically Significant Areas (EBSAs). Report number 2020/023*.

Rumrill, S. S. (1990). Natural mortality of marine invertebrate larvae. *Ophelia* 32, 163–198. doi: 10.1080/00785236.1990.10422030

Saura, S., Bodin, Ö, and Fortin, M. J. (2014). EDITOR'S CHOICE: Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *J. Appl. Ecol.* 51, 171–182. doi: 10.1111/1365-2664.12179

Saura, S., and Rubio, L. (2010). A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. *Ecography* 33, 523–537. doi: 10.1111/j.1600-0587.2009.05760.x

Saura, S., and Torné, J. (2009). Conefor Sensinode 2.2: A software package for quantifying the importance of habitat patches for landscape connectivity. *Environ. Model. Softw.* 24, 135–139. doi: 10.1016/j.envsoft.2008.05.005

Schill, S. R., Raber, G. T., Roberts, J. J., Treml, E. A., Brenner, J., and Halpin, P. N. (2015). No reef is an island: Integrating coral reef connectivity data into the design of regional-scale marine protected area networks. *PLoS One* 10:1–24. doi: 10.1371/journal.pone.0144199

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., et al. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Mar. Ecol. Prog. Ser.* 554, 1–19. doi: 10.3354/meps11792

Shanks, A. L. (2009). Pelagic larval duration and dispersial distance revisited. *Biol. Bull.* 216, 373–385. doi: 10.2307/25548167

Siegel, D. A., Kinlan, B. P., Gaylord, B., and Gaines, S. D. (2003). Lagrangian descriptions of marine larval dispersion. *Mar. Ecol. Prog. Ser.* 260, 83–96. doi: 10.3354/meps260083

Snauffer, E. L., Masson, D., and Allen, S. E. (2014). Modelling the dispersal of herring and hake larvae in the Strait of Georgia for the period 2007-2009. *Fish. Oceanogr.* 23, 375–388. doi: 10.1111/fog.12072

Soontiens, N., and Allen, S. E. (2017). Modelling sensitivities to mixing and advection in a sill-basin estuarine system. *Ocean Model.* 112, 17–32. doi: 10.1016/j.ocemod.2017.02.008

Soontiens, N., Allen, S. E., Latornell, D., Le Souëf, K., Machuca, I., Paquin, J.-P., et al. (2016). Storm Surges in the Strait of Georgia Simulated with a Regional Model. *Atmosphere-Ocean* 54, 1–21. doi: 10.1080/07055900.2015.1108899

Stark, K., Thompson, P., Yakimishyn, J., Lee, L., Adamczyk, E., Hessing-Lewis, M., et al. (2020). Beyond a single patch: local and regional processes explain diversity patterns in a seagrass epifaunal metacommunity. *Mar. Ecol. Prog. Ser.* 655, 91–106. doi: 10.3354/meps13527

Stier, A. C., Lee, S. C., and O'Connor, M. I. (2019). Temporal variation in dispersal modifies dispersal-diversity relationships in an experimental seagrass metacommunity. *Mar. Ecol. Prog. Ser.* 613, 67–76. doi: 10.3354/meps12908

Sunday, J. M., Popovic, I., Palen, W. J., Foreman, M. G. G., and Hart, M. W. (2014). Ocean circulation model predicts high genetic structure observed in a long-lived pelagic developer. *Mol. Ecol.* 23, 5036–5047. doi: 10.1111/mec.12924

Thomas, C. J., Lambrechts, J., Wolanski, E., Traag, V. A., Blondel, V. D., Deleersnijder, E., et al. (2014). Numerical modelling and graph theory tools to study ecological connectivity in the Great Barrier Reef. *Ecol. Modell.* 272, 160–174. doi: 10.1016/j.ecolmodel.2013.10.002

Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., et al. (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecol. Lett.* 23, 1314–1329. doi: 10.1111/ele.13568

Thompson, P. L., Rayfield, B., and Gonzalez, A. (2016). Loss of habitat and connectivity erodes species diversity, ecosystem functioning, and stability in metacommunity networks. *Ecography* 40, 98–108. doi: 10.1111/ecog.02558

Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-41695-z

Treml, E. A., Ford, J. R., Black, K. P., and Swearer, S. E. (2015). Identifying the key biophysical drivers, connectivity outcomes, and metapopulation consequences of larval dispersal in the sea. *Mov. Ecol.* 3, 17. doi: 10.1186/s40462-015-0045-6

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., and Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integr. Comp. Biol.* 52, 525–537. doi: 10.1093/icb/ics101

Urban, D. L., Minor, E. S., Treml, E. A., and Schick, R. S. (2009). Graph models of habitat mosaics. *Ecol. Lett.* 12, 260–273. doi: 10.1111/j.1461-0248.2008.01271.x

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Hered.* 89, 438–450. doi: 10.1093/jhered/89.5.438

Waycott, M., Duarte, C. M., Carruthers, T. J. B., Orth, R. J., Dennison, W. C., Olyarnik, S., et al. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci.* 106, 12377–12381. doi: 10.1073/pnas.0905620106

Werner, F. E., Cowen, R. K., and Paris, C. B. (2007). Coupled Biological and Physical Models. *Oceanography* 20, 54–69. doi: 10.1016/S0967-0645(00)00079-5

Whippo, R., Knight, N. S., Prentice, C., Cristiani, J., Siegle, M. R., and O'Connor, M. I. (2018). Epifaunal diversity patterns within and among seagrass meadows suggest landscape-scale biodiversity processes. *Ecosphere* 9, e02490. doi: 10.1002/ecs2.2490

White, J. W., Morgan, S. G., and Fisher, J. L. (2014). Planktonic larval mortality rates are lower than widely expected. *Ecology* 95, 3344–3353. doi: 10.1890/13-2248.1

Worcester, S. E. (1994). Adult rafting versus larval swimming: dispersal and recruitment of a botryllid ascidian on eelgrass. *Mar. Biol.* 121, 309–317. doi: 10.1007/BF00346739

Wren, J. L. K., Kobayashi, D. R., Jia, Y., and Toonen, R. J. (2016). Modeled population connectivity across the Hawaiian archipelago. *PLoS One* 11:1–25. doi: 10.1371/journal.pone.0167626

Yamada, K., Tanaka, Y., Era, T., and Nakaoka, M. (2014). Environmental and spatial controls of macroinvertebrate functional assemblages in seagrass ecosystems along the Pacific coast of northern Japan. *Glob. Ecol. Conserv.* 2, 47–61. doi: 10.1016/j.gecco.2014.08.003

Yeager, L. A., Geyer, J. K., and Fodrie, F. J. (2019). Trait sensitivities to seagrass fragmentation across spatial scales shape benthic community structure. *J. Anim. Ecol.* 88, 1743–1754. doi: 10.1111/1365-2656.13067

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fmars.2021.608665

Quantifying Patterns in Fish Assemblages and Habitat Use Along a Deep Submarine Canyon-Valley Feature Using a Remotely Operated Vehicle

Benjamin J. Saunders[1], Ronen Galaiduk[2], Karina Inostroza[3], Elisabeth M. V. Myers[4], Jordan S. Goetze[1,5], Mark Westera[3], Luke Twomey[6], Denise McCorry[7] and Euan S. Harvey[1]*

[1] School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia, [2] Australian Institute of Marine Science, Indian Ocean Marine Research Centre (IOMRC), The University of Western Australia, Crawley, WA, Australia, [3] BMT Commercial Australia, Osborne Park, WA, Australia, [4] New Zealand Institute for Advanced Study, Massey University, Albany, New Zealand, [5] Department of Biodiversity, Conservation and Attractions, Marine Science Program, Biodiversity and Conservation Science, Kensington, WA, Australia, [6] Western Australian Marine Science Institution, Crawley, WA, Australia, [7] Woodside Energy Ltd., Perth, WA, Australia

OPEN ACCESS

Edited by:
Pierre Petitgas, Institut Français de Recherche pour l'Exploitation de la Mer (Ifremer), France

Reviewed by:
Chih-Lin Wei, National Taiwan University, Taiwan
Marta M. Varela, Spanish Institute of Oceanography, Spain

***Correspondence:**
Benjamin J. Saunders
Ben.Saunders@curtin.edu.au

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: *21 September 2020*
Accepted: *16 August 2021*
Published: *07 September 2021*

Citation:
Saunders BJ, Galaiduk R, Inostroza K, Myers EMV, Goetze JS, Westera M, Twomey L, McCorry D and Harvey ES (2021) Quantifying Patterns in Fish Assemblages and Habitat Use Along a Deep Submarine Canyon-Valley Feature Using a Remotely Operated Vehicle. Front. Mar. Sci. 8:608665. doi: 10.3389/fmars.2021.608665

The aim of this study was to document the composition and distribution of deep-water fishes associated with a submarine canyon-valley feature. A work-class Remotely Operated Vehicle (ROV) fitted with stereo-video cameras was used to record fish abundance and assemblage composition along transects at water depths between 300 and 900 metres. Three areas (A, B, C) were sampled along a submarine canyon-valley feature on the continental slope of tropical north-western Australia. Water conductivity/salinity, temperature, and depth were also collected using an ROV mounted Conductivity Temperature and Depth (CTD) instrument. Multivariate analyses were used to investigate fish assemblage composition, and species distribution models were fitted using boosted regression trees. These models were used to generate predictive maps of the occurrence of four abundant taxa over the survey areas. CTD data identified three water masses, tropical surface water, South Indian Central Water (centred ~200 m depth), and a lower salinity Antarctic Intermediate Water (AAIW) ~550 m depth. Distinct fish assemblages were found among areas and between canyon-valley and non-canyon habitats. The canyon-valley habitats supported more fish and taxa than non-canyon habitats. The fish assemblages of the deeper location (~700–900 m, Area A) were different to that of the shallower locations (~400–700 m, Areas B and C). Deep-water habitats were characterised by a *Paraliparis* (snail fish) species, while shallower habitats were characterised by the family Macrouridae (rat tails). Species distribution models highlighted the fine-scale environmental niche associations of the four most abundant taxa. The survey area had a high diversity of fish taxa and was dominated by the family Macrouridae. The deepest habitat had a different fish fauna to the shallower areas. This faunal break can be attributed to the influence of AAIW. ROVs provide a platform

on which multiple instruments can be mounted and complementary streams of data collected simultaneously. By surveying fish *in situ* along transects of defined dimensions it is possible to produce species distribution models that will facilitate a greater insight into the ecology of deep-water marine systems.

Keywords: deep-water, habitat, ROV, stereo-video, CTD, species distribution model, submarine canyon, north-western Australia

INTRODUCTION

The deep sea is the largest environment on earth (Levin et al., 2001) and plays a pivotal role in cycling nutrients and water at global scales. It is also a major source of chemosynthetic primary productivity (Armstrong et al., 2012; Danovaro et al., 2014; Jobstvogt et al., 2014), and it plays a major role in climate regulation by absorbing heat from the atmosphere and sequestering carbon at the seafloor (Armstrong et al., 2012; Rogers, 2015). A number of commercially important industries such as fisheries, oil and gas, minerals and pharmaceuticals operate in deep-sea ecosystems (Armstrong et al., 2012). Deep-sea environments are characterised by a complex suit of geomorphic features such as underwater shoals, banks and canyons (Agapova et al., 1979; Heap and Harris, 2008) which provide diverse and structurally complex environments for marine organisms. The ocean is a major source of undocumented biodiversity, with experts predicting that between one to two-thirds of all marine species are undescribed, many of which reside in the deep sea (Appeltans et al., 2012).

Conducting research in deep-sea environments is particularly challenging due to a lack of natural light, high water pressure, low temperature, low oxygen levels and the need for large vessels and advanced technologies to reach these depths. Despite their remoteness, deep-water habitats are susceptible to numerous anthropogenic impacts on global and local scales, including climate change (Hoegh-Guldberg and Bruno, 2010; Rogers, 2015), overfishing (Clark, 2001), plastic pollution (Woodall et al., 2014), exploration and development activities such as drilling (Jones, 2009) or accidents such as oil spills (White et al., 2012). It is therefore important that quantitative data on these ecosystems is obtained in order to assess their vulnerability and ability to recover from disturbance, and to mitigate anthropogenic impacts on these ecosystems.

Much of the research on deep-water fish populations has utilised destructive sampling methods such as bottom trawling (Williams et al., 2001; Tolimieri and Anderson, 2010; Cruz-Acevedo et al., 2018). Trawl surveys are often completed across large spatial scales and work to aggregate samples, resulting in a decreased ability to provide fine-scale descriptions of organisms and their habitat associations (Cappo et al., 2004). A combination of improved technology and concerns over the use of destructive sampling techniques, has increased the use of non-destructive, camera-based surveys. Methods such as landers and baited remote underwater video (BRUV) are well suited to sampling deep-water fish assemblages given they are not limited by depth and use bait to attract fish to the camera system (Priede and Bagley, 2000; Zintzen et al., 2012; McLean et al., 2015). While remote video is a cost effective and statistically powerful way of sampling fish diversity across a gradient of habitats and depths, it is less suitable for finer scale sampling (<100 s of m) and only provides relative estimates of abundance due to variation in the distance a bait plume travels and the attraction of different fish species to the bait (Cappo et al., 2004; Watson et al., 2010; Galaiduk et al., 2017b). Technologies involved with the use of remote operated vehicles (ROVs) have also developed rapidly, allowing time efficient data collection at a fine-spatial resolution, across a wide range of depths (Jones, 2009; Sward et al., 2019; McLean et al., 2020). They also provide a platform on which multiple scientific instruments can be mounted, to facilitate multi-purpose surveys.

Observations made using ROVs provide contextual information about fishes that relate aspects of their population structure and function to habitat, which traditional trawl methods would otherwise overlook (Adams et al., 1995; Macreadie et al., 2018). ROVs have been used to study the impact of deep-water fishing (Puig et al., 2012; Bo et al., 2014), assess benthic associations of fishes with oil and gas structures (Hudson et al., 2005; Jones, 2009; Bond et al., 2018; McLean et al., 2018b; Schramm et al., 2020a, 2021), provide behavioural observations (Lorance and Trenkel, 2006; Gates et al., 2017), and to collect fragile specimens (Macreadie et al., 2018). Stereo-video, the use of two cameras to facilitate accurate estimates of the length of fish (Harvey et al., 2001) and to standardise a sampling area (Harvey et al., 2004), has also developed rapidly and it is now possible to attach a stereo-video system to an ROV and complete transect based fauna surveys (Schramm et al., 2020b). In addition, GPS position overlay can provide information on the precise location of observations. These complementary data streams are well suited to a spatial analysis framework such as species distribution modelling (SDMs), which allows the fitted models to be extrapolated into un-sampled locations using benthic environmental predictors derived from acoustic surveys. These models can help us to understand the ecology of these understudied species and map their environmental niche associations for future studies and management applications.

Although many studies have examined the biodiversity patterns of shallow water fishes, comparatively little is known about how these patterns change with increasing depth. It is common to observe a decrease in species richness of fishes with increasing depth (Moranta et al., 1998; Lorance et al., 2002; Tolimieri, 2007; Wellington et al., 2018), though this pattern can be reversed depending on the scale of depth examined (e.g., McClatchie et al., 1997; Mindel et al., 2016). Our study area was a canyon-valley feature located on the continental slope in the northwest shelf region of Western

Australia. The study area traversed two Key Ecological Features (KEFs): (1) Canyons linking the Cuvier Abyssal Plain and the Cape Range Peninsula, and (2) Continental Slope Demersal Fish Communities (DSEWPaC, 2012). KEFs are elements of the Australian Commonwealth marine environment that are considered to be of regional importance for either a region's biodiversity or its ecosystem function and integrity (DSEWPaC, 2012). Previous deep water research in Northwest Australia has identified high fish species and family level richness when compared to assemblages across similar depth ranges elsewhere in the world (Williams et al., 2001), and in a pattern similar to previous studies fish species richness declined with increasing depth from 78 to 825 m (McLean et al., 2018b). Our study set out to assess fish distribution and diversity, to better define these patterns and to add to our understanding of the deep-water ecology of the Northwest system, and the canyon and fish community KEFs.

To increase efficiency in the field, we used a multi-task stereo-ROV platform to survey fish assemblages in the northwest of Australia between the depths of 420–870 m. To the best of our knowledge, there have been no published studies that combine stereo-video technology with ROV transects to assess fish assemblages in deeper continental slope waters. Using a stereo-ROV we collected data on habitat, fish assemblages and water quality simultaneously. Our objective was to describe the abundance, composition and size of deep-water fish assemblages in three areas along a canyon-valley feature. We also fitted SDMs to document environmental niche associations of the four most abundant fish taxa and assessed the applicability of this approach to improve our understanding of the spatial ecology of deep-sea fishes.

MATERIALS AND METHODS

Site Description

The study was conducted on the continental slope in the Greater Enfield region, located approximately 45 km from the North West Cape of Western Australia (**Figure 1**) at depths between 420 and 870 m. Three areas (A, B, C) ranging from deepest (A, 870 m maximum) to shallowest (C, 420 m minimum depth) were surveyed along a continuous submarine canyon-valley feature. The feature is not included in the National Submarine Canyon Database, and in comparison to those listed in the database is a relatively small tributary to the north of the Cape Range Canyon KEF. Area A was the deepest site (between 790 and 870 m) and had a wide (250–500 m) and topographically shallow (20–50 m) valley feature. The valley feature at Area B was narrower (600–250 m wide), topographically shallow (50–80 m) and traversed water depths of between 590 and 690 m. Area C was the shallowest survey site, in water depths ranging from 420 to 560 m, and where the canyon was more pronounced, being the narrowest of the three canyons and topographically the deepest (250–300 m wide, 200–250 m deep). The survey encompassed flat seabed and canyon-valley features as identified *a priori* from multi-beam bathymetry and derivatives. Following the criteria described in Huang et al. (2018) at Area C the feature is a shelf incising canyon, which is narrow, steep walled and deep. However, at areas B and C the feature is wider and topographically shallower, so at these areas it has transitioned into a valley on the continental slope (Huang et al., 2018). All three areas surveyed were predominantly soft sediment bottom, with sparse higher rugosity hard bottom features. For simplicity when describing statistical analysis, the term "canyon" is used to describe both the canyon and valley sections of the feature throughout the methods and results sections.

Survey Method

The study was conducted over three consecutive days from 31 October to 2 November 2015. The survey was conducted using a Centurion QX 312 work class ROV fitted with stereo-video cameras to conduct transects with a standardised width of 5 m. A single downward facing video camera was used to simultaneously record benthic habitat. Water quality data (conductivity/salinity, temperature, and depth; CTD) was collected with a Sea-Bird Electronics (SBE) 19plus V2 SeaCAT Profiler CTD, mounted to the ROV platform. The ROV operated at approximately 1.5 m above the seabed and at a forward speed of approximately 0.5 knots. The total transect length covered by the ROV at Area A was approximately 11,330 m, Area B 3,470 m, and C 5,990 m.

The stereo-video system was made up of two Sony HDR CX550 handycams within custom built underwater housings with a depth rating to 2,000 m. The stereo-cameras were mounted onto a rigid base bar with a separation of 700 mm between the cameras and were inwardly converged at an angle of 8°, following the principles outlined by Harvey and Shortis (1995) and Goetze et al. (2019). This configuration provided an overlapping field of view from approximately 0.5 m in front of the cameras and accurate length measurements out to 8 m (Harvey et al., 2010). Stereo-video footage was recorded at high definition with a 1,920 × 1,080 resolution. A second stereo-video system was also fitted as redundancy in case of failure of the primary system. The stereo-video systems were mounted as low as possible onto the ROV (**Figure 2**), and camera systems were interfaced with and powered from the ROV systems. The ROV communication channels allowed the systems to be remotely controlled, and a live standard definition video stream from each camera was fed to the surface with live position, depth, date, and time overlays. To maximise illumination of the field of view, a combination of high output Light Emitting Diode (LED) and High Intensity Displacement (HID) lighting was used. These were placed as high as possible above the camera system on the ROV to reduce backscatter from suspended particles in the water column.

Identification of Taxa, Three Dimensional Positioning and Length Measurements

Video footage was analysed using EventMeasure Stereo software (SeaGIS, 2014). Identifications of fish were made by one experienced researcher based upon morphological features. In some cases, identification to species level was not possible as identifying features could not be distinguished on the video footage. In such cases, a precautionary approach was taken and

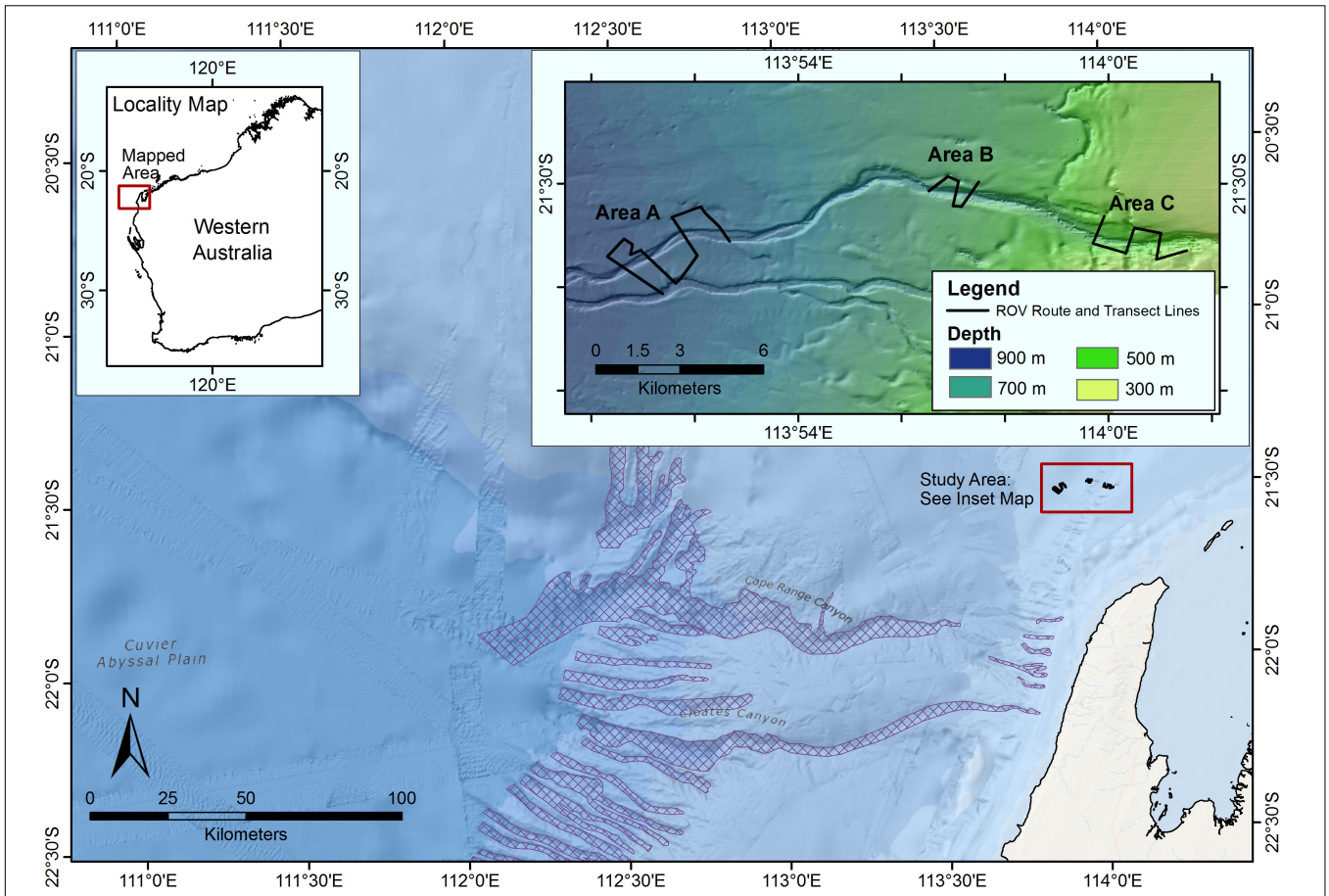

FIGURE 1 | Context map illustrating the geographical location of the Greater Enfield survey area. Inset map illustrating the transect layout at each of the three survey Areas along a canyon-valley feature. Submarine canyons that are included in the Australian national submarine canyon database (Huang et al., 2014) are highlighted in cross-hatch (Source: Geoscience Australia).

taxa were identified to the lowest taxonomic resolution possible. Distinct taxa were given a unique number, which facilitated the assessment of diversity.

The calibrated stereo-video system (see Boutros et al., 2015) allowed accurate and precise measurements of the fork length (tip of nose to the middle caudal fin rays) of fishes, and the identification of the three dimensional position of each fish in relation to the ROV. To ensure fish were within the transect area (a 5 m belt), length measurements further than 2.5 m to either side or greater than 7 m in front of the ROV (based on the minimum visibility), were automatically rejected by the EventMeasure software.

Data Presentation and Statistical Analysis

The collection of data over transects of known position and dimensions allowed the use of two statistical approaches: Firstly, a multivariate analysis of variance approach was used to investigate differences in the fish assemblage structure between the three areas, and between canyon and non-canyon habitats. Secondly, a species distribution modelling approach was used to investigate environmental drivers of the distribution of taxa that were characteristic of each area, and to identify one taxa that was common and widely distributed over all three areas. These models were used to generate continuous predictive maps of the occurrence of these taxa over the survey areas.

Approach 1, Assemblage Patterns

Definition of canyon and non-canyon sections

Using bathymetry and slope data and the known depth profile of the ROV track log, transects were separated into two habitat types: canyon and non-canyon. Canyon habitats were defined as benthic habitat with increased depth, slope and structure, while the flat seabed surrounding the feature was classified as non-canyon habitat. Where a transect crossed over a canyon-valley habitat, the depth was noted outside of the canyon on both sides and the maximum depth inside the canyon-valley feature was also recorded (depths labelled A, B, and C, respectively, on **Supplementary Figure 1.1**). Taking a conservative approach and to avoid areas of transition between the two habitat types, a buffer section between canyon-valley and non-canyon was excluded from analyses. This buffer was calculated as 10% of the difference between the maximum canyon-valley depth and the

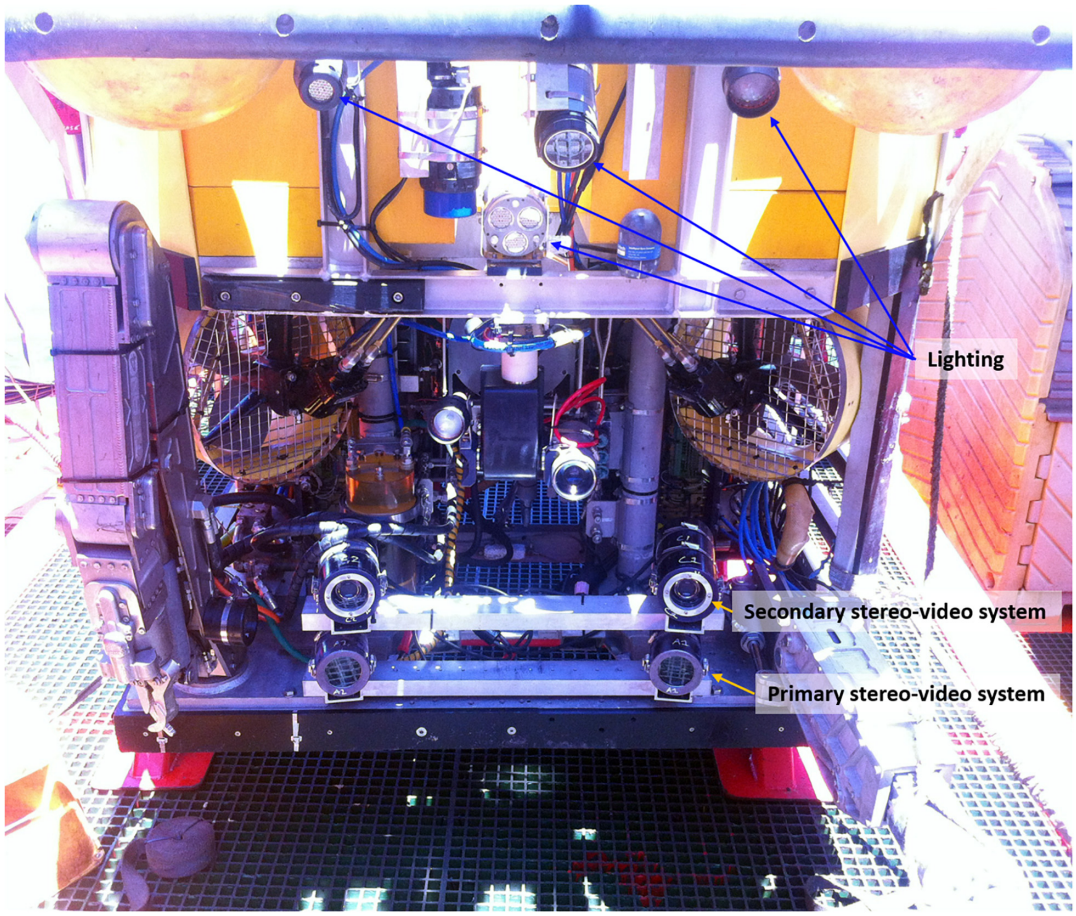

FIGURE 2 | Front view of the QX312 ROV showing the position of the stereo-video camera systems and the lighting used during the survey. CTD was mounted to the rear of the ROV.

depths outside the canyon-valley feature (C - A and C - B on **Supplementary Figure 1.1**). Video footage from the remaining depth of the canyon-valley feature (orange line in **Supplementary Figure 1.1**) was used to determine fish assemblage structure inside canyon-valley habitats.

Within each surveyed area (A, B, C), every time the transect crossed a canyon-valley feature, the video footage within the canyon-valley habitat was treated as a single replicate. As such, Area A had six replicate transect sections within canyon-valley habitat (**Supplementary Figure 1.2**), while Areas B and C both had five replicate transect sections within canyon-valley habitat (**Supplementary Figure 1.2**). To ensure comparability between habitats, similar replication occurred for non-canyon habitats such that Area A had six replicate transect sections and Area B had five. However, much of the area surveyed at Area C was steeply sloping and only two sections could be classified as non-canyon (**Supplementary Figure 1.2**; **Supplementary Table 1.1**).

Standardisation of data for assemblage analysis

To ensure that transect sections were statistically comparable, transect length and densities of fishes underwent three levels of standardisation prior to statistical analyses: Firstly, sections of transect were removed when the seabed was not within the field of view (**Supplementary Figure 1.2**). Secondly, the length of transects between canyon-valley and non-canyon habitats were standardised to one another. To achieve this, the total length of transect within each habitat was determined within each area, and then averaged. A note was taken of the habitat type with the shortest average transect length at each area. This shorter average length was then used to randomly select replicate segments of transects of that length in the other habitat in that area. If a transect section was shorter than the average length, its original length was used (**Supplementary Table 1.1**). Lastly, measures of fish assemblages from each transect section (as defined above) were standardised to a 250 m length by 5 m wide transect, or 1,250 m^2. This transect length was used to increase the number of observations per transect for deep-sea fishes that are relatively sparse compared to shallower water fishes, but is similar to the length used other studies (Watson and Ormond, 1994; Newman et al., 1997; Westera et al., 2003). This allowed for fish assemblages within canyon-valley features of different sizes to be compared to each other and to non-canyon habitats. These standardised transects formed the replicates of the "canyon" and "non-canyon" levels within the factor "Habitats" in the statistical analysis design.

The fish assemblage abundance data and the total number of fish for each transect were standardised in this way. However, the number of taxa for each transect was not standardised to a 1,250 m^2 transect area. This was to avoid artificially inflating the number of taxa observed in any transect. The influence of varying transect length on the number of taxa was assessed by including transect length as a covariate in the initial analysis. The interactions between transect length and the factors of Area and Habitat were not significant (all $p > 0.20$). Therefore we considered that transect length was not a confounding variable for tests of the number of taxa, and was not included in further statistical analysis.

Assemblage analysis approach

To investigate whether there were differences in the fish assemblage composition between the areas and habitats, a two-factor statistical design was used. The two factors tested were Areas (fixed factor, orthogonal; Area A vs. Area B vs. Area C) and Habitats (fixed factor, orthogonal; canyon vs. non-canyon). The standardised fish assemblage composition data was analysed using PERMANOVA (non-parametric analysis of variance; Anderson et al., 2008) add-in to the PRIMER v6 statistical software (Clarke and Gorley, 2006). The data were investigated for homogeneity of variance for the factors Habitat and Area using PERMDISP (permutational analysis of multivariate dispersions, Anderson, 2006). No taxa were numerically dominant and data was not transformed as the data met the assumption of homogeneity of variance. The multivariate fish assemblage composition was tested using PERMANOVA based upon a Bray Curtis similarity matrix. Number of taxa and the total number of fish were analysed separately using a univariate PERMANOVA based on a Euclidean distance resemblance matrix (Anderson, 2006; Clarke et al., 2006a,b). Where a significant difference ($p < 0.05$) was detected for the factor Area, *post hoc* tests were conducted to determine which of the Areas were significantly different.

To illustrate the multivariate patterns in the fish assemblage, a Principle Coordinate Ordination (PCO) was presented with vectors overlaid (Anderson et al., 2008). The vectors illustrate the strength and direction of the spearman rank correlation between the density of the taxa and the PCO axes. Taxa with a correlation stronger than 0.4 to either PCO axes were plotted. Results of the univariate analyses (number of taxa and total number of fish) were presented using bar graphs of means and standard error (SE).

Approach 2, Species Distribution Models

We developed individual Species Distribution Models (SDMs) for four fish taxa that were observed most frequently in the ROV videos. These models can help understand the ecology of these understudied taxa and map their environmental niche associations for future studies. The SDMs were developed using boosted regression trees (BRT) and "gbm" package applied in R software (R Core Team, 2014). BRT is a machine learning algorithm for additive numerical optimisation of the loss function to iteratively increase the predictive performance of the final model while gradually emphasising poorly modelled observations in the existing collection of trees (Elith et al., 2006). In the last decade, it has gained popularity within the marine spatial community because it is insensitive to outliers, missing data, or monotone transformations, and can be easily used with any type of predictors such as numeric, binary, categorical (Pittman et al., 2007; Oyafuso et al., 2017; Stamoulis et al., 2018). Fitted values in the final model are computed as the sum of all trees multiplied by the learning rate and are much more stable and accurate than those from a single decision tree model (Elith et al., 2006).

The occurrence data for each of the four taxa was extracted from direct observations of each occurrence from the underwater video recording along the ROV tracks and were using in BRT fitting as records of taxa presence. As BRT models also require a sample of observations to characterise the available environment to discriminate used from available habitat (pseudo-absences, Boyce et al., 2002; Phillips et al., 2009; Franklin, 2010) we derived pseudo-absences for all taxa by randomly sampling all the available data points along the ROV tracks where the study taxa were not recorded. Because presence records for all study taxa were low, we created a final ratio of 1:1 of the observed presences and pseudo-absences of each study taxa along the tracks to effectively estimate unbiased parameters for rare populations (Fielding and Bell, 1997; Franklin, 2010; Galaiduk et al., 2017b). The final datasets were partitioned into training (75%) and testing (25%) data for individual modelled taxa and tested for spatial autocorrelation between observations. The explanatory variables were a set of 9 functionally relevant environmental predictors with Spearman's rank correlation between them <0.7 which is considered to be acceptable for spatial models (Moore et al., 2011; Galaiduk et al., 2017a; **Table 1**) that describe the structure, complexity and type of benthic habitat derived from the bathymetry data using Spatial Analyst toolkit in ArcGIS 10.3. The 10th variable, "Habitat type," was categorical, and described occurrence of benthic biota and signs of bioturbation. This was derived through unsupervised classification procedures in ERDAS using direct observation along the ROV transects and post processed with the Spatial Analyst toolkit in ArcGIS (**Table 1**).

To determine the effect of environmental predictors and their importance on the probability of occurrence of four fish taxa, we fitted BRT models on training datasets for these taxa following the procedures outlined in Elith et al. (2008). Optimal model settings were chosen using 10-fold cross-validation by optimising learning rate, bagging rate and tree complexity (Leathwick et al., 2006). The optimal model was considered a model that produced the lowest cross-validated residual deviance with at least 1,000 fitted trees. Selected models were then simplified to remove less informative predictors which achieved more parsimonious models without degradation of model fits (Elith et al., 2008). The importance of predictor variables in the simplified BRT models was determined using the variable importance score based on the improvements of all splits associated with a given variable across all trees when this variable was added in the model (Leathwick et al., 2006).

To assess the predictive performance, and discrimination and accuracy of fitted models, a set of common evaluation metrics of predictive performance was calculated on the test datasets.

TABLE 1 | Summary of the environmental predictors extracted from the hydroacoustic survey used for fitting boosted regression trees.

Environmental predictor	Description
Depth	Elevation in metres relative to the Australian Height Datum.
Eastness	Trigonometric transformation of a circular azimuthal direction of the slope [*sin*(aspect)]. Values close to 1 represent east-facing slope, close to −1 if the aspect is westward.
Northness	Trigonometric transformation of a circular azimuthal direction of the slope [*cos*(aspect)]. Values close to 1 represent north-facing slope, close to −1 if the aspect is southward.
Slope	First derivative of elevation. Average change in elevation, steepness of the terrain, % rise.
Plan curvature	Secondary derivative of elevation. Measure of concave/convexity perpendicular to the slope.
Profile curvature	Secondary derivative of elevation. Measure of concave/convexity parallel to the slope.
Curvature	Combined index of profile (parallel to the slope) and plan (perpendicular to the slope) curvature relative to the analysis window.
Range 5, 50	Maximum minus the minimum elevation in the local neighbourhood (fine and coarse scale local relief). Calculated at radii of 5 and 50 cells.
Habitat	A spatial polygons layer with 3 categories: presence of biota/presence of biota and bioturbation/bare seafloor.

We used Receiver Operating Characteristic (ROC) and the area under the curve (AUC) as graphical means to test the sensitivity (true positive rate) and specificity (false positive rate) of a model output (Fielding and Bell, 1997). The AUC is a measure of the ability of a model to discriminate between a presence or absence observation (Elith et al., 2006). In addition, we calculated a threshold dependent Kappa statistic which is commonly used in ecological studies with presence-absence data and provides an index between 0 and 1 of how much a model predicted actual classes versus a guess (Cohen, 1960). A probability threshold that balances sensitivity and specificity was chosen as it provides a measure of how well the model predicts both presences and absences (Liu et al., 2005). After evaluation, the final models for individual taxa were predicted on 4 × 4 m grid using all available observations.

RESULTS

Oceanographic Situation

The CTD profiles revealed a steady decline in water temperature with depth across all Areas, from ∼25°C at the surface to ∼6°C at the deepest point (∼800 m) (**Figure 3A**). Salinity increased from ∼35 ppt at the surface to a maximum of ∼35.5 ppt at ∼150–200 m, and then decreased to a minimum of ∼34.6 ppt at depths greater than ∼500 m across all three surveyed Areas (**Figure 3A**).

The relationship between temperature and salinity (T-S plot; **Figure 3B**) at the deepest site (Area A) was used to evaluate the localised water masses, with reference to recent evaluations of deep-water hydrography off the Gascoyne region (Woo and Pattiaratchi, 2008). Based on the analyses of Woo and Pattiaratchi (2008), three water masses were identified. Firstly, a lower salinity tropical surface water (TSW) with a temperature range of ∼22–25°C was found in the upper ∼100 m of water column. Secondly, a higher salinity South Indian Central Water (SICW) with a temperature range of 12–22°C was centred on ∼200 m depth. Lastly, a lower salinity Antarctic Intermediate Water (AAIW) with a temperature range of ∼6–9°C was identified ∼550 m depth.

Near-bottom time series profile from Area A (depths ∼800–875 m) showed a temperature range of ∼5.7–6.3°C and relatively uniform salinity (∼34.6; **Figure 4**). At Area B, near-bottom (∼560–680 m) temperatures ranged ∼6.6–7.8°C and salinity was relatively uniform (∼34.6; **Figure 4**). The shallowest transect (∼350–560 m) in Area C showed a temperature range of ∼6.7–13°C and a salinity range of ∼34.6–35.1 (**Figure 4**). In general, the coldest, most dense water was found as the ROV traversed the deepest portions of the canyon-valley features.

General Description of Fish Assemblages

Across the three areas surveyed, the total transect length was ∼20,790 m giving a total transect area of 103,950 m^2 or 10.4 hectares. Along the entire transect, a total of 610 individual fish were recorded belonging to 80 unique taxa and 41 families. A full list of the taxa recorded by area, the total number of each taxa and their mean lengths are shown in **Supplementary Table 1.2**. Many of the taxa and individual fish recorded were small bodied (<30 cm fork length; **Supplementary Table 1.2**), however, a large 63 cm morid cod (Moridae sp3), was recorded, along with a number of larger bodied elasmobranchs. A 1.3 m whaler shark (*Carcharhinus* sp1) was the largest elasmobranch identified. The western gulper shark (*Centrophorus westraliensis*) had the largest mean fork length (86 ± 2.5 cm SE). The indigo legskate (*Sinobatis caerulea*) was the second largest taxa recorded on average at 74 ± 11.3 cm disc length (excluding the tail). One large 71 cm Chimaera *Hydrolagus lemures* (blackfin ghost shark) was also measured (**Supplementary Table 1.2**).

Approach 1, Assemblage Analysis

For total number of fish, a significant difference was found (at $\alpha = 0.05$) between Areas and between Habitats, but there was no interaction between the two factors (**Table 2**). On average, the greatest number of fish per transect (1,250 m^2) was recorded at Area B (**Figure 5**; Area B vs. Area A $t_{(18)} = 2.56$, $p = 0.003$, Area B vs. Area C $t_{(12)} = 2.20$, $p = 0.04$). The total number of fish recorded per transect were similar at Areas A and C (**Figure 5**, $t_{(14)} = 1.59$, $p = 0.134$). On average, a significantly greater number of fish were recorded in canyon-valley Habitats

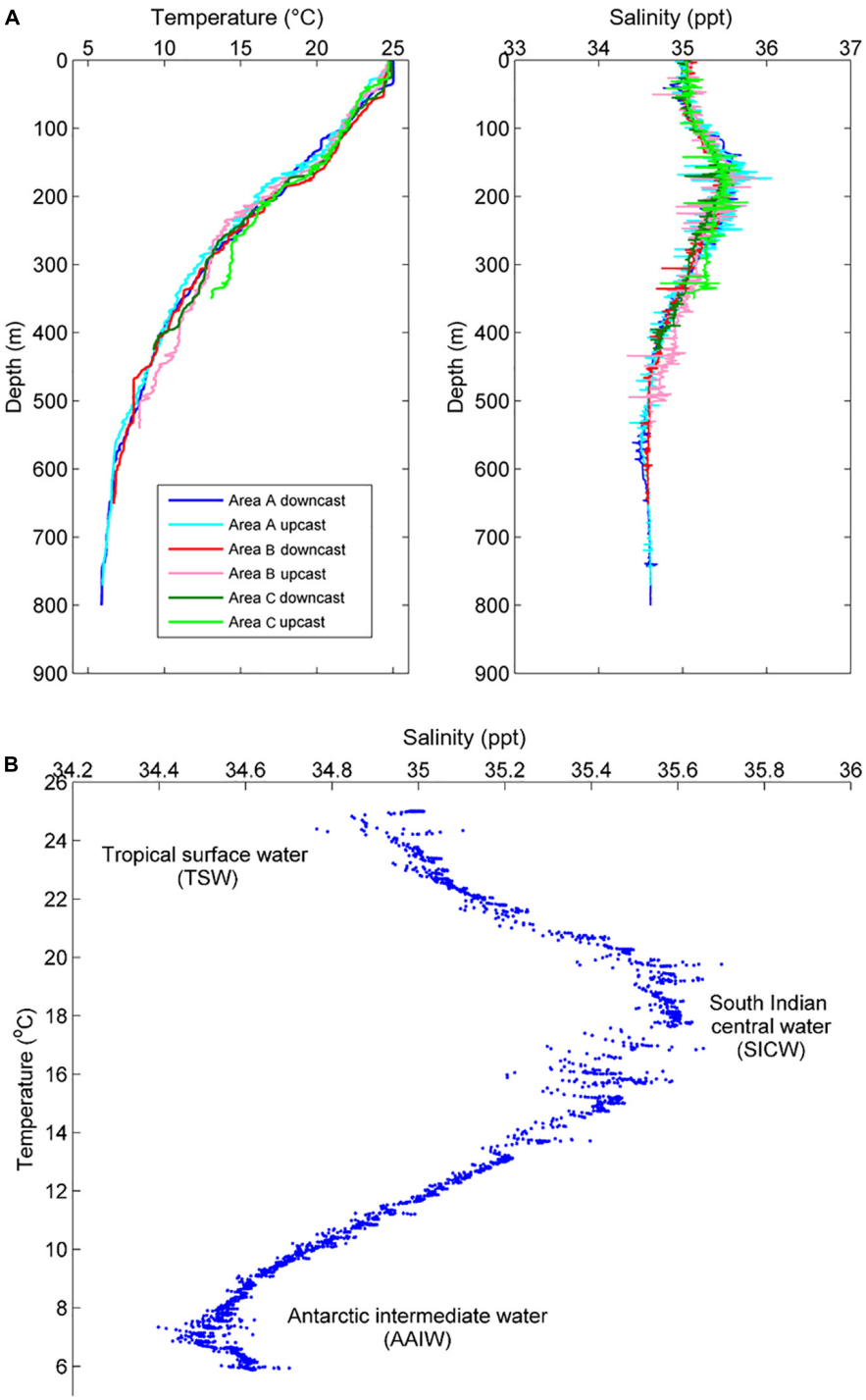

FIGURE 3 | (A) Vertical profiles of temperature and salinity across all Areas, and **(B)** temperature-salinity measured on the downcast at Area A.

compared to non-canyon Habitats (**Table 2**; **Figure 5**). Canyon-valley Habitats were also found to support a significantly greater number of taxa on average than non-canyon Habitats (**Table 2**; **Figure 5**). No significant difference was found in number of taxa between the survey Areas (**Table 2**), although the composition of the assemblages did vary.

Significant differences in the multivariate fish assemblage composition were found between both Areas and Habitats, but no interaction between these factors (**Table 2**). *Post hoc* tests on the factor Area revealed that all Areas differed to one another (all $t > 1.14$, all $p < 0.03$). The multivariate PCO showed three distinct groupings (**Figure 6**). The PCO shows transects

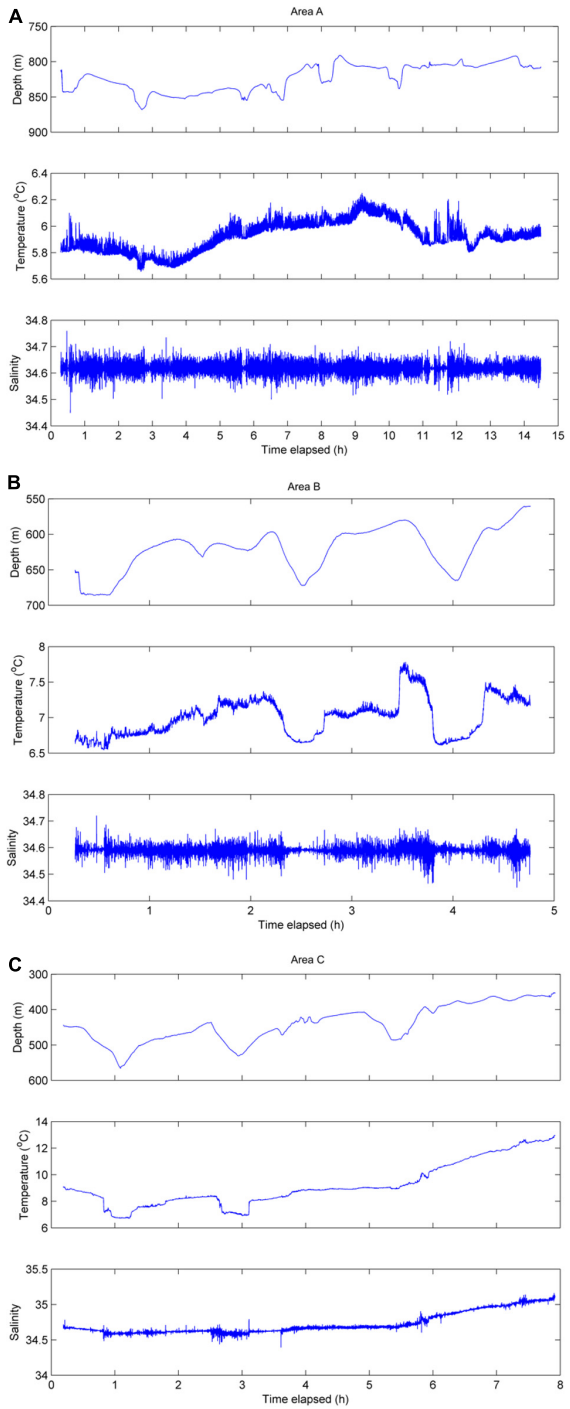

FIGURE 4 | Time series of near-bottom data collected along Area A transect **(A)**, Area B transect **(B)**, and Area C transect **(C)**.

at Area A clustered to the left of the plot were correlated to unidentified anguilliform (eel-like) fishes, and *Paraliparis* sp1 (a snail fish), as both fish were abundant in this area. In comparison, *Paraliparis* sp1 made up only 2% of the individuals recorded at Area B and were absent at Area C (**Supplementary Table 1.2**).

TABLE 2 | PERMANOVA test results for the statistical analysis of the fish assemblage composition between Areas (A, B, C) and Habitat (canyon, non-canyon).

	df	SS	MS	Pseudo-F	P(perm)	Permutations
Number of fish						
Area	2	496.90	248.45	5.482	**0.016**	9941
Habitat	1	431.56	431.56	9.522	**0.008**	9866
Area × Habitat	2	31.84	15.92	0.351	0.753	9958
Residual	22	997.07	45.32			
Total	27	2033.40				
Number of taxa						
Area	2	24.52	12.26	1.654	0.211	9951
Habitat	1	55.42	55.42	7.478	**0.012**	9810
Area × Habitat	2	16.67	8.33	1.125	0.340	9956
Residual	22	163.03	7.41			
Total	27	282.11				
Fish assemblage composition						
Area	2	21340.00	10670.00	3.618	< **0.001**	9915
Habitat	1	5793.40	5793.40	1.964	**0.023**	9932
Area × Habitat	2	6921.50	3460.70	1.173	0.250	9870
Residual	22	64888.00	2949.40			
Total	27	99509.00				

Statistically significant P values ($p < 0.05$) are highlighted in bold.

Samples to the right side of the PCO plot are arranged into two clusters (**Figure 6**). The bottom grouping is characterised by four Macrouridae (rat tail) taxa which were characteristic of Areas B and C. The grouping toward the upper right side of the plot is characterised by a mixed grouping of taxa. *Chaunax* sp1 (a coffin fish) were more abundant at both Areas B and C than at Area A. *Ariosoma* sp1 (a conger eel), *Coelorinchus* sp1 (a rat tail), and Moridae sp2 (a morid cod) were recorded only at Area B (**Supplementary Table 1.2**). One sample from Area C is grouped to the left of the plot with Area A samples. This transect contained a single fish, an unidentified anguilliform fish.

Approach 2, Species Distribution Models

Model Performance

Boosted regression trees models provided "good" model predictions (AUC = 0.80–0.89) for two taxa and "poor" predictions (AUC = 0.60–0.69; **Table 3**) for the other two taxa according to the criteria of Hosmer et al. (2013). Similar trends in the performance of fitted models were evident for all the other evaluation metrics with sensitivity, specificity and the total proportion of correct predictions being highest for the *Macrouridae* sp5 and *Paraliparis* sp1. These performance measures were further corroborated by Kappa statistics with models for Macrouridae sp2 and Macrouridae sp4 performing "fair" (Kappa = 0.21–0.40) and models developed for Macrouridae sp5 and *Paraliparis* sp1 providing "substantial" (Kappa = 0.61–0.80; **Table 3**) predictions (Landis and Koch, 1977).

Habitat Associations

Depth had the greatest influence on the probability of occurrence of all taxa (**Figure 7**). For Macrouridae sp2, Macrouridae sp5 and

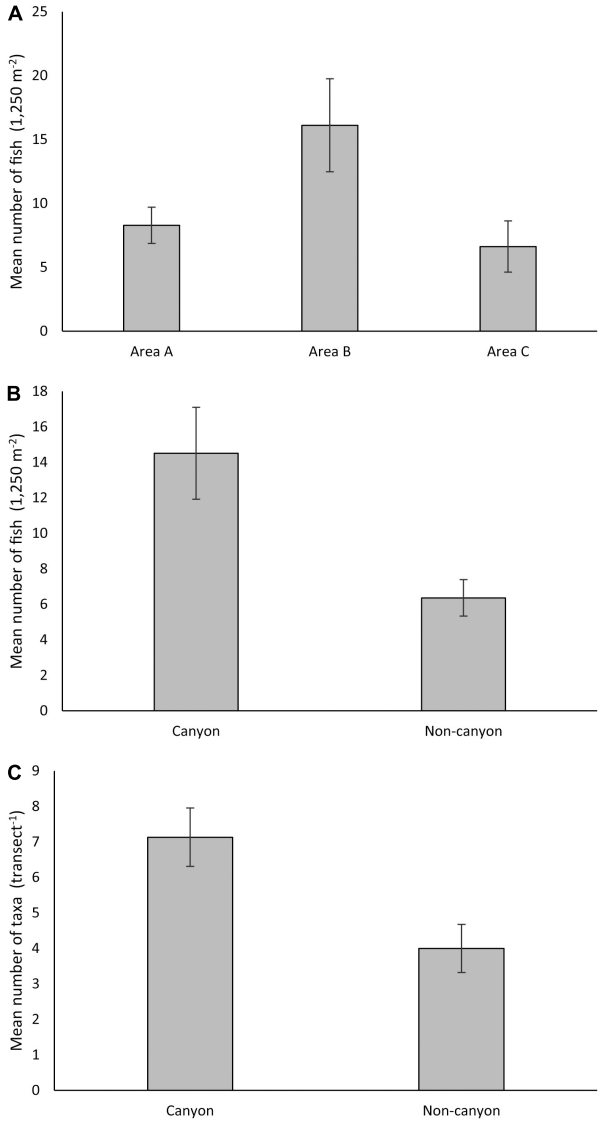

FIGURE 5 | The mean ± SE of **(A)** number of fish recorded (1,250 m^{-2}) at each Area, **(B)** the number of fish recorded (1,250 m^{-2}) in each Habitat, and **(C)** number of taxa recorded (transect^{-1}) in each Habitat.

Paraliparis sp1, higher probability of occurrence was associated with water depth over 600 m. In contrast, Macrouridae sp4 was predicted to be most likely to occur at depths between 500 and 600 m (**Figure 7**). Various indices of topographic complexity of the relief (plan, profile, curvature and range 50) were also important in influencing the probability of occurrence of Macrouridae sp2, Macrouridae sp4 and Macrouridae sp5. Higher probability of occurrence of Macrouridae sp2 and Macrouridae sp4 was predicted on the seabed surrounding the canyon-valley slopes, whereas Macrouridae sp5 was predicted to be deeper and associated with both flat areas of low structural complexity, and within the feature on canyon-valley slopes and the canyon floor (**Figures 7**, **8**). A higher probability of occurrence of *Paraliparis* sp1 was predicted on deep non-canyon habitats (**Figures 7**, **8**).

Spatial Patterns of Distributions

A high probability of occurrence of *Macrouridae* sp2, *Paraliparis* sp1 and *Macrouridae* sp5 was predicted in the deepest area A (**Figure 8**; **Supplementary Material 2, 3, 5**). A high probability of occurrence of *Macrouridae* sp5 was predicted in Area A, and also over the deeper areas (>500 m) and within the canyon-valley of Area B. In contrast, higher probability of occurrence of *Macrouridae* sp2 and *Macrouridae* sp4 in area B was predicted for regions outside the canyon-valley and on its slopes, respectively (**Figure 8**). *Macrouridae* sp4 was also predicted to occur within the deeper canyon-valley of Area C and the shallower regions of Area A, reflecting its predicted relationship with depth (**Figure 8**; **Supplementary Material 4**).

DISCUSSION

Deep-water environments and subsea canyons support diverse fish assemblages which are different to nearby shallow water assemblages (Williams et al., 2001; Fernandez-Arcaya et al., 2017). Here, we combined novel survey methods and advanced analytical approaches to document patterns of deep-water fish assemblage composition and spatial distributions. We also showed that subsea canyon environments support fish assemblages characterised by a higher number of taxa and overall number of individuals when compared to the areas outside the canyon, which likely reflects the heterogeneity of habitat and oceanographic conditions within the canyon (Klinck, 1996; McClain and Barry, 2010). A combination of non-extractive survey methods and SDMs provided a useful tool for improving our understanding of fine-scale spatial distribution patterns of deep-water fishes and the associated environmental drivers.

The survey area had a high diversity of taxa for a deep-sea environment with 80 taxa of fish from 41 families recorded, which is consistent with previous observations of fish diversity in this region (Williams et al., 1996, 2001). The family Macrouridae (rat tails) dominated the fish assemblage across the survey area, which is also supported by earlier research (Williams et al., 1996, 2001). We recorded 11 Macrouridae taxa whereas Williams et al. (1996) reported 16 species of Macrouridae using trawls within the same latitudinal and depth range (21–22°S, 300–900 m depth). There was a consistent pattern between areas in that the canyon-valley features supported a greater number of fish and a higher diversity of taxa than non-canyon habitat. This is likely due to variation in the habitat structure of the canyon-valley feature (McClain and Barry, 2010). At Areas A and B, the canyon-valley habitat was wide (250–500 m) and topographically shallow (20–50 m) in comparison to Area C which was a more typical canyon feature. Submarine canyon habitats can support higher abundances of benthic feeding fishes than nearby continental shelf slopes (De Leo et al., 2010). Greater prey availability in sediment infauna and epifauna may support the increased abundance of fishes, particularly Macrouridae, which were characteristic of canyon-valley habitats (De Leo et al., 2010).

A different fauna characterised the deeper Area A compared to Areas B and C. This was likely to be due to the influence of

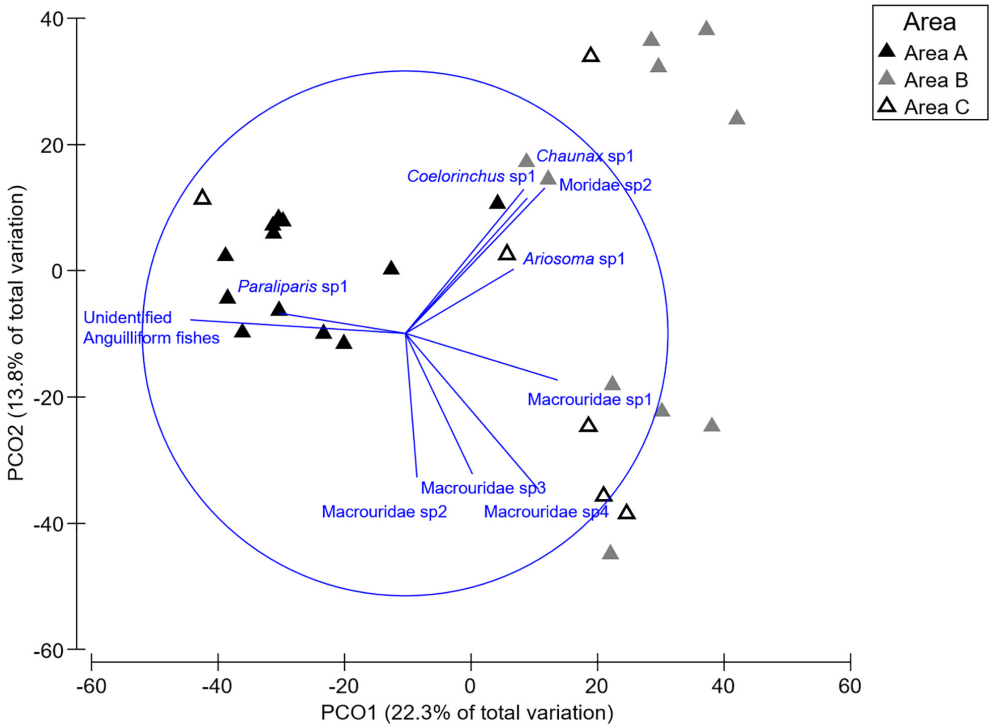

FIGURE 6 | Principle component ordination (PCO) illustrating the variation between transects by Area. The overlaid vectors illustrate the strength and the direction of the correlation to the PCO axes of influential fish species.

TABLE 3 | Summary of model predictive performance for each fish species.

Model	Threshold	AUC	Omission rate	Sensitivity	Specificity	Proportion correct	Kappa
Macrouridae sp2	0.46	0.66	0.38	0.62	0.7	0.65	0.28
Macrouridae sp4	0.57	0.69	0.33	0.67	0.67	0.67	0.32
Macrouridae sp2	0.49	0.84	0.17	0.83	0.8	0.82	0.63
Paraliparis sp1	0.56	0.89	0.1	0.89	0.87	0.88	0.74

Presences and absences for assessing sensitivity and specificity were determined using a threshold that balances sensitivity and specificity.

the AAIW deep water current at area A. Submarine canyons are important in linking continental slope and shelf areas and water masses. Canyon features change nutrient dynamics through increased upwelling and interactions with alongshore currents (Klinck, 1996), and are important for supplying nutrients to deep ocean through down-canyon transport (Fernandez-Arcaya et al., 2017). Area A was differentiated from Areas B and C by a higher abundance of the species group of unidentified anguilliform fishes and also a higher abundance of *Paraliparis* sp1 (a snail fish). Species from the genus *Paraliparis* generally inhabit the deep sea, and have been recorded in this region at depths of 1,030 m (Williams et al., 1996), and 2,821 m in New Zealand (Roberts et al., 2015). A similar pattern of a faunal break between fish assemblages on the upper and the mid-slope was reported by Last et al. (2011), which can be attributed to the influence of the AAIW (Williams et al., 2001). This deep-water current has a significant influence on the distribution of fish communities in the south-west region between the Great Australian Bight and north of the Ningaloo Reef where its core is known to fluctuate with depth from approximately 875 m at 27.50°S to 520 m around 21.50°S (Williams et al., 1996, 2001; Woo and Pattiaratchi, 2008). A full assessment of the influence of this current on fish assemblages within our study area would require further specifically targetted research.

Our study contributes additional information to understanding fish communities and their relationship to benthic habitats in two KEFs for the Australian northwest marine region; Canyons linking the Cuvier Abyssal Plain and the Cape Range Peninsula and Continental Slope Demersal Fish Communities (DSEWPaC, 2012). Deep-sea KEFs are often data-poor and their definition is often based on limited data (Falkner et al., 2009). The Canyons linking the Cuvier Abyssal Plain and the Cape Range Peninsula are representative features of the region but are not unique in a wider Australian context (Falkner et al., 2009). In terms of the Continental Slope Demersal Fish Communities KEF, Last et al. (2005) recorded 500 species making it the most biodiverse slope region in Australia. In

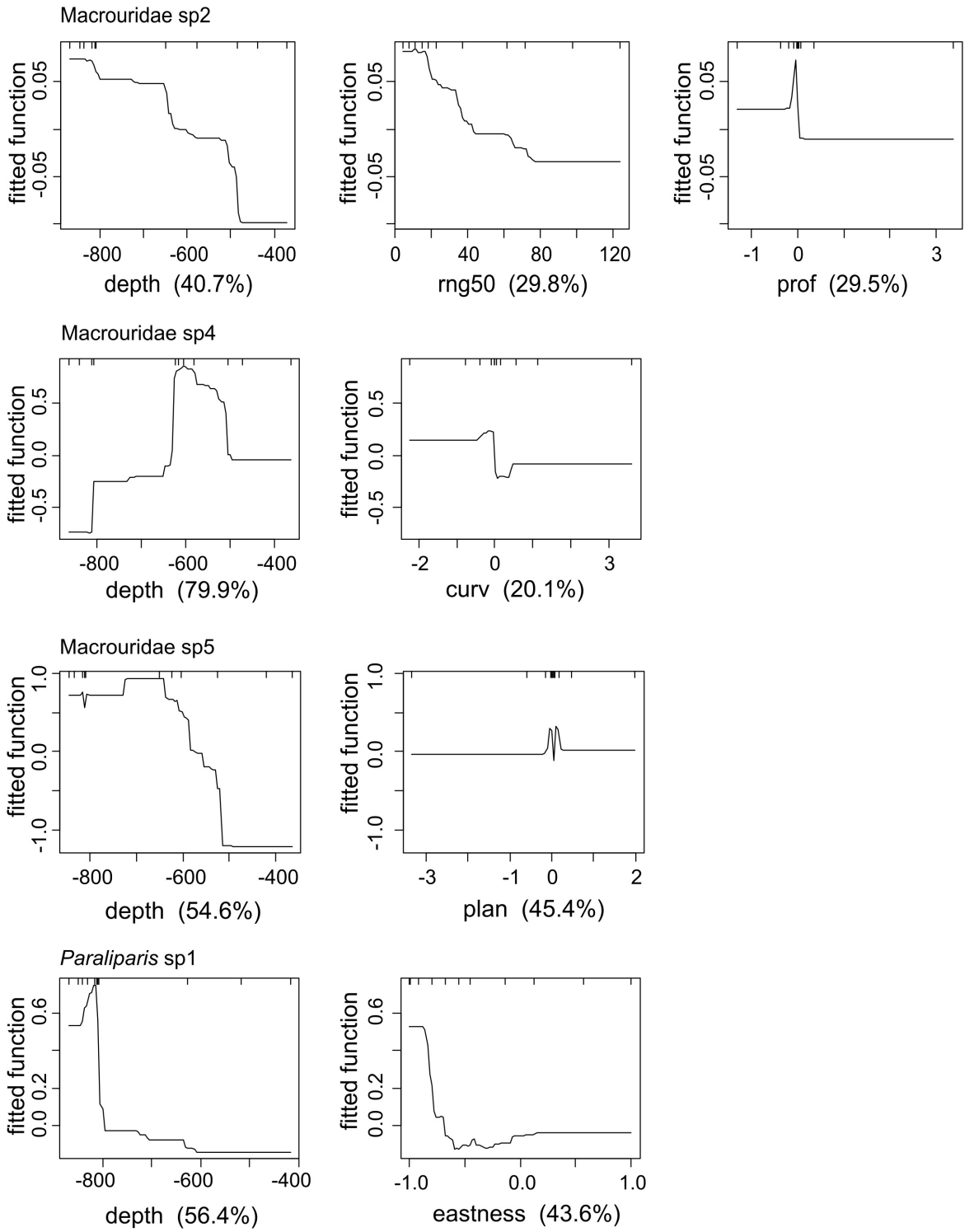

FIGURE 7 | Partial dependence plots for Boosted Regression Tree (BRT) analyses relating species occurrence to the environmental predictors. The relative importance of each variable is shown in parentheses on the *x*-axis.

comparison, we recorded 80 taxa, but our study focus was on a tributary to a canyon system, and not the entire continental slope assessed by Last et al. (2005).

We elucidated the distribution patterns of the four most abundant taxa within the assemblages using SDMs. SDMs are a useful tool for exploring habitat associations of the deep-sea fishes

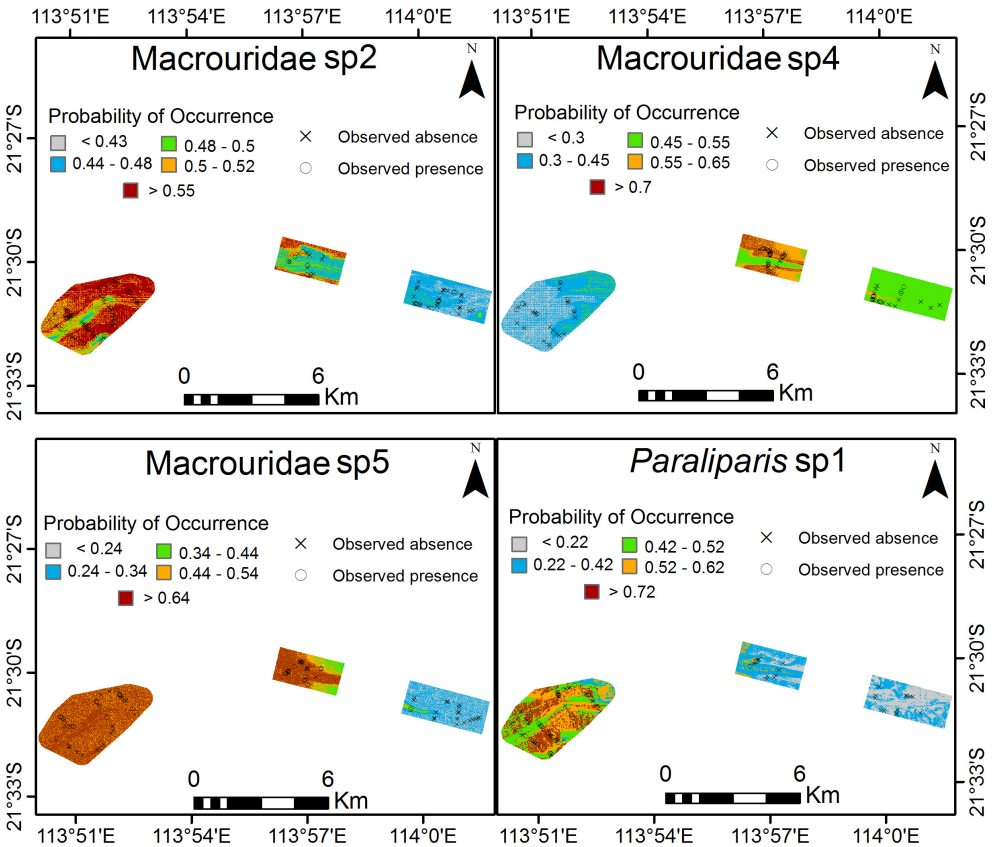

FIGURE 8 | Predicted probability of occurrence for each species across 3 Areas as a result of Boosted Regression Trees analysis. Higher probability of occurrence is in red. For high resolution images please see **Supplementary Materials 2** through **5**.

because they provide insights into the ecology of these rarely observed taxa and data can be extrapolated into unsurveyed areas. Often depth and habitat complexity drive environmental niche associations of demersal fishes (Monk et al., 2010; Moore et al., 2013; Galaiduk et al., 2018). This was observed in our study, where depth was the most important environmental predictor for all modelled taxa. Our models identified *Paraliparis* sp1 to be primarily associated with water depth over 800 m. In contrast, the three other taxa were predicted to occur at water depth greater than 500 m with Macrouridae sp4 predicted to mainly occupy depths between 500 and 600 m. Habitat complexity calculated at different scales was also an important predictor of habitat associations of the modelled taxa, suggesting that these taxa use various sized nooks and crevices (as estimated from the bathymetry derivatives) for shelter and protection from predation (Kelley et al., 2006). Macrouridae sp2 and Macrouridae sp4 were predicted to occupy the canyon slopes, whereas Macrouridae sp5 was predicted to be more associated with areas within the canyon and the canyon floor, suggesting within-family partitioning of the available environmental niche space to avoid resource overlap (Ross, 1986).

Biological data collection in deep-sea environments is expensive and labour-intensive. Large vessels capable of deploying multiple data collection platforms (e.g., ROVs, CTD probes, sediment corers, and camera systems) through the water column to the seabed, often to depths of 1,000's of metres, are required (McLean et al., 2020). In this study, we streamlined the collection of multiple data types simultaneously (stereo-video imagery of fishes, benthic habitat imagery, and CTD) by deploying multiple instruments on a single ROV platform. This efficient configuration is recommended for future environmental studies of both shallow and deep seas as it minimises the time required to collect data. The methods are particularly applicable to baseline studies that support environmental impact assessments for offshore facilities (e.g., oil and gas) and inform the planning and management of marine reserves.

Remotely Operated Vehicle operated surveys are an industry-supported method with potential for large data collection in remote marine environments rarely accessible to researchers (McLean et al., 2017, 2018a; Macreadie et al., 2018). ROVs present a platform which can be fitted with multiple tools to collect a variety of environmental data from challenging remote environments (McLean et al., 2017, 2018a; Macreadie et al., 2018). Previously, sampling has relied on extractive trawls, or more recently deep-water baited cameras (Priede and Bagley, 2000; Jamieson et al., 2009; Marouchos et al., 2011; Wellington et al., 2021). These techniques are better suited to describing broader-scale spatial patterns (100 s of m) than the fine-scale

sampling (<10 s of m) using the transect and SDM approach described in this study. The ROV transect approach allows for the *in situ* observation of fishes (Macreadie et al., 2018), and the investigation of fine-scale fish-habitat relationships through SDMs. It also eliminates the bias associated with the use of baited video systems (i.e., the bait plume effect; see Galaiduk et al. (2017b) and Monk et al. (2012) for a comparison between methods and applications in SDMs).

The ROV, however, has its own biases associated with noise, and the bright lighting required to collect quality video footage in dark environments which may affect the behaviour of mobile fauna (Ryer et al., 2009). We did not observe adverse behavioural reactions or fleeing, however, it is possible that certain taxa may simply avoid the ROV entirely and were thus not recorded (Linley et al., 2013). It is likely that avoidance by mobile fauna is greatest over soft sediment habitats such as those sampled during this study (Schramm et al., 2020a). These artefacts and avoidance may have implications for SDMs, however, they were consistent throughout the survey and should not affect statistical comparisons between Habitats or Areas. Traditional extractive methods for sampling deep-sea environments have biases and limitations relating to the gear and the spatial resolution of samples obtained. Previous studies from the region used a large netted paired-warp trawl rig (Williams et al., 1996, 2001) that may have under-sampled small species (Williams et al., 2001). Smaller meshed nets used in this region have caught a different suite of fishes, including more Ophidiidae (cusk eels) and Congridae (conger eels) (Williams et al., 2001). Many of the fishes measured during our ROV survey were between 10 and 30 cm in length, and so may have been under represented in previous trawl surveys. Similarly, in our stereo-video ROV survey the positive identification of many small bodied (<10 cm) individuals was difficult because morphological features were indistinct. Therefore, both stereo-video ROV and trawl methods may under sample the diversity of small bodied fishes.

CONCLUSION

The submarine canyon-valley feature sampled here supports a characteristically rich fish assemblage with a greater number of taxa and individuals than areas outside the feature. This canyon-valley feature is a tributary to the larger Cape Range canyon system which may play an ecological role in linking the Cuvier Abyssal Plain and the Cape Range Peninsula (Huang et al., 2018). Here, we present a novel application of ROV stereo-video transects in deep water, demonstrating the utility of this technique for facilitating fine-scale fish species distribution modelling that has the potential to feed into spatial management of deeper offshore waters.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data are commercial in confidence. Requests to access the datasets should be directed to environment@bmtglobal.com.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because at the time that the study was conducted ethical review and approval was not required for observational studies.

AUTHOR CONTRIBUTIONS

BS, EH, LT, MW, KI, and DM conceived and designed the study. BS, KI, and LT collected the data. BS, RG, and KI analysed the data. BS, RG, EM, JG, and KI developed the initial version of the manuscript. All authors contributed to interpreting the data and developing the full manuscript.

FUNDING

This research was supported by funding and data from the Woodside-operated Greater Enfield Project, a Joint Venture between Woodside Energy Ltd. and Mitsui E&P Australia Pty Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.608665/full#supplementary-material

REFERENCES

Adams, P. B., Butler, J. L., Baxter, C. H., Laidig, T. E., Dahlin, K. A., and Waldo Wakefield, W. (1995). Population estimates of Pacific coast groundfishes from video transects and swept-area trawls. *Fish. Bull.* 93, 446–455.

Agapova, G. V., Budanova, L. Y., Zenkevich, N. L., Larina, N. I., Litvin, V. M., Marova, N. A., et al. (1979). *Geomorphology of the ocean floor, Geofizika okeana*. Moscow: Geofizika okeanskogo dna, Neprochnov, Izd.

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253. doi: 10.1111/j.1541-0420.2005.00440.x

Anderson, M. J., Gorley, R. N., and Clarke, K. R. (2008). *PERMANOVA+ for PRIMER: Guide for Software and Statistical Methods*. Plymouth: University of Auckland and Plymouth Marine Laboratory.

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., et al. (2012). The magnitude of global marine species diversity. *Curr. Biol.* 22, 2189–2202.

Armstrong, C. W., Foley, N. S., Tinch, R., and van den Hove, S. (2012). Services from the deep: Steps towards valuation of deep sea goods and services. *Ecosyst. Serv.* 2, 2–13. doi: 10.1016/j.ecoser.2012.07.001

Bo, M., Bava, S., Canese, S., Angiolillo, M., Cattaneo-Vietti, R., and Bavestrello, G. (2014). Fishing impact on deep Mediterranean rocky habitats as revealed by ROV investigation. *Biol. Conserv.* 171, 167–176. doi: 10.1016/j.biocon.2014.01.011

Bond, T., Partridge, J. C., Taylor, M. D., Cooper, T. F., and McLean, D. L. (2018). The influence of depth and a subsea pipeline on fish assemblages and commercially fished species. *PLoS One* 13:e0207703.

Boutros, N., Shortis, M. R., and Harvey, E. S. (2015). A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. *Limnol. Oceanogr. Methods* 13, 224–236. doi: 10.1002/lom3.10020

Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. (2002). Evaluating resource selection functions. *Ecolog. Model.* 157, 281–300. doi: 10.1016/s0304-3800(02)00200-4

Cappo, M., Speare, P., and De'ath, G. (2004). Comparison of baited remote underwater video stations (BRUVS) and prawn (shrimp) trawls for assessments of fish biodiversity in inter-reefal areas of the Great Barrier Reef Marine Park. *J. Exp. Mar. Bio. Ecol.* 302, 123–152. doi: 10.1016/j.jembe.2003.10.006

Clark, M. (2001). Are deepwater fisheries sustainable?—the example of orange roughy (Hoplostethus atlanticus) in New Zealand. *Fish. Res.* 51, 123–135. doi: 10.1016/s0165-7836(01)00240-5

Clarke, K. R., Chapman, M. G., Somerfield, P. J., and Needham, H. R. (2006a). Dispersion-based weighting of species counts in assemblage analyses. *Mar. Ecol. Prog. Ser.* 320, 11–27. doi: 10.3354/meps320011

Clarke, K. R., and Gorley, R. N. (2006). *Primer*. Plymouth: PRIMER-e.

Clarke, K. R., Somerfield, P. J., and Chapman, M. G. (2006b). On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J. Exp. Mar. Biol. Ecol.* 330, 55–80. doi: 10.1016/j.jembe.2005.12.017

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cruz-Acevedo, E., Tolimieri, N., and Aguirre-Villaseñor, H. (2018). Deep-sea fish assemblages (300-2100 m) in the eastern Pacific off northern Mexico. *Mar. Ecol. Prog. Ser.* 592, 225–242. doi: 10.3354/meps12502

Danovaro, R., Snelgrove, P. V. R., and Tyler, P. (2014). Challenging the paradigms of deep-sea ecology. *Trends Ecol. Evol.* 29, 465–475. doi: 10.1016/j.tree.2014.06.002

De Leo, F. C., Smith, C. R., Rowden, A. A., Bowden, D. A., and Clark, M. R. (2010). Submarine canyons: hotspots of benthic biomass and productivity in the deep sea. *P. Roy. Soc. B-Biol. Sci.* 277, 2783–2792. doi: 10.1098/rspb.2010.0462

DSEWPaC. (2012). *Marine bioregional plan for the North-West Marine Region prepared under the Environment Protection and Biodiversity Conservation Act 1999*. Canberra: Australian Government Department of Sustainability Environment Water Populations and Communities.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.

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

Falkner, I., Whiteway, T., Przeslawski, R., and Heap, A. D. (2009). *Review of Ten Key Ecological Features (KEFs) in the North-west Marine Region. Geoscience Australia, Record 2009/13*. Canberra: Geoscience Australia, 117.

Fernandez-Arcaya, U., Ramirez-Llodra, E., Aguzzi, J., Allcock, A. L., Davies, J. S., Dissanayake, A., et al. (2017). Ecological Role of Submarine Canyons and Need for Canyon Conservation: A Review. *Front. Mar. Sci.* 4:5.

Fielding, A. H., and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 1997, 38–49. doi: 10.1017/s0376892997000088

Franklin, J. (2010). *Mapping species distributions: spatial inference and prediction*. Cambridge, MA: Cambridge University Press.

Galaiduk, R., Radford, B. T., and Harvey, E. S. (2018). Utilizing individual fish biomass and relative abundance models to map environmental niche associations of adult and juvenile targeted fishes. *Sci. Rep.* 8:9457.

Galaiduk, R., Radford, B. T., Saunders, B. J., Newman, S. J., and Harvey, E. S. (2017a). Characterizing ontogenetic habitat shifts in marine fishes: advancing nascent methods for marine spatial management. *Ecol. Appl.* 27, 1776–1788. doi: 10.1002/eap.1565

Galaiduk, R., Radford, B. T., Wilson, S. K., and Harvey, E. S. (2017b). Comparing two remote video survey methods for spatial predictions of the distribution and environmental niche suitability of demersal fishes. *Sci. Rep.* 7:17633.

Gates, A. R., Morris, K. J., Jones, D. O. B., and Sulak, K. J. (2017). An association between a cusk eel (Bassozetus sp.) and a black coral (Schizopathes sp.) in the deep western Indian Ocean. *Mar. Biodivers.* 47, 971–977. doi: 10.1007/s12526-016-0516-z

Goetze, J. S., Bond, T., McLean, D. L., Saunders, B. J., Langlois, T. J., Lindfield, S., et al. (2019). A field and video analysis guide for diver operated stereo-video. *Methods Ecol. Evol.* 10, 1083–1090. doi: 10.1111/2041-210x.13189

Harvey, E., Fletcher, D., and Shortis, M. (2001). A comparison of the precision and accuracy of estimates of reef-fish lengths determined visually by divers with estimates produced by a stereo-video system. *Fish. Bull.* 99, 63–71.

Harvey, E., and Shortis, M. (1995). A system for stereo-video measurement of sub-tidal organisms. *Mar. Technol. Soc. J.* 1995:6318.

Harvey, E. S., Fletcher, D., Shortis, M., and Kendrick, G. (2004). A comparison of underwater visual distance estimates made by scuba divers and a stereo-video system: implications for underwater visual census of reef fish abundance. *Mar. Freshwater Res.* 55, 573–580. doi: 10.1071/mf03130

Harvey, E. S., Goetze, J. S., McLaren, B., Langlois, T., and Shortis, M. R. (2010). Influence of range, angle of view, image resolution and image compression on underwater stereo-video measurements: high-definition and broadcast-resolution video cameras compared. *Mar. Technol. Soc. J.* 44, 75–85. doi: 10.4031/mtsj.44.1.3

Heap, A. D., and Harris, P. T. (2008). Geomorphology of the Australian margin and adjacent seafloor. *Aust. J. Earth Sci.* 55, 555–585. doi: 10.1080/08120090801888669

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

Hosmer, J. D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons, Inc., 528.

Huang, Z., Nichol, S. L., Harris, P. T., and Caley, M. J. (2014). Classification of submarine canyons of the Australian continental margin. *Mar. Geol.* 357, 362–383. doi: 10.1016/j.margeo.2014.07.007

Huang, Z., Schlacher, T. A., Nichol, S., Williams, A., Althaus, F., and Kloser, R. (2018). A conceptual surrogacy framework to evaluate the habitat potential of submarine canyons. *Prog. Oceanogr.* 169, 199–213. doi: 10.1016/j.pocean.2017.11.007

Hudson, I. R., Jones, D. O. B., and Wigham, D. B. (2005). A review of the uses of work-class ROVs for the benefits of science: Lessons learned from the SERPENT project. *Underwat. Technol.* 26, 83–88. doi: 10.3723/175605405784426637

Jamieson, A. J., Fujii, T., Solan, M., and Priede, I. G. (2009). HADEEP: Free-Falling Landers to the Deepest Places on Earth. *Mar. Technol. Soc. J.* 43, 151–160. doi: 10.4031/mtsj.43.5.17

Jobstvogt, N., Hanley, N., Hynes, S., Kenter, J., and Witte, U. (2014). Twenty thousand sterling under the sea: Estimating the value of protecting deep-sea biodiversity. *Ecol. Econ.* 97, 10–19. doi: 10.1016/j.ecolecon.2013.10.019

Jones, D. O. B. (2009). Using existing industrial remotely operated vehicles for deep-sea science. *Zool. Scr.* 38, 41–47.

Kelley, C., Moffitt, R. B., and Smith, J. R. (2006). *Mega-to micro-scale classification and description of bottomfish essential fish habitat on four banks in the Northwestern Hawaiian Islands*. Available online at: https://repository.si.edu/bitstream/handle/10088/33874/Atoll_2006_18.pdf (accessed August 27, 2018).

Klinck, J. M. (1996). Circulation near submarine canyons: A modeling study. *J. Geophys. Res.* 101, 1211–1223. doi: 10.1029/95jc02901

Landis, J. R., and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977, 363–374. doi: 10.2307/2529786

Last, P. R., Lyne, V. D., Yearsley, G. K., Gledhill, D. C., Gomon, M. F., Rees, A. J. J., et al. (2005). *Validation of national demersal fish datasets for the regionalisation of the Australian continental slope and outer shelf (>40m depth)*. Hobart: National Oceans Office.

Last, P. R., White, W. T., Gledhill, D. C., Pogonoski, J. J., Lyne, V., and Bax, N. J. (2011). Biogeographical structure and affinities of the marine demersal ichthyofauna of Australia. *J. Biogeogr.* 38, 1484–1496. doi: 10.1111/j.1365-2699.2011.02484.x

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., and Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* 321, 267–281. doi: 10.3354/meps321267

Levin, L. A., Etter, R. J., Rex, M. A., Gooday, A. J., Smith, C. R., Pineda, J., et al. (2001). Environmental Influences on Regional Deep-Sea Species Diversity.

Annu. Rev. Ecol. Syst. 32, 51–93. doi: 10.1146/annurev.ecolsys.32.081501.114002

Linley, T. D., Alt, C. H. S., Jones, D. O. B., and Priede, I. G. (2013). Bathyal demersal fishes of the Charlie-Gibbs Fracture Zone region (49°--54° N) of the Mid-Atlantic Ridge: III. Results from remotely operated vehicle (ROV) video transects. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* 98, 407–411. doi: 10.1016/j.dsr2.2013.08.013

Liu, C., Berry, P. M., Dawson, T. P., and Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28, 385–393. doi: 10.1111/j.0906-7590.2005.03957.x

Lorance, P., Souissi, S., and Uiblein, F. (2002). Point, alpha and beta diversity of carnivorous fish along a depth gradient. *Aquat. Living Resour.* 15, 263–271. doi: 10.1016/s0990-7440(02)01189-0

Lorance, P., and Trenkel, V. M. (2006). Variability in natural behaviour, and observed reactions to an ROV, by mid-slope fish species. *J. Exp. Mar. Bio. Ecol.* 332, 106–119. doi: 10.1016/j.jembe.2005.11.007

Macreadie, P. I., McLean, D. L., Thomson, P. G., Partridge, J. C., Jones, D. O. B., Gates, A. R., et al. (2018). Eyes in the sea: Unlocking the mysteries of the ocean using industrial, remotely operated vehicles (ROVs). *Sci. Total Environ.* 634, 1077–1091. doi: 10.1016/j.scitotenv.2018.04.049

Marouchos, A., Sherlock, M., Barker, B., and Williams, A. (2011). "Development of a stereo deepwater Baited Remote Underwater Video System (DeepBRUVS),". *OCEANS 2011 IEEE - Spain*. New York, NY: IEEE 2011, 1–5.

McClain, C. R., and Barry, J. P. (2010). Habitat heterogeneity, disturbance, and productivity work in concert to regulate biodiversity in deep submarine canyons. *Ecology* 91, 964–976. doi: 10.1890/09-0087.1

McClatchie, S., Millar, R. B., Webster, F., Lester, P. J., Hurst, R., and Bagley, N. (1997). Demersal fish community diversity off New Zealand: Is it related to depth, latitude and regional surface phytoplankton? *Deep Sea Res. Part I* 44, 647–667. doi: 10.1016/s0967-0637(96)00096-9

McLean, D. L., Green, M., Harvey, E. S., Williams, A., Daley, R., and Graham, K. J. (2015). Comparison of baited longlines and baited underwater cameras for assessing the composition of continental slope deepwater fish assemblages off southeast Australia. *Deep Sea Res. Part I* 98, 10–20. doi: 10.1016/j.dsr.2014.11.013

McLean, D. L., Macreadie, P., and White, D. J. (2018a). Understanding the Global Scientific Value of Industry ROV Data, to Quantify Marine Ecology and Guide Offshore Decommissioning Strategies. *Offshore Technol.* 2018:28312.

McLean, D. L., Parsons, M. J. G., Gates, A. R., Benfield, M. C., Bond, T., Booth, D. J., et al. (2020). Enhancing the Scientific Value of Industry Remotely Operated Vehicles (ROVs) in Our Oceans. *Front. Mar. Sci.* 2020:7.

McLean, D. L., Partridge, J. C., Bond, T., Birt, M. J., Bornt, K. R., and Langlois, T. J. (2017). Using industry ROV videos to assess fish associations with subsea pipelines. *Cont. Shelf Res.* 141, 76–97.

McLean, D. L., Taylor, Partridge, J. C., Gibbons, B., Langlois, T. J., Malseed, B. E., et al. (2018b). Fish and habitats on wellhead infrastructure on the north west shelf of Western Australia. *Cont. Shelf Res.* 164, 10–27. doi: 10.1016/j.csr.2018.05.007

Mindel, B. L., Neat, F. C., Trueman, C. N., Webb, T. J., and Blanchard, J. L. (2016). Functional, size and taxonomic diversity of fish along a depth gradient in the deep sea. *PeerJ* 4:e2387. doi: 10.7717/peerj.2387

Monk, J., Ierodiaconou, D., Harvey, E., Rattray, A., and Versace, V. L. (2012). Are we predicting the actual or apparent distribution of temperate marine fishes? *PLoS One* 7:e34558. doi: 10.1371/journal.pone.0034558

Monk, J., Ierodiaconou, D., Versace, V. L., Bellgrove, A., Harvey, E., Rattray, A., et al. (2010). Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. *Mar. Ecol. Prog. Ser.* 420, 157–174. doi: 10.3354/meps08858

Moore, C. H., Drazen, J. C., Kelley, C. D., and Misa, W. F. X. E. (2013). Deepwater marine protected areas of the main Hawaiian Islands: establishing baselines for commercially valuable bottomfish populations. *Mar. Ecol. Prog. Ser.* 476, 167–183. doi: 10.3354/meps10132

Moore, C. H., Van Niel, K., and Harvey, E. S. (2011). The effect of landscape composition and configuration on the spatial distribution of temperate demersal fish. *Ecography* 34, 425–435. doi: 10.1111/j.1600-0587.2010.06436.x

Moranta, J., Stefanescu, C., Massutí, E., Morales-Nin, B., and Lloris, D. (1998). Fish community structure and depth-related trends on the continental slope of the Balearic Islands (Algerian basin, western Mediterranean). *Mar. Ecol. Prog. Ser.* 171, 247–259. doi: 10.3354/meps171247

Newman, S. J., Williams, D. M., and Russ, G. R. (1997). Patterns of zonation of assemblages of the Lutjanidae, Lethrinidae and Serranidae (Epinephelinae) within and among mid-shelf and outer-shelf reefs in the central Great Barrier Reef. *Mar. Freshwater Res.* 48, 119–128. doi: 10.1071/mf96047

Oyafuso, Z. S., Drazen, J. C., Moore, C. H., and Franklin, E. C. (2017). Habitat-based species distribution modelling of the Hawaiian deepwater snapper-grouper complex. *Fish. Res.* 195, 19–27. doi: 10.1016/j.fishres.2017.06.011

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., et al. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. doi: 10.1890/07-2153.1

Pittman, S. J., Christensen, J. D., Caldow, C., Menza, C., and Monaco, M. E. (2007). Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecol. Model.* 204, 9–21. doi: 10.1016/j.ecolmodel.2006.12.017

Priede, I. G., and Bagley, P. M. (2000). In situ studies on deep-sea demersal fishes using autonomous unmanned lander platforms. *Oceanogr. Mar. Biol.* 38, 357–392.

Puig, P., Canals, M., Company, J. B., Martín, J., Amblas, D., Lastras, G., et al. (2012). Ploughing the deep sea floor. *Nature* 489, 286–289. doi: 10.1038/nature11410

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Roberts, C. D., Stewart, A. L., and Struthers, C. D. (eds) (2015). *The fishes of New Zealand*, Vol. 4. New Zealand: Te Papa Press.

Rogers, A. D. (2015). Environmental Change in the Deep Ocean. *Ann. Rev. Env. Resour.* 2015:21415. doi: 10.1146/annurev-environ-102014-021415

Ross, S. T. (1986). Resource Partitioning in Fish Assemblages: A Review of Field Studies. *Copeia* 1986, 352–388. doi: 10.2307/1444996

Ryer, C. H., Stoner, A. W., Iseri, P. J., and Spencer, M. L. (2009). Effects of simulated underwater vehicle lighting on fish behavior. *Mar. Ecol. Prog. Ser.* 391, 97–106. doi: 10.3354/meps08168

Schramm, K. D., Harvey, E. S., Goetze, J. S., Travers, M. J., Warnock, B., and Saunders, B. J. (2020b). A comparison of stereo-BRUV, diver operated and remote stereo-video transects for assessing reef fish assemblages. *J. Exp. Mar. Bio. Eco.* 524, 151273. doi: 10.1016/j.jembe.2019.151273

Schramm, K. D., Marnane, M. J., Elsdon, T. S., Jones, C., Saunders, B. J., Goetze, J. S., et al. (2020a). A comparison of stereo-BRUVs and stereo-ROV techniques for sampling shallow water fish communities on and off pipelines. *Mar. Env. Res.* 162:105198. doi: 10.1016/j.marenvres.2020.105198

Schramm, K. D., Marnane, M. J., Elsdon, T. S., Jones, C. M., Saunders, B. J., Newman, S. J., et al. (2021). Fish associations with shallow water subsea pipelines compared to surrounding reef and soft sediment habitats. *Sci. Rep.* 11:6238.

SeaGIS (2014). *TransectMeasure – Single camera biological analysis tool*. http://www.seagis.com.au/transect.html [Accessed November 28, 2015].

Stamoulis, K. A., Delevaux, J. M., Williams, I. D., Poti, M., Lecky, J., Costa, B., et al. (2018). Seascape models reveal places to focus coastal fisheries management. *Ecol. Appl.* 28, 910–925. doi: 10.1002/eap.1696

Sward, D., Monk, J., and Barrett, N. (2019). A Systematic Review of Remotely Operated Vehicle Surveys for Visually Assessing Fish Assemblages. *Front. Mar. Sci.* 6:134.

Tolimieri, N. (2007). Patterns in species richness, species density, and evenness in groundfish assemblages on the continental slope of the U.S. Pacific coast. *Environ. Biol. Fishes* 78, 241–256. doi: 10.1007/s10641-006-9093-5

Tolimieri, N., and Anderson, M. J. (2010). Taxonomic distinctness of demersal fishes of the California current: moving beyond simple measures of diversity for marine ecosystem-based management. *PLoS One* 5:e10653. doi: 10.1371/journal.pone.0010653

Watson, D. L., Harvey, E. S., Fitzpatrick, B. M., Langlois, T. J., and Shedrawi, G. (2010). Assessing reef fish assemblage structure: how do different stereo-video techniques compare? *Mar. Biol.* 157, 1237–1250. doi: 10.1007/s00227-010-1404-x

Watson, M., and Ormond, R. F. G. (1994). Effect of an artisanal fishery on the fish and urchin populations of a Kenyan coral reef. *Mar. Ecol. Prog. Ser.* 1994, 115–129. doi: 10.3354/meps111115

Wellington, C. M., Harvey, E. S., Wakefield, C. B., Abdo, D., and Newman, S. J. (2021). Latitude, depth and environmental variables influence deepwater fish assemblages off Western Australia. *J. Exp. Mar. Biol. Ecol.* 539:151539. doi: 10.1016/j.jembe.2021.151539

Wellington, C. M., Harvey, E. S., Wakefield, C. B., Langlois, T. J., Williams, A., White, W. T., et al. (2018). Peak in biomass driven by larger-bodied mesopredators in demersal fish communities between shelf and slope habitats at the head of a submarine canyon in the south-eastern Indian Ocean. *Cont. Shelf Res.* 167, 55–64. doi: 10.1016/j.csr.2018.08.005

Westera, M., Lavery, P., and Hyndes, G. (2003). Differences in recreationally targeted fishes between protected and fished areas of a coral reef marine park. *J. Exp. Mar. Bio. Ecol.* 294, 145–168. doi: 10.1016/s0022-0981(03)00268-5

White, H. K., Hsing, P.-Y., Cho, W., Shank, T. M., Cordes, E. E., Quattrini, A. M., et al. (2012). Impact of the Deepwater Horizon oil spill on a deep-water coral community in the Gulf of Mexico. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20303–20308.

Williams, A., Koslow, J. A., and Last, P. R. (2001). Diversity, density and community structure of the demersal fish fauna of the continental slope off western Australia (20 to 35°S). *Mar. Ecol. Prog. Ser.* 212, 247–263. doi: 10.3354/meps212247

Williams, A., Last, P. R., Gomon, M. F., and Paxton, J. R. (1996). Species composition and checklist of the demersal ichthyofauna of the continental slope off Western Australia (20-35 S). *Rec. West. Aust. Mus.* 18, 135–156.

Woo, M., and Pattiaratchi, C. (2008). Hydrography and water masses off the western Australian coast. *Deep Sea Res. Part I* 55, 1090–1104. doi: 10.1016/j.dsr.2008.05.005

Woodall, L. C., Sanchez-Vidal, A., Canals, M., Paterson, G. L. J., Coppock, R., Sleight, V., et al. (2014). The deep sea is a major sink for microplastic debris. *R. Soc. Open Sci.* 1:140317.

Zintzen, V., Anderson, M. J., Roberts, C. D., Harvey, E. S., Stewart, A. L., and Struthers, C. D. (2012). Diversity and composition of demersal fishes along a depth gradient assessed by baited remote underwater stereo-video. *PLoS One* 7:e48522.

Conflict of Interest: KI, MW, and LT was employed by the company BMT Commercial Australia. DM was employed by the company Woodside Energy Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/fmars.2021.706666

The Extent of Seasonally Suitable Habitats May Limit Forage Fish Production in a Temperate Estuary

Mary C. Fabrizio[1], Troy D. Tuckey[1], Aaron J. Bever[2] and Michael L. MacWilliams[2]*

[1] *Department of Fisheries Science, Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, United States,*
[2] *Anchor QEA, LLC, San Francisco, CA, United States*

The sustained production of sufficient forage is critical to advancing ecosystem-based management, yet factors that affect local abundances and habitat conditions necessary to support aggregate forage production remain largely unexplored. We quantified suitable habitat in the Chesapeake Bay and its tidal tributaries for four key forage fishes: juvenile spotted hake *Urophycis regia*, juvenile spot *Leiostomus xanthurus*, juvenile weakfish *Cynoscion regalis*, and bay anchovy *Anchoa mitchilli*. We used information from monthly fisheries surveys from 2000 to 2016 coupled with hindcasts from a spatially interpolated model of dissolved oxygen and a 3-D hydrodynamic model of the Chesapeake Bay to identify influential covariates and construct habitat suitability models for each species. Suitable habitat conditions resulted from a complex interplay between water quality and geophysical properties of the environment and varied among species. Habitat suitability indices ranging between 0 (poor) and 1 (superior) were used to estimate seasonal and annual extents of suitable habitats. Seasonal variations in suitable habitat extents in Chesapeake Bay, which were more pronounced than annual variations during 2000–2016, reflected the phenology of estuarine use by these species. Areas near shorelines served as suitable habitats in spring for juvenile spot and in summer for juvenile weakfish, indicating the importance of these shallow areas for production. Tributaries were more suitable for bay anchovy in spring than during other seasons. The relative baywide abundances of juvenile spot and bay anchovy were significantly related to the extent of suitable habitats in summer and winter, respectively, indicating that Chesapeake Bay habitats may be limiting for these species. In contrast, the relative baywide abundances of juvenile weakfish and juvenile spotted hake varied independently of the spatial extent of suitable habitats. In an ecosystem-based approach, areas that persistently provide suitable conditions for forage species such as shoreline and tributary habitats may be targeted for protection or restoration, thereby promoting sufficient production of forage for predators. Further, quantitative habitat targets or spatial thresholds may be developed for habitat-limited species using estimates of the minimum habitat area required to produce a desired abundance or biomass; such targets or thresholds may serve as spatial reference points for management.

Keywords: habitat suitability, abundance, Chesapeake Bay, weakfish, spot, bay anchovy, spotted hake, boosted regression trees

OPEN ACCESS

Edited by:
Wen-Cheng Wang,
National Taiwan Normal University, Taiwan

Reviewed by:
Dong Liang,
University of Maryland Center for Environmental Science (UMCES), United States
Timothy Targett,
University of Delaware, United States

****Correspondence:***
Mary C. Fabrizio
mfabrizio@vims.edu

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: *07 May 2021*
Accepted: *16 September 2021*
Published: *12 October 2021*

Citation:
Fabrizio MC, Tuckey TD, Bever AJ and MacWilliams ML (2021) The Extent of Seasonally Suitable Habitats May Limit Forage Fish Production in a Temperate Estuary. Front. Mar. Sci. 8:706666. doi: 10.3389/fmars.2021.706666

INTRODUCTION

Trophic interactions among aquatic predators and prey are rarely incorporated in stock assessments (Skern-Mauritzen et al., 2016), yet quantification of such interactions is critical to advancing ecosystem-based management. In the mid-Atlantic, predators such as summer flounder *Paralichthys dentatus,* striped bass *Morone saxatilis,* and bluefish *Pomatomus saltatrix* use estuarine and offshore habitats to support their critical life functions, and their seasonal migrations and trophic interactions maintain connectivity between coastal and offshore ecosystems (Scharf et al., 2004; Latour et al., 2008; Overton et al., 2008). Although feeding habits of many predators are well studied, the distribution and abundance of prey species that comprise the forage base have received less attention (but see Arbeider et al., 2019; Woodland et al., 2021). Estuaries that support relatively high forage production offer spatially extensive habitat conditions that sustain recruitment, growth, and survival; conversely, forage production may be low in estuaries where habitat conditions are degraded. The relationship between population abundance and the extent of suitable habitats has been reported for many species (Holbrook et al., 2000; Parsons et al., 2014; Sundblad et al., 2014; Weber et al., 2017), but has not been widely explored for forage species. Such a relationship, however, may reveal conditions under which habitats limit forage production.

In this study, we focused on forage fishes in Chesapeake Bay because of the availability of temporally and spatially rich data for these taxa and because ecosystem-based approaches to management are currently pursued in the region (Atlantic States Marine Fisheries Commission [ASMFC], 2018; Freitag et al., 2018; Leslie, 2018). The health and sustainability of iconic fisheries in this system depend on sufficient production and availability of forage as well as effective management and protection from anthropogenic degradation of habitats. With the exception of Woodland et al. (2021), habitat conditions necessary to support forage production in this system are not well understood. Our objectives were to (1) quantify suitable habitats for several forage fishes in Chesapeake Bay, and (2) assess the relationship between the extent of suitable habitats and annual abundance of these species. We considered four numerically dominant forage fishes (Tuckey and Fabrizio, 2020): juvenile (age 0) spotted hake *Urophycis regia*, juvenile spot *Leiostomus xanthurus*, juvenile weakfish *Cynoscion regalis* and bay anchovy *Anchoa mitchilli.* Small-bodied fishes such as these are important components of the diets of resident and transient predators in Chesapeake Bay (Buchheister and Latour, 2011, 2015). Because we selected taxonomically and ecologically disparate species, we expected that suitable habitats would be defined by habitat features that differed among species. If the extent of suitable habitats limits the production of forage fishes in Chesapeake Bay, then we would expect annual patterns in forage fish abundances to exhibit patterns similar to those for suitable habitats.

Static features such as substrate type are often used to characterize fish habitats because such features affect distribution and habitat use (Day et al., 1989; Fabrizio et al., 2013). Dynamic environmental conditions such as salinity, temperature, dissolved oxygen (DO), and depth also contribute to variations in the distribution and abundance of estuarine and coastal species. In river-dominated estuaries, river flow affects salinity and alters the extent of suitable habitats for juvenile fishes (Kostecki et al., 2010), many of which may serve as forage for predators. Temperature is a key determinant of habitat suitability for ectotherms because temperature governs critical processes such as metabolic rates, movement, and growth (Little et al., 2020). Low DO conditions are believed to limit suitable habitats for fishes, especially during summer when some estuarine and coastal systems exhibit prolonged seasonal hypoxia. In particular, abundance of fish is low in hypoxic (<2 mg O_2/l) waters (Craig and Crowder, 2005; Tyler and Targett, 2007; Zhang et al., 2009; Buchheister et al., 2013; Glaspie et al., 2019), suggesting that individuals actively avoid hypoxic habitats. Other habitat features such as bottom-current velocities, water column stability, and salinity stratification contribute to hydrodynamic complexity of estuarine systems and as such, may shape variations in the spatial distribution and abundance of estuarine organisms (Manderson et al., 2011; Jenkins et al., 2015; Bever et al., 2016). Indeed, hydrodynamic models can be used to estimate habitat volume for estuarine species when coupled with information on physiological tolerances and bioenergetics requirements (e.g., Schlenger et al., 2013). Hydrodynamic models have also been used to assess the effect of sea-level rise on fishes that depend on marsh habitats for juvenile growth and survival (Fulford et al., 2014), to assess the extent of suitable habitats for fishes in coastal environments (e.g., Le Pape et al., 2003; Bever et al., 2016; MacWilliams et al., 2016), and to evaluate potential impacts of climate change on extent of habitats (Crear et al., 2020a,b).

Fish-habitat relationships are best derived from observations across broad spatial scales and long time periods (Gray et al., 2011; Lecours et al., 2015), thus, we quantified these relationships for each of the four species in Chesapeake Bay and its subestuaries during the 17-year period, 2000–2016. We developed an integrated modeling framework to couple information on the abundance of forage fishes with environmental conditions estimated from two models of Chesapeake Bay, rather than considering only those habitat features measured at the time of fish sampling. The primary data were monthly catches from fishery-independent surveys of forage fishes, hindcasts of dynamic environmental conditions (covariates describing salinity, temperature, current speed, depth, and DO conditions), and estimates of static habitat conditions (sediment composition and distance to shore). We applied a data-driven approach, boosted regression tree analysis (Elith et al., 2008), to select a subset of habitat covariates that were most influential in explaining fish relative abundance. Non-parametric suitability models using the histogram approach (Tanaka and Chen, 2015; Guan et al., 2016) were then constructed using the selected influential environmental covariates. Higher suitability is ascribed to conditions in which greater abundances of organisms are observed, and as such, habitat suitability models are process-based models. We used spatial distributions of the environmental covariates to examine seasonal habitat suitability throughout Chesapeake Bay and tributaries during 2000–2016, because most

of the species we studied are seasonal migrants; all of the species studied use the Chesapeake Bay as a nursery, but the nursery function is temporally restricted. Finally, we assessed the role of habitat area in driving forage fish abundance across the 17 years of study.

MATERIALS AND METHODS

Estimation of Relative Abundance

Geo-referenced catches of forage fishes were obtained from two bottom-trawl surveys: the Virginia Institute of Marine Science Juvenile Fish Trawl Survey (hereafter, Virginia survey) and the Maryland Department of Natural Resources Blue Crab Summer Trawl Survey (hereafter, Maryland survey). The sampling domain of the Virginia survey includes waters greater than 1.2 m throughout Virginia tidal waters of the Chesapeake Bay and its major tributaries (**Figure 1**). Each month, from January to December, the Virginia survey sampled fishes from 111 stations selected from a random stratified survey design (**Supplementary Table 1**). A 30′ semi-balloon bottom trawl was deployed for 5 min at each site; protocol details are available in Tuckey and Fabrizio (2016). The Maryland survey is primarily a shallow-water survey (mean depth = 2.1 m; 99.7% of sites < 5 m deep) that samples fishes from fixed sites in tributaries and sounds of the Maryland portion of the Chesapeake Bay (**Figure 1**). A 16′ semi-balloon

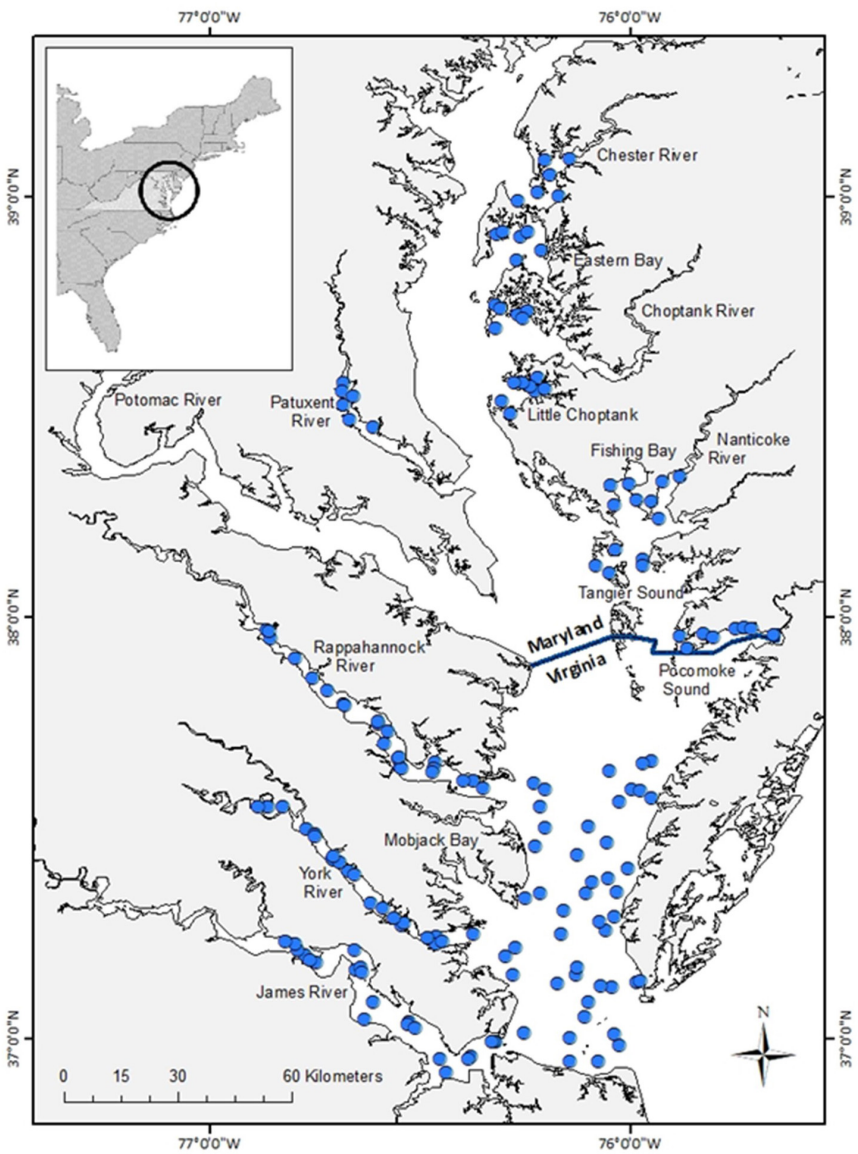

FIGURE 1 | Sites (filled circles) sampled to assess relative abundance of forage fishes in Chesapeake Bay, 2000–2016. Sites in Virginia waters were sampled monthly from a random stratified survey design; sites depicted in the figure are from a representative month and year (October 2020). Fixed sites were sampled monthly between May and October in Maryland waters of Chesapeake Bay.

otter trawl is towed for 6 min at each site. In 2000 and 2001, sampling was conducted monthly from May through October at 37 sites in the Chester River, Choptank River, Eastern Bay, Patuxent River, Pocomoke Sound, and Tangier Sound. In 2002 and thereafter, 16 additional sites were sampled in Fishing Bay, the Little Choptank River, and the Nanticoke River (57 sites total; **Figure 1** and **Supplementary Table 1**). No sampling occurred in Maryland waters in May 2006.

For each trawl tow, we expressed the relative index of abundance of juvenile (age-0) spotted hake, juvenile spot, juvenile weakfish, and all life stages of bay anchovy as the catch per unit effort (CPUE), where effort was estimated by the area swept by the net. Area swept was calculated by multiplying the geodetic distance of the tow by the effective net width, estimated as 55% of the headline spread based on gear tests performed in a flume tank. To relate forage fish abundance to suitable habitat areas and to ensure that CPUE best represented relative abundance, we restricted consideration of CPUE measures to those seasons in which individuals were available to the gear: spring (March, April, May) for juvenile spotted hake; summer (June, July, August) and fall (September, October, November) for juvenile spot and juvenile weakfish; and spring, summer, fall, and winter (December, January, February) for bay anchovy. Note that no sampling was completed in Maryland in winter, thus the bay anchovy CPUE index in winter was based on catches from Virginia waters only.

A Bayesian hierarchical method (Conn, 2010) was used to estimate baywide relative abundance for each species using the survey CPUEs. This method uses the coefficient of variation to weight the individual data sources and extracts a single annual index to represent the pattern exhibited by the multiple indices under the assumption that component indices are subject to process error (from variation in catchability, spatial distribution, etc.) and sampling error (i.e., within-survey variance; Conn, 2010). Annual baywide indices of relative abundance and their associated 95% credible intervals were estimated for the season(s) of interest for each forage species. We used WinBugs accessed through an R script (R Core Team, 2019) to perform these calculations.

Habitat Covariates

Static Habitat Covariates

Two static variables, distance to shore (km) and percent fine sediment, were determined for locations at the midpoint of each trawl tow (**Table 1**). Fringing marsh and other shallow water areas may provide resources that enhance survival and growth of forage fish (e.g., refuge from predators and provisioning of food; Manderson et al., 2004; França et al., 2009; Boutin and Targett, 2019), and as such, distance to shore may influence fish habitat use. The shortest distance to the nearest shoreline was calculated, and in some cases, the nearest shoreline was an in-Bay island (**Figure 2A**). Seabed percent fine sediment, a key feature of fish habitat (Kritzer et al., 2016), was determined from a baywide surface grain-size distribution map (**Figure 2B**) based on observed surface seabed grain size (Moncure and Nichols, 1968; Byrne et al., 1983; Kerhin et al., 1988; Velinsky et al., 1994; Maryland Geological Survey, 1996; Reid et al., 2005).

TABLE 1 | Static and dynamic habitat features considered for optimization of boosted regression trees (BRTs) for forage fishes in Chesapeake Bay, 2010–2012.

Type	Habitat covariate	Units
Static	Sediment composition (percent fine sediment)	%
Static	Distance to shore	m
Dynamic	Water depth	m
Dynamic	Bottom dissolved oxygen	mg O_2/l
Dynamic	Tidal-averaged depth-averaged salinity	PSU
Dynamic	Tidal-averaged surface salinity	PSU
Dynamic	Tidal-averaged bottom salinity	PSU
Dynamic	Tidal-averaged salinity stratification	PSU
Dynamic	Tidal-averaged depth-averaged temperature	°C
Dynamic	Tidal-averaged bottom temperature	°C
Dynamic	Tidal-averaged surface temperature	°C
Dynamic	Tidal-averaged temperature stratification	°C
Dynamic	Tidal-averaged depth-averaged current speed	m/s
Dynamic	Maximum depth-averaged current speed	m/s
Dynamic	Tidal-averaged surface current 1 m below surface	m/s
Dynamic	Tidal-averaged bottom current, 1 m above bottom	m/s
Dynamic	Tidal-averaged vertical stratification in current speed	m/s
Dynamic	Tidal-averaged horizontal gradient in current speed	m/s/m
Dynamic	Percent of time bottom waters < 10°C	%
Dynamic	Percent of time bottom waters between 10° and 20°C	%
Dynamic	Percent of time bottom waters > 20°C	%
Dynamic	Percent of time bottom waters < 10 PSU	%
Dynamic	Percent of time bottom waters between 10 and 20 PSU	%
Dynamic	Percent of time bottom waters > 20 PSU	%

With the exception of the six 'percent time' covariates, the same covariates were used to fit the BRTs to the 2000–2016 observations.

Bottom-Water Dissolved Oxygen

Dissolved oxygen concentrations (mg O_2/l) were hindcast for bottom waters of the Chesapeake Bay and its tributaries using the methods of Du and Shen (2014) modified to include observations from monthly fisheries surveys, quarter-hourly records from Maryland data buoys (Maryland Eyes on the Bay), quarter-hourly records from the Virginia Estuarine and Coastal Observing System (VECOS), and monthly to bi-monthly surveys from the Chesapeake Bay Program's Water Quality Monitoring Program. Monthly mean bottom DO conditions were calculated for each observed location in each year; we used these values at a 1-km^2 resolution to represent bottom DO conditions and then spatially interpolated values for grid cells that did not have estimated DO values. Using a 5-km search radius, we assigned bottom DO values to each 1-km^2 grid cell in one of two ways: if only a single grid cell in the search radius had an estimated DO value, that value was used; if more than one grid cell in the search radius had estimated DO values, then inverse distance weighting was used to obtain a value for the grid cell in question. The search expanded to 10 km in cases where no bottom DO values were available within 5 km. Daily interpolated bottom DO values were then estimated for 2000 to 2016 by linear regression through time in each grid cell using the monthly bottom DO values. Bottom DO observations from a subset of fisheries surveys (2010–2012; n = 4,604) used to develop the model revealed that at least 98% of hindcasts were reliable, that is, values in the normoxic

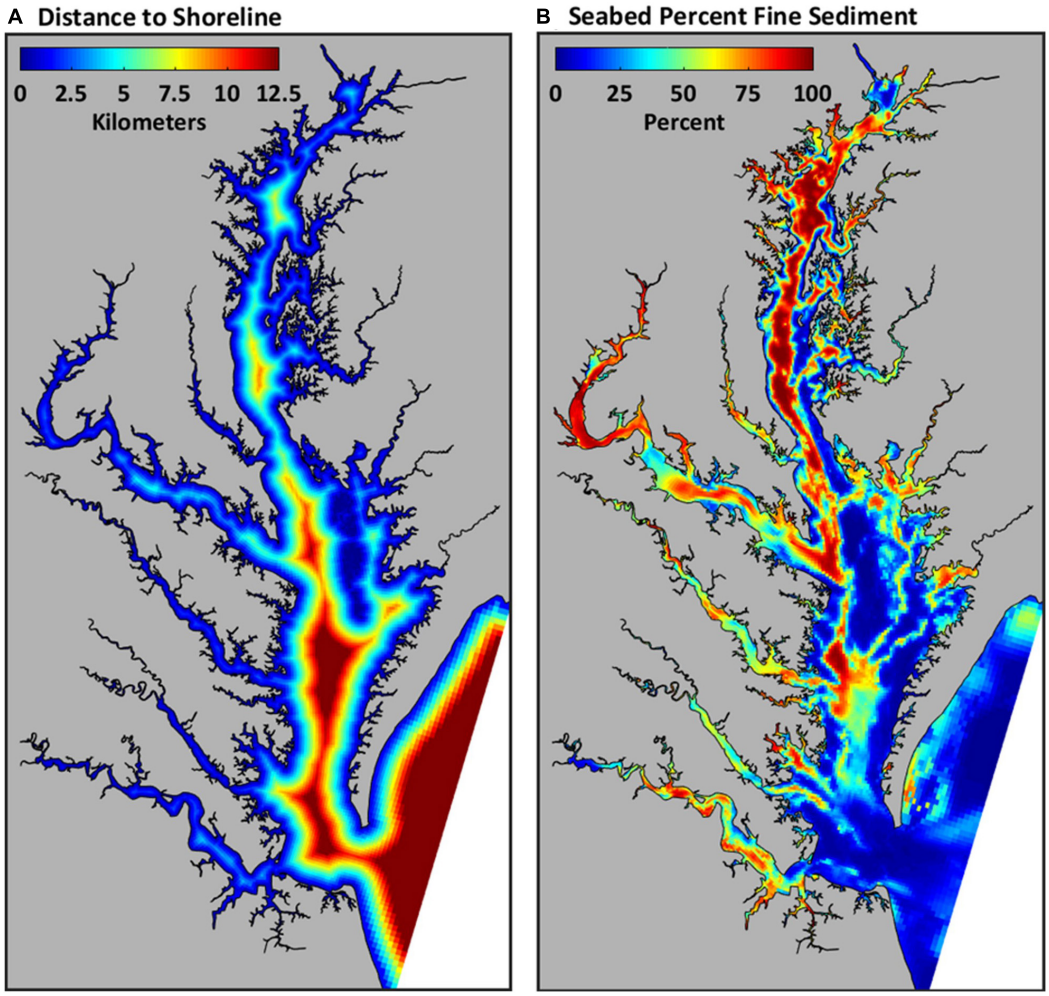

FIGURE 2 | Distance to shoreline (km) **(A)** and sediment composition (% fines) **(B)** of the seabed in Chesapeake Bay.

range (DO > 5 mg O_2/l) were hindcast at 99% of locations where normoxic conditions were observed, and values indicating hypoxia (DO < 2 mg O_2/l) were hindcast at 98.4% of locations with observed hypoxic conditions.

Dynamic Habitat Covariates

Estimates of temperature, salinity, depth, and current speed were obtained from a three-dimensional model of the Chesapeake Bay developed using the UnTRIM hydrodynamic model (Casulli and Zanolli, 2002, 2005). This model takes advantage of the grid flexibility allowed in an unstructured mesh by gradually varying grid cell sizes, beginning with large grid cells in the Atlantic Ocean and transitioning to finer grid resolution in the smaller channels of the tributaries and the northern portion of the estuary (**Figure 3**). This approach allows for the model to accurately capture the bathymetry and shoreline at multiple spatial scales while maintaining suitable simulation times using a single high-end workstation computer. Further description of the inputs for the hydrodynamic model are provided in the **Supplementary Material**, along with a summary of the model validation to the data most relevant to this study. This validation demonstrated that the model accurately estimated temperature and salinity in the Bay and major tributaries under a wide range of environmental conditions (**Supplementary Figures 1, 2**); model accuracy was similar to that reported for a suite of Chesapeake Bay models evaluated by Irby et al. (2016).

A large number of environmental variables were initially considered for use in developing fish habitat suitability models to eliminate *a priori* specification of environmental conditions that may be important for describing abundance and distribution of forage fishes. Twenty-two dynamic environmental variables were calculated, along with sediment composition and distance to shore (**Table 1**). Environmental covariates were extracted from the hydrodynamic model and the DO model at the time and location of the individual tows (midpoint of the tow) to allow us to couple fisheries observations with hindcasts of environmental covariates. We considered hindcasts of environmental covariates at multiple temporal and spatial scales because such measures may provide more accurate predictions of habitat suitability (Lecours et al., 2015). Dynamic variables were extracted from

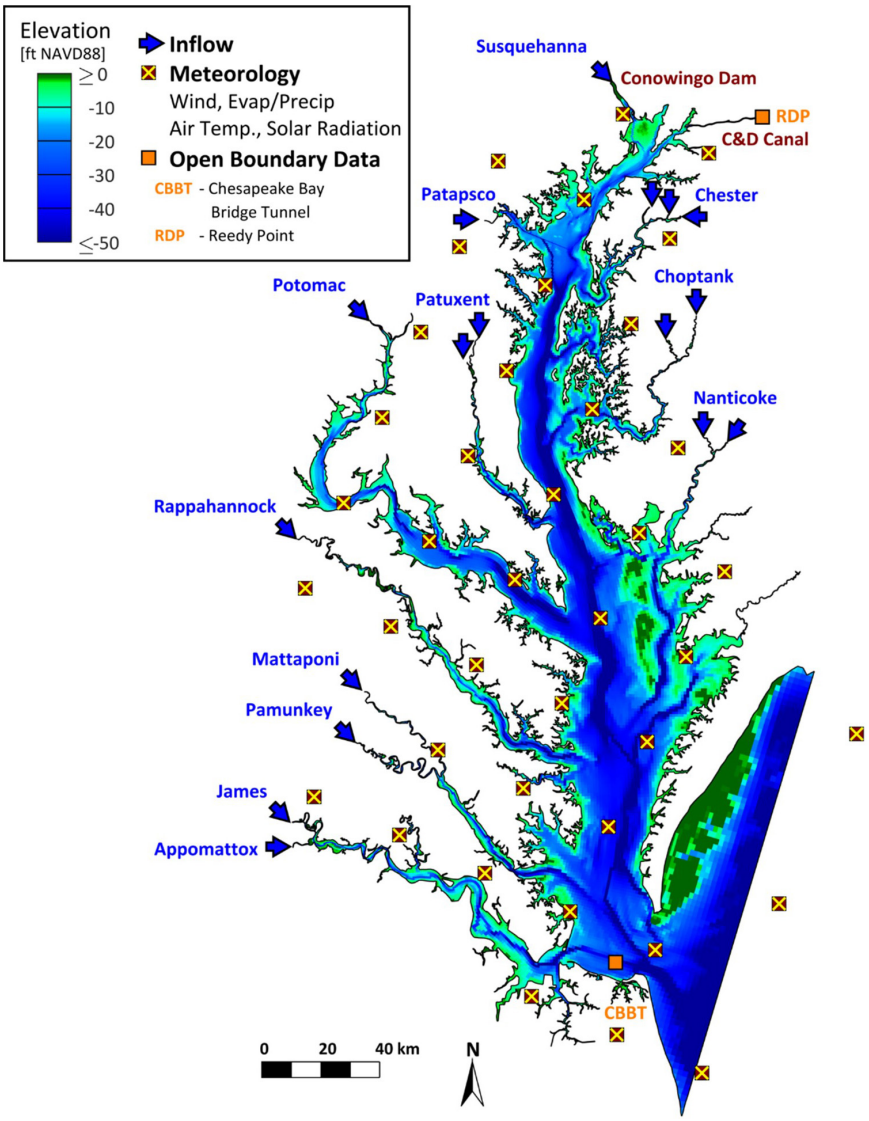

FIGURE 3 | Model domain and boundary conditions for the 3-dimensional UnTRIM Chesapeake Bay model used to hindcast dynamic environmental conditions for forage fishes, 2000–2016.

the hydrodynamic model as instantaneous values at the time of each tow (i.e., the time of capture), and as the tidal-averaged (24.8 h) values in the tidal cycle preceding the capture event. Conditions preceding capture may influence the distribution of fishes (e.g., Jaonalison et al., 2020), and preliminary investigations with boosted regression trees suggested that models that considered tidal-averaged conditions performed better than those with comparable instantaneous values; henceforth, we considered tidal-averaged conditions. Depth-averaged conditions were obtained for salinity, temperature, and current speed (**Table 1**). Tidal-averaged current speeds provide a metric of flow which may be used by some species to aid movements within the estuary (Brady and Targett, 2013). We also considered covariates describing near-bed conditions (1 m above the seabed), and maximum depth-averaged current speed. Because vertical or horizontal gradients in current speed may act to aggregate food near complex currents or fronts, we also considered these covariates. The vertical gradient in the current speed was calculated as the difference between the current speed one meter above the bottom and one meter below the surface; the horizontal gradient in the current speed was calculated as the maximum difference in current speed between adjacent model grid cells divided by the distance between the model grid cells.

Although habitat conditions near the seabed may be most relevant for understanding fish-habitat relationships for demersal species such as spotted hake, habitat use may also reflect overall conditions in the water-column because even demersal fishes are not confined to near-bed habitats. For example, salinity stratification may influence the supply of food or DO to the

near-bed region sampled by the trawl, and may be an indicator of forage fish occurrence. Thus, we used covariates describing salinity and temperature stratification, that were calculated as the difference between instantaneous surface (top 1 m) and near-bed (1 m above seabed) conditions. In addition, because favorable habitats may be characterized by a range of conditions, we considered covariates based on the percent of time that near-bed conditions fell within a given range, for example, the percent of time that salinity exceeded 20 PSU. The percent of time within a given salinity or temperature range was calculated over the same time interval used for tidal averaging. We identified three salinity ranges (<10 PSU; 10–20 PSU; >20 PSU), and three temperature ranges (<10°C; 10–20°C; >20°C) consistent with observed patterns in fish communities in Chesapeake Bay (Tuckey and Fabrizio, personal observation).

Selection of Influential Habitat Covariates

Boosted regression trees (BRTs) were used to select a subset of influential habitat covariates that explained variations in the nominal catch rates of fish (Breiman et al., 1984). The regression tree algorithm uses recursive partitioning to explain variation in the response (nominal catch rates), that is, observations are repeatedly split into increasingly homogeneous groups based on threshold values of the predictors (habitat covariates; Breiman et al., 1984). Cross-validation was used to assess model fit and to ensure that the resultant trees were applicable to out-of-sample observations; cross-validation was achieved by fitting the tree to a subset of the data (the training set) and fit was assessed using the remaining data (the test set). Furthermore, the performance of regression tree algorithms may be improved with ensemble methods such as boosting, which aggregates multiple trees to enhance the stability of the resultant model (Knudby et al., 2010). We used a Poisson response to model the number of fish captured per tow with the *gbm.step* procedure (R package 'dismo,' Elith et al., 2008; Elith and Leathwick, 2017; R Core Team, 2019). Catches from the Virginia survey were used without modification (numbers per 5-min tow), but catches from the Maryland survey were expressed in 5-min-tow equivalencies rounded to the nearest integer. All habitat covariates were standardized to permit direct comparison of covariate importance (Schielzeth, 2010).

Optimization of Boosted Regression Tree Models

Prior to fitting the BRTs to the 2000–2016 observations, we optimized the model-fitting parameters of the BRTs by exploring the combination of parameter values that produced the lowest deviance of the cross-validated data sets (Elith et al., 2008; Cameron et al., 2014). To determine optimal parameters values for the BRTs and because optimization is computationally intensive, we used a subset of observations (2010–2012; n = 4,604 tows) that represented notably different environmental conditions (2011 was a wet year compared with 2010) as well as large differences in the observed relative abundance of forage species. BRT parameters were optimized separately for spotted hake, spot, weakfish, and bay anchovy using the *gbm.step* procedure in R (*dismo* package, R Core Team, 2019). Model fitting failed to produce at least 1,000 trees for bay anchovy using all of the 2010-2012 observations, so we optimized BRT parameters using data from each season separately. For bay anchovy in fall, BRT modeling yielded less than 1,000 trees and was thus unreliable (Elith et al., 2008) and not considered further.

Optimization focused on selection of the learning rate and tree complexity, two model-fitting parameters assigned by the analyst. The learning rate determines how quickly the model approximates the observed data (Miller et al., 2016), and the tree complexity represents the level of interaction possible among the predictors. The third parameter selected by the analyst is the bag fraction, or the proportion of observations used to train the model. Observations in the training subset are selected randomly without replacement for each model run and the remaining observations are used for cross-validation. Preliminary investigations suggested that a bag fraction of 0.75 was reasonable. Using this bag fraction, we fitted a series of trees to a range of learning rates (lr = 0.0005, 0.0050, 0.0075, 0.0100, 0.0200, 0.0300, 0.0400, 0.0500, 0.0750) and tree complexities (tc = 1, 3, 5, 10), similar to Cameron et al. (2014). For each species, we considered only those BRTs for which at least 1,000 trees were fit and selected parameter values that reduced the cross-validated deviance (Elith et al., 2008). We identified the optimal tree complexity for each species by graphically examining the change in cross-validated deviance across learning rates. Next, using the selected tree complexity, we identified the learning rate that produced the minimum cross-validated deviance.

To select influential habitat covariates, we fitted BRT models for each species for the period 2000–2016 (N_{total} = 25,333 tows; $N_{Virginia}$ = 20,326 tows, $N_{Maryland}$ = 5,007 tows) using a bag fraction of 0.75 and values of the optimized species-specific learning rates and tree complexities determined by optimization. Optimized model-fitting parameters (lr and tc) varied among species: juvenile spotted hake lr = 0.01 and tc = 10; juvenile spot lr = 0.02 and tc = 10; juvenile weakfish lr = 0.005 and tc = 10; and bay anchovy (spring, summer, winter) lr = 0.02 and tc = 3. In addition, optimization runs indicated that the six covariates describing percent time were least informative, so these were not considered further. Therefore, a suite of 18 covariates (16 dynamic, 2 static; **Table 1**) were considered for the BRT models. Estimates of variable influence and scree plots produced by the *gbm.step* procedure (R Core Team, 2019) allowed us to identify and select a subset of influential covariates for each species, and for bay anchovy for each season (spring, summer, winter). Care was taken to consider only those covariates that did not exhibit high correlations with other influential covariates, that is, only those covariates with $r^2 < 0.8$ were considered in the subsequent calculation of habitat suitability indices (HSIs). This approach avoids overweighting of the HSI for a particular habitat condition.

Habitat Suitability Models

Habitat suitability models were used to assign habitat suitability scores and to quantify the extent of suitable habitat for forage fishes throughout the Chesapeake Bay and its tributaries

from 2000 to 2016 for each season. Habitat suitability models were estimated with the non-parametric histogram approach because this approach makes no assumption about the nature of the relationship between environmental features and fish abundance (Guan et al., 2016). Briefly, thresholds of environmental conditions that resulted in a gradient of suitability from least suitable (0) to most suitable (1) were identified for each influential habitat covariate for each species. The HSI, calculated as the mean of two or more covariate-specific suitability indices (SIs), also ranged between 0 and 1 to ease interpretation.

Suitability Indices

We estimated SIs for the range of observed values for each of the influential habitat covariates identified by the species-specific BRTs. We used the Tanaka and Chen (2015) approach to estimate SIs but applied a disjoint clustering method to identify 'natural clusters' of the habitat covariates for the histogram approach; we implemented this method with the *FastClus* procedure in SAS/STAT®. Tanaka and Chen (2015) fixed the number of individual bins to 10 for each habitat covariate, but we found that this resulted in bins with few observations (<5) or narrowly defined environmental limits (e.g., bottom temperature between 16.2 and 16.3°C). Thus, we allowed the number of bins to vary (but not exceed 10), and restricted cluster sizes to a minimum of 40 observations; in all cases, the smallest cluster included at least 54 observations, allowing a reasonable description of average abundance (and hence, relative suitability) in each cluster. SIs were estimated for each cluster and habitat covariate using

$$SI_{ij} = \frac{CPUE_{ij} - CPUE_{i,\min}}{CPUE_{i,\max} - CPUE_{i,\min}}$$

where SI_{ij} is the suitability index for cluster j of habitat covariate i, $CPUE_{ij}$ is the average catch (fish/km^2) observed in cluster j of habitat covariate i, and $CPUE_{i,min}$ and $CPUE_{i,max}$ are the minimum and maximum average catches observed across all clusters of habitat covariate i (Tian et al., 2009; Chang et al., 2012). This formulation allows the SIs to range between 0 and 1.0, with 1.0 indicating the most suitable condition and 0 the least. More explicitly, each covariate cluster, defined by a range of values, was associated with an SI score.

Habitat Suitability Indices

Habitat suitability indices were calculated by expressing the HSI as an average of multiple covariate-specific SIs; we restricted the number of covariates in the HSI to those that were most influential as determined by BRTs (**Table 2**). The HSI for a given set of covariates may be expressed as an arithmetic or geometric mean of the individual SIs (e.g., Brown et al., 2000; Tanaka and Chen, 2015). A single averaging approach to estimate the HSIs, however, may not be appropriate for all species (e.g., Yu et al., 2019), so we explored both models of the mean. The arithmetic mean model for the HSI is

$$HSI_{am} = \frac{SI_1 + SI_2 + SI_3 + \cdots SI_p}{p}$$

where SI_1 is the suitability index for habitat covariate 1, SI_2 is the suitability index for covariate 2, and so forth; and p is the number of covariates considered (e.g., Hess and Bay, 2000). The geometric mean model for the HSI is

$$HSI_{gm} = \sqrt[p]{SI_1 \times SI_2 \times SI_3 \times \cdots SI_p}$$

(e.g., Layher and Maughan, 1985; Lauver et al., 2002; Tian et al., 2009). The geometric mean index applies the concept of a 'limiting factor' whereby a low SI for a single covariate results in a low HSI_{gm} (Zajac et al., 2015). HSI calculations were performed in SAS® or Matlab (MathWorks Inc.).

Habitat suitability index models were calibrated by using fish catches at each tow location and graphically examining the relationship between the HSIs and the average relative abundance for each species-season combination (e.g., Tanaka and Chen, 2015). We used the 5% trimmed mean as a measure of the average because trimmed means are insensitive to the occasional extreme catches observed for some species. For a properly calibrated HSI, the mean relative abundance of forage fish is expected to increase as habitat conditions approach optimal for the species, that is, as the HSI increases from 0 to 1.0.

Verification of Modeling Approach

We verified the use of BRTs for selection of covariates and evaluated the reliability of the two averaging formulations of HSI for forage fishes using bootstrap resampling (Efron and Tibshirani, 1986). About 70% of the fisheries observations (N = 18,121) comprised the training data set, and the remaining ∼30% (N = 7,212) was used as the test (or verification) data set. The *SurveySelect* procedure in SAS/STAT® was implemented to randomly select samples without replacement using a stratified design to ensure representation across years, seasons, and geographic areas (Maryland and Virginia). Training and test data sets were constructed separately for each species, and bay anchovy sets were constructed separately for spring, summer, and winter. For each training data set, we fitted BRTs using the same bag fraction and species-specific lr and tc as before; we then selected influential covariates, and modeled the HSI_{am} and HSI_{gm}. Note that the BRTs for each data set may have indicated a different number of influential covariates, as well as a different suite of influential covariates, than what was identified by the original BRT model fitted to observations from 25,333 tows. The resulting HSI models were applied to each of the test data sets to estimate the predicted HSIs. Due to computational intensity, 10 cross-validation data sets were generated (consistent with Pennino et al., 2020). The expected performance of the HSI_{am} and HSI_{gm} for each species across all seasons was evaluated with the root mean square error (RMSE), calculated as the standard deviation of the residuals (where residuals represent the difference between the predicted HSI and observed HSI for each location where fish were sampled). For bay anchovy, we estimated RMSEs for spring, summer, and winter individually. We used a paired t-test implemented in the *glm* procedure in SAS® to assess differences in mean RMSEs, and retained the HSI formulation that exhibited the lower mean RMSE for further analyses.

TABLE 2 | Summary of key results for juvenile spotted hake, juvenile spot, juvenile weakfish, and bay anchovy from Chesapeake Bay, 2000–2016; only those seasons during which fish were available to the gear are shown.

Influential habitat covariates	Season	Mean extent of suitable habitat, km^2 (*SE*) [range]	Change in extent of suitable habitat from 2000 to 2016	Relationship between abundance and extent of suitable habitat?
Juvenile spotted hake				
• Water depth • Tidal-averaged bottom temperature • Tidal-averaged salinity vertical stratification • Maximum depth-averaged current speed	Spring	2,046.0 (60.9) [1,512.4–2,471.5]	No change	No
Juvenile spot				
• Distance to shore • Water depth • Tidal-averaged temperature stratification • Bottom DO • Tidal-averaged current speed horizontal gradient	Summer	4,650.7 (113.6) [4,107.6–5,670.0]	No change	Yes
	Fall	775.4 (54.2) [491.7–1,310.8]	No change	No
Juvenile weakfish				
• Distance to shore • Water depth • Tidal-averaged bottom temperature • Tidal-averaged current speed stratification	Summer	4,394.8 (186.0) [3,080.5–5,559.7]	Increase	No
	Fall	2,914.0 (59.7) [2,457.9–3,271.5]	Increase	No
Bay anchovy				
• Distance to shore • Percent fine sediment • Water depth • Tidal-averaged temperature stratification • Bottom DO	Spring	4,976.9 (88.6) [4,368.1–5,748.0]	Increase	No
• Distance to shore • Percent fine sediment • Water depth • Tidal-averaged bottom temperature • Tidal-averaged surface salinity • Tidal-averaged salinity vertical stratification • Tidal-averaged current speed horizontal gradient	Summer	3,809.0 (97.9) [3,071.7–4,373.0]	Increase	No
• Distance to shore • Percent fine sediment • Water depth • Tidal-averaged surface salinity • Bottom DO • Tidal-averaged current speed horizontal gradient	Winter	2.906.3 (146.8) [2,082.2–4,283.1]	No change	Yes

*Model fitting failed for bay anchovy when data were considered across seasons, so we fit boosted regression trees (BRTs) for each season separately; for bay anchovy in fall, BRT modeling failed to produce at least 1,000 trees and was thus not considered further. DO is dissolved oxygen. Extent of suitable habitat was calculated for each season as the sum of the areas throughout Chesapeake Bay with habitat suitability indices ≥ 0.5; SE is the standard error of the mean. The change in extent of suitable habitats evaluates the monotonic change (increase, decrease, no change) from 2000 to 2016; the relationship between abundance and extent of suitable habitat is based on results from a nonparametric regression (see **Table 3**).*

Estimation of the Extent of Suitable Habitat

The extent of suitable habitat for each species was quantified (objective 1) by calculating HSIs from the environmental covariates at each hydrodynamic model grid cell for each season from 2000 through 2016. To facilitate GIS visualization of seasonal habitat conditions and calculation of the area of suitable habitat for each species, we used the median of the daily values of the covariates to represent the seasonal average for a given model grid cell, season, and year. Median values for each habitat covariate at each model grid cell were then used to estimate the HSIs for each species. In this manner, we mapped the species-specific seasonal HSIs at the spatial resolution of the hydrodynamic model because processes operating at small

TABLE 3 | Non-parametric regression analysis results for juvenile spotted hake, juvenile spot, juvenile weakfish, and bay anchovy from Chesapeake Bay, 2000–2016.

Species – season	*N*	Relative abundance (Y) and suitable habitat extent (X_1)		Suitable habitat extent (Y) and time (X_1)	
		F	*P*	*F*	*P*
Juvenile spotted hake – spring	17	0.01	0.91	0.21	0.65
Juvenile spot – summer	17	4.57	0.05	0.01	0.93
Juvenile spot – fall	17	0.01	0.93	1.11	0.31
Juvenile weakfish – summer	17	2.75	0.12	10.39	<0.01
Juvenile weakfish – fall	17	0.12	0.73	8.68	0.01
Bay anchovy – spring	17	0.67	0.43	10.12	<0.01
Bay anchovy – summer	17	0.06	0.80	24.37	<0.01
Bay anchovy – winter	16	19.98	<0.01	0.17	0.69

The model fitted to the data was $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$ where Y_i is the rank-transformed response, i = 1 to N, N is the sample size, β_0 is the overall average response (intercept), β_1 is the regression coefficient (slope), X_{1i} is the value of the predictor for observation i, and ε_i is the unexplained random error. Extent of suitable habitat was calculated for each season as the sum of the areas throughout Chesapeake Bay with HSI values ≥ 0.5.

spatial scales may be masked when environmental conditions are averaged over large spatial scales (Windle et al., 2012). The areas of the individual grid cells where HSI exceeded a given threshold of suitability (i.e., HSI ≥ 0.5, HSI ≥ 0.6, HSI ≥ 0.7, HSI ≥ 0.8) were summed to obtain an estimate of the extent of 'suitable' habitat throughout the Chesapeake Bay and its tributaries for each species, season, and year. Preliminary graphical investigations revealed that annual patterns in suitable habitat extents were similar among the 0.5, 0.6, and 0.7 thresholds; the 0.8 threshold yielded values that were too low to be useful. The 0.5 threshold, used to describe habitat suitability for the eastern oyster *Crassostrea virginica* in Chesapeake Bay (Theuerkauf and Lipcius, 2016) and for pelagic sharks in Australia (Birkmanis et al., 2020), was used for subsequent analyses and presentation.

Relationship Between Suitable Habitat Extent and Relative Abundance of Forage Species

We hypothesized that annual and seasonal changes in the area of suitable habitat affect the abundance of forage species in this temperate ecosystem. To address this hypothesis, we related the annual time series of suitable habitat with annual estimates of baywide relative abundance for each species (objective 2). We limited the exploration of these relationships to the season during which a particular species was most vulnerable to the trawl gear: juvenile spotted hake in spring, juvenile spot in summer and fall, juvenile weakfish in summer and fall, and bay anchovy in spring, summer, and winter. We rank transformed abundance indices because of the small number of observations ($n = 17$ years except for bay anchovy in winter when $n = 16$) and fit non-parametric regressions to model the relationship between rank abundance and extent of suitable habitat. For this analysis, we assumed the relationship was stationary, that is, the effect of suitable habitat extent on the abundance of forage fishes remained stable from 2000 to 2016 (e.g., Zeng et al., 2018). In addition, we tested the null hypothesis that the seasonal extent of suitable habitat was constant in Chesapeake Bay between 2000 and 2016; as before, we rank transformed the extent of suitable habitat (defined as areas with HSI ≥ 0.5). Computations for non-parametric regression analyses were performed with the *rank* and *glm* procedures in SAS®. Residual plots indicated a reasonable fit of the rank regression model to the data.

RESULTS

Influential Habitat Covariates

Environmental conditions and habitat features that comprised suitable habitats varied among species and ranged between four and seven, depending on species (**Table 2**). Water depth and one of the current speed metrics were consistently identified as influential covariates for all species; one of the temperature covariates was influential in describing suitable habitats for forage fishes in spring, summer, and fall, but was not selected for describing suitable habitats in winter (**Table 2**). At least one salinity metric defined suitable habitats for juvenile spotted hake and bay anchovy, and distance to shore explained suitable habitats for juvenile spot, juvenile weakfish, and bay anchovy (**Table 2**). Bottom DO conditions delineated suitable habitats for juvenile spot in winter and bay anchovy in winter and spring (**Table 2**). Conditions at sites sampled in Maryland waters differed from those in Virginia waters: in Maryland, sampled habitats tended to be shallower, closer to shore, warmer in summer, and cooler in fall than habitats sampled in Virginia (**Supplementary Figure 3**). Most notably, Maryland sites exhibited lower bottom DO concentrations in summer than Virginia sites (**Supplementary Figure 3**). In addition, relative to sites in Virginia, Maryland sites exhibited lower surface salinities, less stratification in terms of salinity and temperature, lower current speeds and less stratification in current speeds (**Supplementary Figure 3**).

Verification and Calibration of the Habitat Suitability Index Modeling Approach

Bootstrap analyses verified that BRTs were useful for selection of influential covariates; in general, the same or similar covariates were identified as most influential among the 10 bootstrap realizations. For juvenile spotted hake, the HSI formulation based on the geometric mean (HSI_{gm}) provided the best approximation to the original HSI estimated for each sample as indicated by the significantly lower RMSE ($t = 4.56$, $P < 0.01$). Unlike results for juvenile spotted hake, the HSI based on the arithmetic mean (HSI_{am}) performed better for bay anchovy, regardless of season ($t_{spring} = -3.08$, $P_{spring} < 0.01$; $t_{summer} = -4.01$, $P_{summer} < 0.01$; $t_{winter} = -3.27$, $P_{winter} < 0.01$). Although we found no evidence for a difference in the mean RMSEs of the HSI_{am} and the HSI_{gm} for juvenile spot and juvenile weakfish ($t_{spot} = -1.13$, $P = 0.27$; $t_{weakfish} = -1.65$, $P = 0.12$), we used the HSI_{am} for these species because the mean RMSE of the HSI_{am} was consistently less than the mean RMSE of the HSI_{gm}.

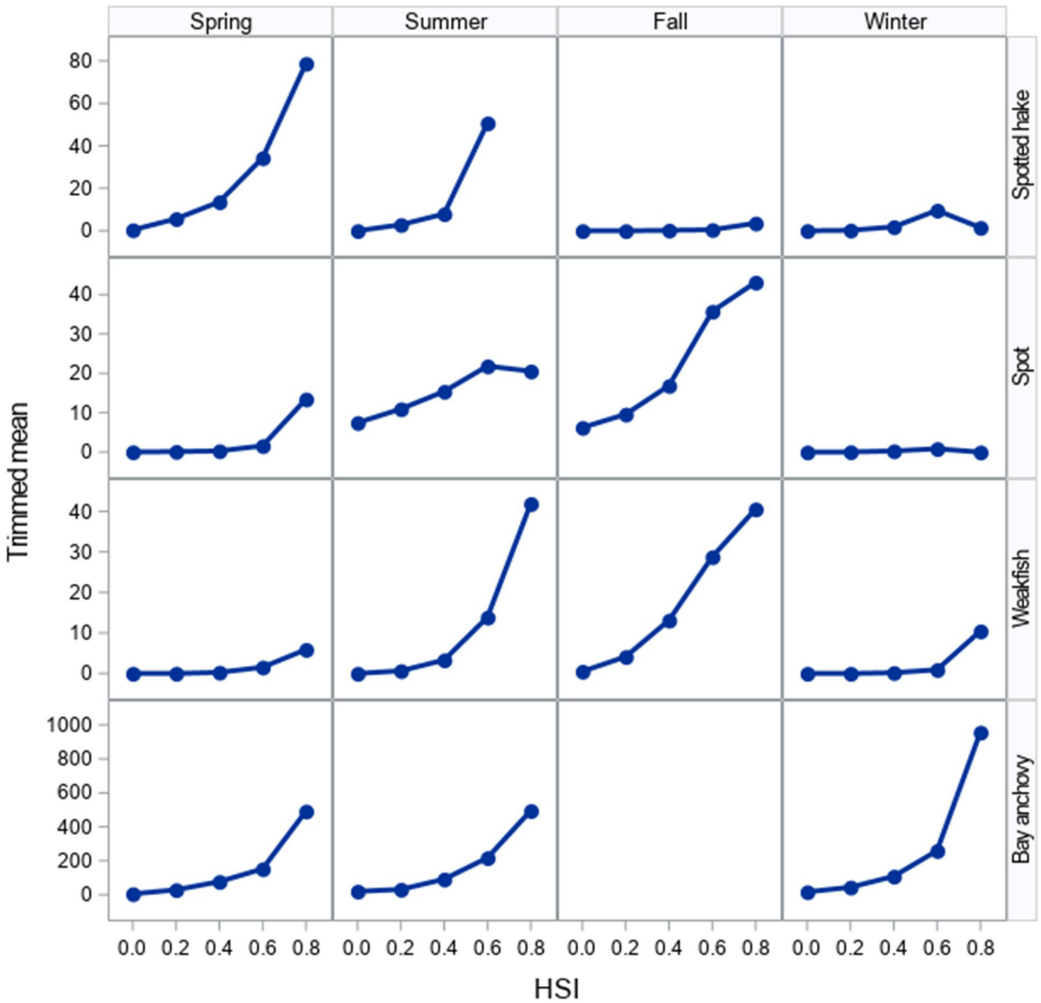

FIGURE 4 | Relationship between HSI and trimmed mean catches of juvenile spotted hake, juvenile spot, juvenile weakfish, and bay anchovy in Chesapeake Bay, 2000–2016. For juvenile spotted hake, the HSI_{gm} is shown, whereas the HSI_{am} is shown for other species. HSI values were binned for ease of plotting such that bin 0.0 includes all observations where $0 \leq HSI < 0.2$, bin 0.2 includes all observations where $0.2 \leq HSI < 0.4$, and so forth. Note differences in *y*-axes.

Relative abundance as measured by the trimmed mean of the four species increased with increasing values of HSIs (**Figure 4**), indicating proper calibration of habitat suitability models. The ranges of observed HSI values across years were 0–0.95 for juvenile spotted hake in spring; 0.12–0.98 for juvenile spot in summer; 0.04–0.86 for juvenile spot in fall; 0.09–0.99 for juvenile weakfish in summer; 0–0.89 for juvenile weakfish in fall; 0–0.92 for bay anchovy in spring; 0–0.92 for bay anchovy in summer; and 0.04–0.98 for bay anchovy in winter.

Suitable Habitat Extent for Forage Fishes

Juvenile Spotted Hake

We detected a strong seasonal pattern in the extent of suitable habitats for juvenile spotted hake in Chesapeake Bay: little to no suitable habitat was available in summer and fall, increased in winter ($mean_{winter}$ = 467.0 km^2 or 4.3% of the total area modeled), and was greatest in spring ($mean_{spring}$ = 2,046.0 km^2 or 18.8% of the total area; **Table 2**, **Figure 5A**, and **Supplementary Figure 4A**). In spring, the extent of suitable habitat was greater in the Virginia portion of the Chesapeake Bay system than in the Maryland portion (**Supplementary Figure 5A**). Suitable habitats for spotted hake in spring were relatively deep, away from shorelines, and where salinity stratification was greater than 4.9 psu (**Figure 5A**); these habitats were characterized by tidal-averaged bottom temperatures ranging between 5.3 and 14.2°C. Annual changes in the extent of suitable habitats varied by as much as 90.3% in winter (**Figure 6A**) and as much as 38.8% in spring (**Figure 6B**) across the time period of study. The extent of suitable winter habitat was higher during years when waters in the region began to warm earlier in the year (e.g., 2012) and was lower when warming was delayed (e.g., 2011; **Figure 6A**).

Juvenile Spot

The extent of suitable habitat for juvenile spot displayed a persistent seasonal pattern, with relatively greater extents of

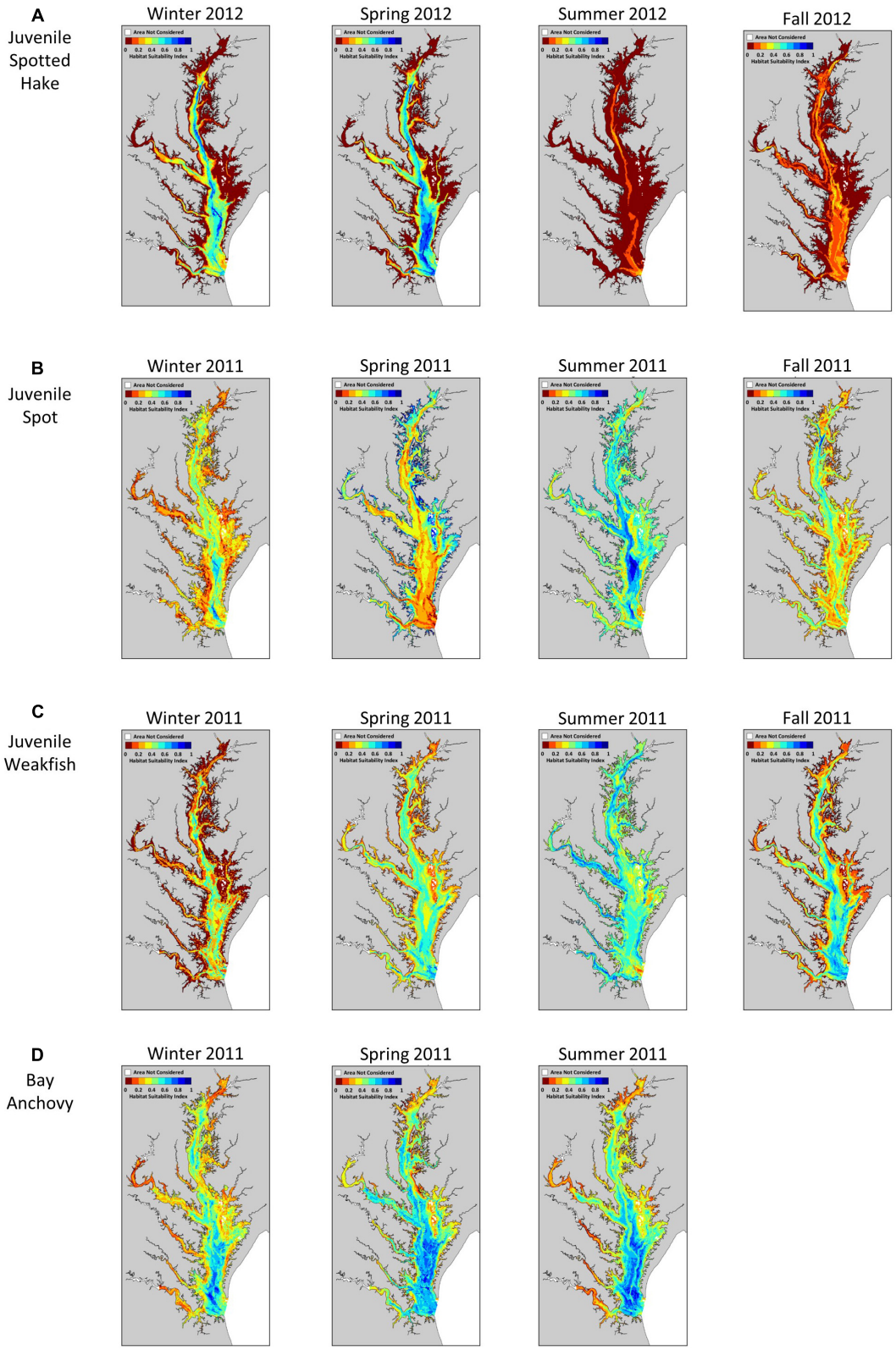

FIGURE 5 | Representative examples of seasonal variation in habitat suitability for **(A)** juvenile spotted hake, **(B)** juvenile spot, **(C)** juvenile weakfish, and **(D)** bay anchovy in Chesapeake Bay. The habitat suitability index ranges from 0 (red) indicating poor habitat to 1 (blue) indicating most suitable habitat.

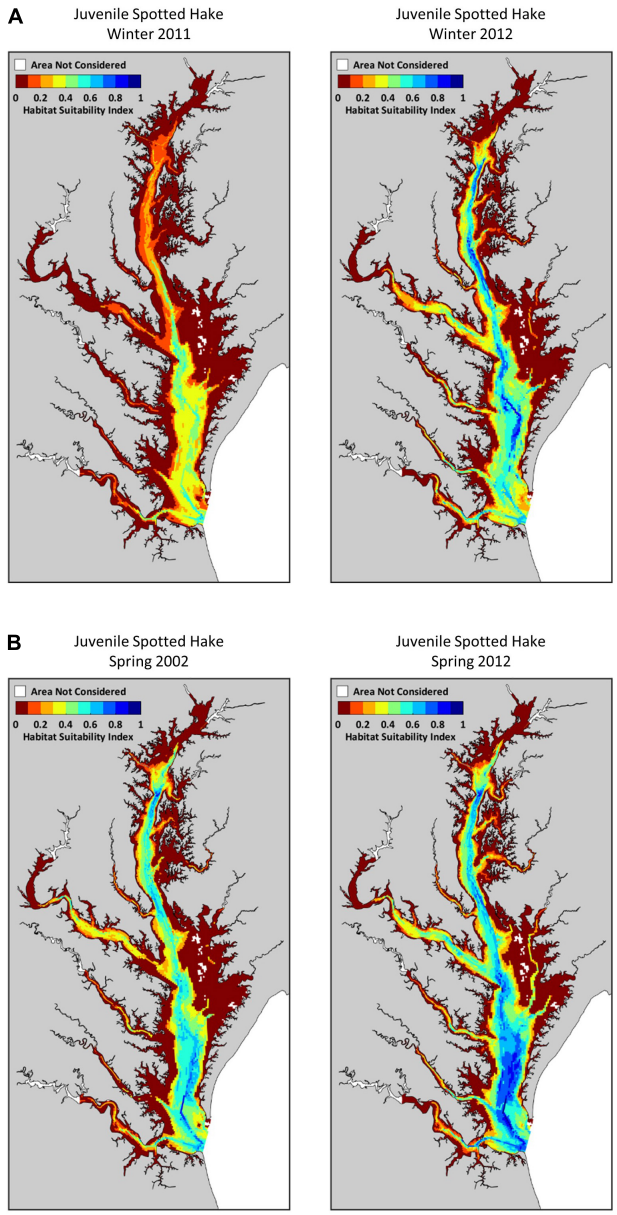

FIGURE 6 | Annual variation in extent of suitable habitats for juvenile spotted hake in Chesapeake Bay: **(A)** winter 2011 vs. 2012 (90.3% increase), and **(B)** spring 2002 vs. 2012 (38.8% increase). The habitat suitability index ranges from 0 (red) indicating poor habitat to 1 (blue) indicating most suitable habitat.

suitable habitats in spring ($mean_{spring}$ = 3,169.4 km^2 or 29.1% of the total area) and summer ($mean_{summer}$ = 4,650.7 km^2 or 42.6% of the total area) and relatively little suitable habitat in fall and winter ($mean_{fall}$ = 775.4 km^2 or 7.1% of the total area; $mean_{winter}$ = 864.0 km^2 or 7.9% of the total area; **Table 2**, **Figure 5B**, and **Supplementary Figure 4B**). Suitable habitats for juvenile spot were primarily found in shallow areas near shorelines in spring, and in the deeper portions of the Chesapeake Bay and its tributaries in summer (**Figure 5B**). In tributaries and embayments such as Mobjack Bay, the extent of suitable habitat for juvenile spot in spring was greater than that in summer by as much as 47.2%. Suitable habitat extents in spring and summer were greater in the Maryland portion of the Chesapeake Bay system compared with the Virginia portion (**Supplementary Figure 5B**). The extent of suitable habitats for spot was driven by distance to shore, depth, temperature stratification, bottom DO, and horizontal gradients in current speed, and was not well described by a single covariate (**Table 2**). Nevertheless, suitable habitats exhibited bottom DO levels less than 4.0 mg O_2/l in summer and less than 5.3 mg O_2/l in fall, suggesting that juvenile spot used habitats that may be considered marginal or unsuitable for other species. Low bottom DO conditions are associated with stratified water column conditions and indeed, juvenile spot were more likely to occur in habitats where tidal-averaged temperature stratification exceeded 2.2°C in summer and 2.7°C in fall.

Juvenile Weakfish

A notable seasonal pattern was evident in the amount of suitable habitat available for juvenile weakfish: little suitable habitat was available in winter ($mean_{winter}$ = 667.3 km^2 or 6.1% of the total area), increased in spring ($mean_{spring}$ = 2,446.3 or 22.4% of the total area) and was greatest in summer ($mean_{summer}$ = 4,394.8 or 40.3% of the total area; **Table 2**, **Figure 5C**, and **Supplementary Figure 4C**). Throughout the Chesapeake Bay system, the extent of suitable habitat in fall was, on average, 19.1% greater than in spring, but this pattern was not observed in 2000, 2010, and 2012 (**Supplementary Figure 4C**). In summer, suitable habitats for juvenile weakfish were found close to shorelines; in fall, suitable habitats were away from shore, near the mouth of the Potomac River and the lower Chesapeake Bay (**Figure 5C**). In Virginia waters, we observed similar extents of suitable habitats in summer ($mean_{summer}$ = 2,056.7 km^2 or 18.9% of the total area) and fall ($mean_{fall}$ = 2,007.7 km^2 or 18.4% of the total area), but in Maryland waters, suitable habitat extents declined by an average of 61.2% from summer ($mean_{summer}$ = 2,338.0 km^2 or 21.4% of the total area) to fall ($mean_{fall}$ = 906.3 km^2 or 8.3% of the total area; **Supplementary Figure 5C**). Regional differences in fall were clearly depicted by seasonal maps: HSI values in the mainstem of the bay were greater in waters south of the Rappahannock River than north of the Rappahannock River (**Figure 5C**). Suitable habitats for juvenile weakfish were characterized by distance to shore, depth, bottom temperature, and current speed stratification (**Table 2**).

Bay Anchovy

The estimated extent of suitable habitat for bay anchovy varied seasonally, with the greatest extent of suitable habitat area in the Chesapeake Bay system in spring ($mean_{spring}$ = 4,976.9 km^2 or 45.6% of the total area), followed by summer ($mean_{summer}$ = 3,809.0 km^2 or 34.9% of the total area) and winter ($mean_{winter}$ = 2,906.3 km^2 or 26.6% of the total area; **Table 2**, **Figure 5D**, and **Supplementary Figure 4D**). Although this seasonal pattern was also observed for Virginia waters ($mean_{spring}$ = 3,117.2 km^2 or 28.6%; $mean_{summer}$ = 2,774.4 km^2 or 25.4%; $mean_{winter}$ = 1,839.8 km^2 or 17.0%), suitable habitat extents in Maryland waters were greatest in spring ($mean_{spring}$ = 1,859.7 km^2 or 17.0%) and lower

but similar in summer and winter ($mean_{summer}$ = 1,034.6 km^2 or 9.5%; $mean_{winter}$ = 1,066.5 km^2 or 9.8%; **Supplementary Figure 5D**). The greatest extents of suitable habitats in tributaries occurred in spring; in summer and winter, the upper reaches of major tributaries were particularly unsuitable for bay anchovy (**Figure 5D**). Suitable habitat conditions occurred in areas that reflected complex relationships between multiple environmental and geophysical covariates and consistently included sediment composition, depth, and distance to shore.

Relative Baywide Abundance and Relationship to Extent of Suitable Habitat

Seasonal indices of relative abundance for forage fishes in Chesapeake Bay varied across years, and interannual patterns were generally similar for the Virginia and Maryland surveys, particularly for juvenile spotted hake in spring, juvenile spot in summer and fall, and juvenile weakfish in summer (**Figure 7**). Inconsistent patterns in interannual changes in relative abundance between Virginia and Maryland waters were observed for juvenile weakfish in fall and bay anchovy in spring and summer (**Figure 7**), suggesting that seasonal processes affecting the distribution of these species varied across regions of the Chesapeake Bay. We note that the mean index of relative abundance for juvenile weakfish in fall was several orders of magnitude lower in Maryland waters than in Virginia. Similarly, the mean indices of relative abundance for bay anchovy in spring and summer were an order of magnitude lower in Maryland waters than in Virginia. Spatial differences in relative abundance reflected the geographic differences in the extents of suitable habitats which were notably lower in Maryland than in Virginia for these species-season combinations (**Supplementary Figures 5C,D**).

Two contrasting relationships were detected between the ranked baywide relative abundance index and the extent of suitable habitat for forage fishes. We observed a significant positive relationship between seasonal ranked baywide relative abundance and extent of suitable habitat for juvenile spot in summer (**Tables 2**, **3** and **Figure 8A**). The baywide relative abundance index for juvenile spot in summer was highly variable with contrasting periods of low and high abundance (**Figure 9A**). Extents of suitable summer habitat for juvenile spot exhibited no significant linear pattern across time (**Tables 2**, **3** and **Figure 9A**). We observed a similar relationship for bay anchovy in winter: the extent of suitable habitat in winter was a significant determinant of the ranked relative abundance of bay anchovy (**Table 3** and **Figure 8B**). The baywide relative abundance of bay anchovy varied widely in winter and exhibited no obvious temporal pattern (**Figure 9B**); extents of suitable winter habitat for bay anchovy showed no systematic change through time (**Tables 2**, **3** and **Figure 9B**).

More commonly, we were unable to detect a significant relationship between the area of suitable habitat and the rank-transformed estimate of baywide relative abundance of forage fishes (**Tables 2**, **3**). For juvenile spotted hake in spring, we found no indication that the extent of suitable habitat was limiting, except perhaps in 2002 when the area of suitable habitat declined below 1,600 km^2 and the ranked abundance index was among the lowest observed in the time series. This, however, may be coincidental. The extent of suitable spring habitat for spotted hake varied without trend since 2000 (**Tables 2**, **3**). We found no evidence of an effect of the extent of suitable habitat on the ranked relative abundance of juvenile spot in fall (**Tables 2**, **3**). The extent of suitable habitat for juvenile spot in fall exhibited no trend through time (**Tables 2**, **3**) and was markedly less than in summer. Between 2000 and 2016, the extent of suitable habitat for juvenile weakfish increased significantly in summer and fall (**Tables 2**, **3** and **Figures 10A,B**). The relative abundance of juvenile weakfish, however, exhibited no detectable response to increases in the extent of suitable habitats in either summer or fall during the study period (**Tables 2**, **3**). Relative abundances of bay anchovy in spring and summer were highly variable (**Figure 7**) and annual estimates were imprecise. Although the extent of suitable spring and summer habitats for bay anchovy increased significantly since 2000 (**Tables 2**, **3** and **Figures 10C,D**), we did not detect a response in the ranked relative abundance of bay anchovy to changes in the extent of suitable habitats in spring or summer (**Tables 2**, **3**).

DISCUSSION

Our modeling framework combined the power of machine learning to identify influential habitat covariates with the flexibility of non-parametric approaches to characterize habitat suitability and the capabilities of GIS to quantify and depict suitable (and unsuitable) habitats for forage fishes in Chesapeake Bay from 2000 to 2016. We coupled catch information from fishery surveys with static features of the environment and outputs from models of dynamic conditions to depict suitable habitats in Chesapeake Bay. In an ecosystem-based approach, these habitats may be targeted for protection (e.g., by limiting habitat alterations or other human activities that may incidentally kill or injure forage fishes) or restoration (e.g., by improving water quality conditions), thereby promoting production of sufficient forage for predators. Importantly, our modeling approach for building forage-fish habitat suitability models for the Chesapeake Bay was calibrated and verified, thereby allowing estimation of habitat suitability for tributaries and embayments in Chesapeake Bay that are not routinely sampled by fishery surveys (e.g., Mobjack Bay, Potomac River). Furthermore, our results allow resource managers to focus protective measures on critical habitat areas in Chesapeake Bay, for example, shallow areas near coastlines in spring and summer which persistently support suitable habitats for multiple forage species. Our finding that shallow areas near coastlines are important habitats for some forage species is consistent with the observed ontogenetic habitat shift of juvenile weakfish, which move from salt marsh tributaries to shallow habitats near coastlines in summer (Boutin and Targett, 2019). We found annual patterns in suitable habitat extent that mirrored those of baywide relative abundance for two of the species examined, juvenile spot in summer and bay anchovy in winter; as such, estimates of the minimum habitat area required to produce a desired abundance (or biomass) of

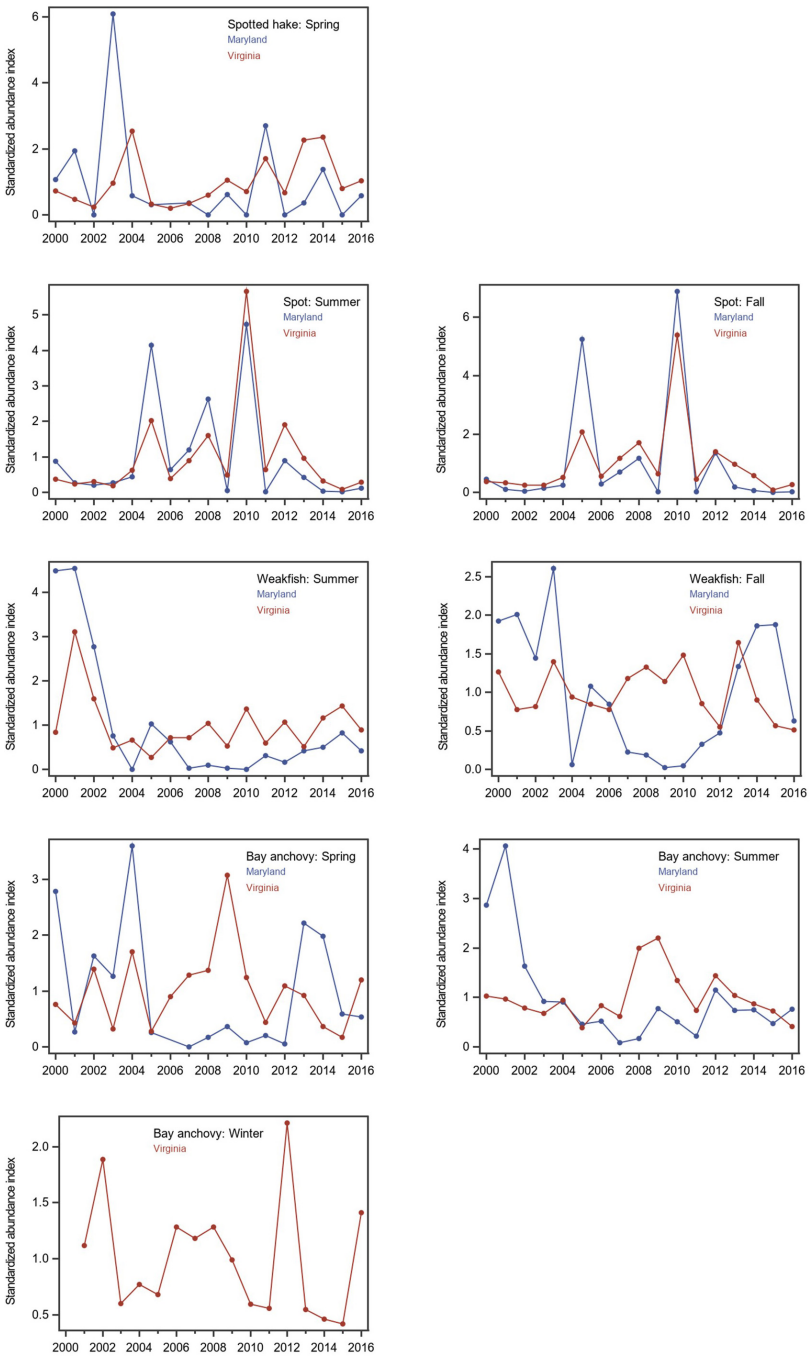

FIGURE 7 | Seasonal standardized abundance indices for forage fishes in Maryland (blue) and Virginia (red) waters of Chesapeake Bay, 2000–2016: juvenile spotted hake in spring, juvenile spot in summer and fall, juvenile weakfish in summer and fall, and bay anchovy in spring, summer, and winter. Seasons were March–May for spring, June–August for summer, September–November for fall, and December–February for winter. Seasonal relative abundance indices (mean catch per unit effort expressed as number of fish/km^2) were standardized to a mean of 1.0 across the 17 years, thus, patterns of abundance can be compared between states, but these standardized abundance indices do not reflect differences in estimated mean catch rates within a given year. Note that bay anchovy were not sampled in Maryland in winter, and thus, only the standardized index for Virginia is depicted.

forage fish can be used to establish quantitative habitat targets (Kritzer et al., 2016) or spatial thresholds that may serve as spatial reference points for management (Reuchlin-Hugenholtz et al., 2016). Quantitative habitat targets and spatial reference points for bay anchovy and juvenile spot in the Chesapeake Bay warrant further consideration.

Suitable seasonal habitat extents for forage species exhibited annual changes reflecting annual-scale heterogeneity in habitat

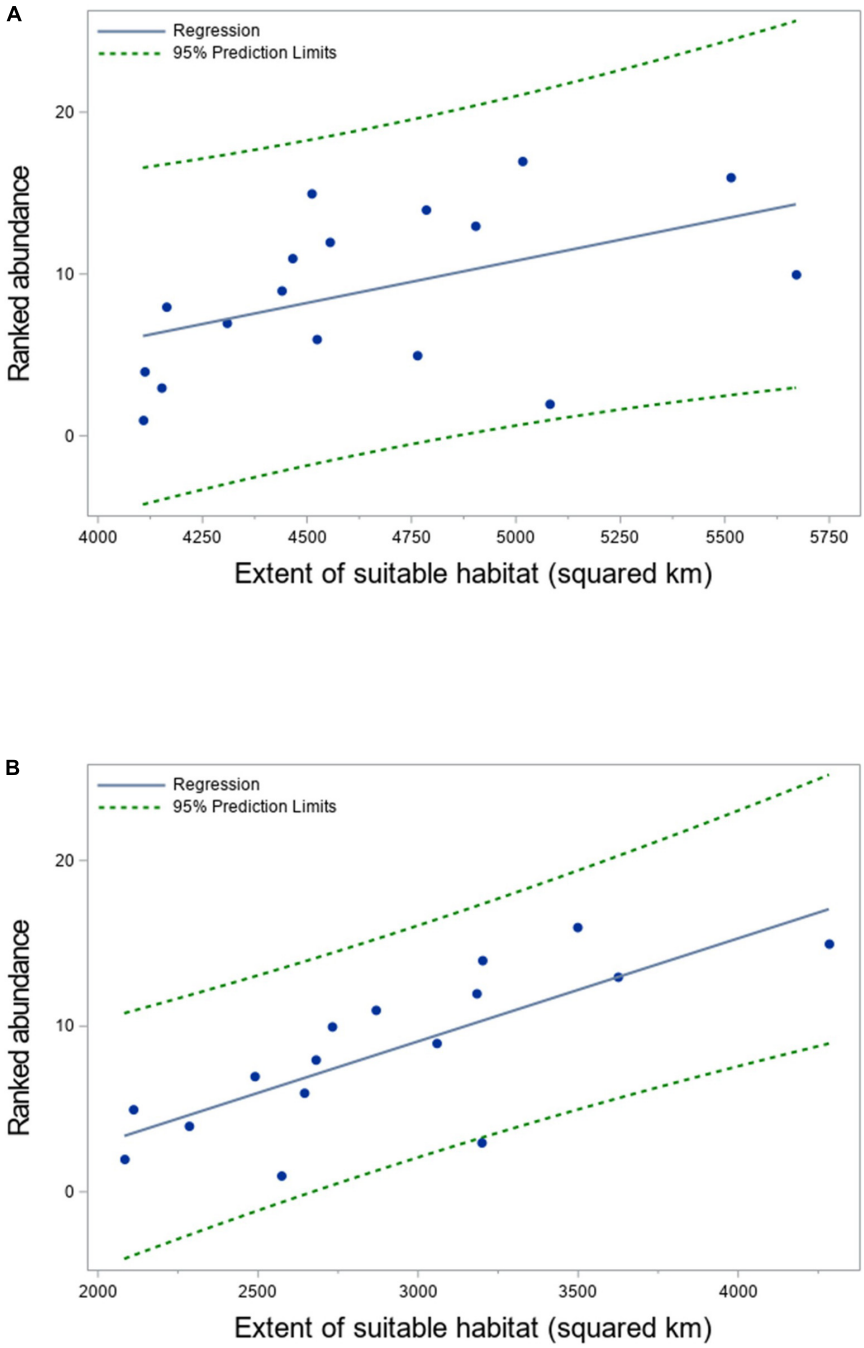

FIGURE 8 | Non-parametric relationship between rank abundance and extent of suitable habitat (km^2) for **(A)** juvenile spot in summer and **(B)** bay anchovy in winter in Chesapeake Bay, 2000–2016. Observations are depicted by filled circles; the solid line is the nonparametric regression fit to the observations, and the dashed line is the 95% prediction limit. The Spearman correlation coefficients were 0.54 for spot in summer and 0.80 for bay anchovy in winter. Values of $HSI_{am} \geq 0.5$ were considered suitable habitat.

conditions in Chesapeake Bay, with persistent seasonal signals that varied among species. Current speed, water depth, and either temperature or DO were important covariates for the four species we examined, and distance to shore was important for three species; thus, suitable habitat conditions resulted from a complex interplay between water quality and the geophysical properties of the habitat. Variations in seasonal extents were more pronounced than annual variations in suitable habitat extent, supporting the notion that the Chesapeake Bay serves as a nursery area for juvenile fishes and that the nursery function is temporally restricted and varies among species (e.g., spring for juvenile spotted hake, and summer for juvenile spot and

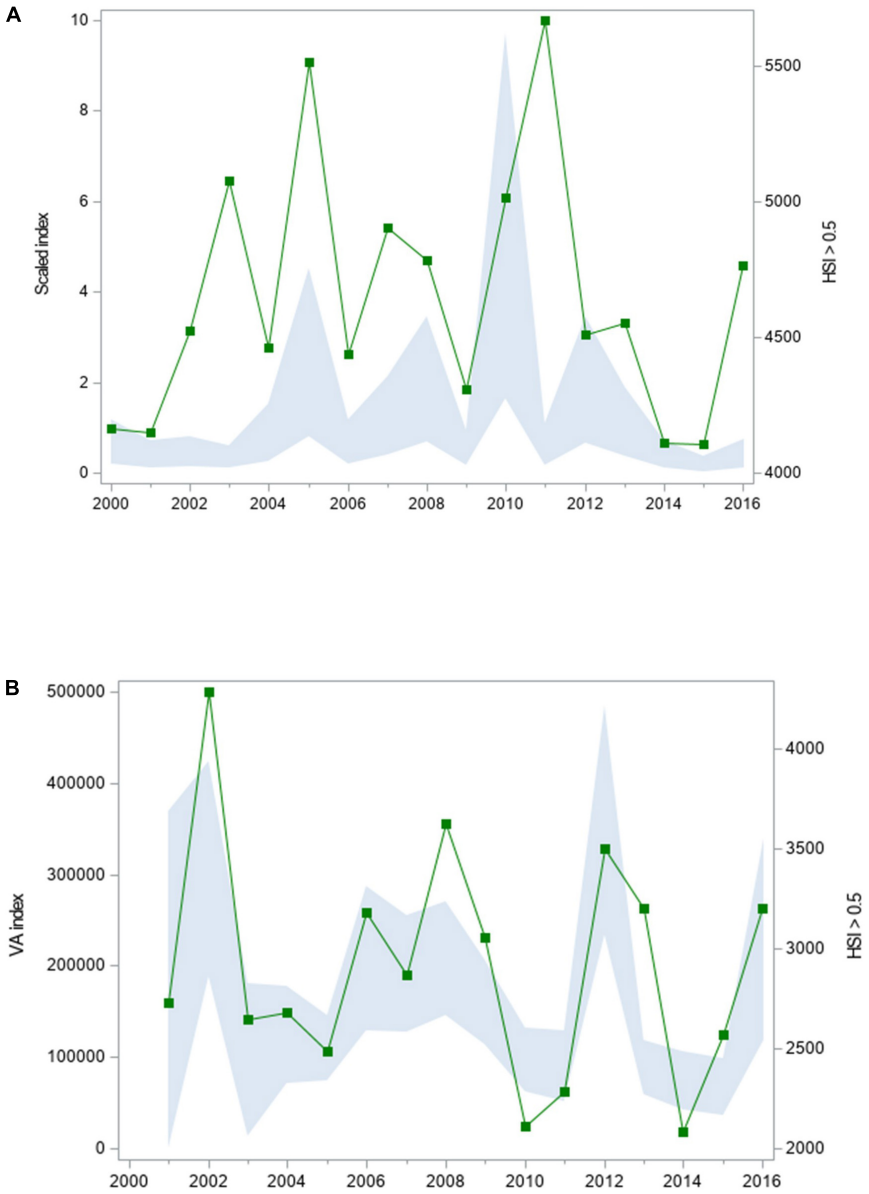

FIGURE 9 | Relative abundance (scaled index) and extent of suitable habitat (km^2) for **(A)** juvenile spot in summer and **(B)** bay anchovy in winter in Chesapeake Bay, 2000–2016. Relative abundance (solid polygon) is depicted with a 95% credible interval; area of suitable habitat is denoted by filled squares. Values of $HSI_{am} \geq 0.5$ were considered suitable habitat.

weakfish). For some species, the extent of suitable seasonal habitat increased since 2000 (summer and fall habitats for juvenile weakfish, and spring and summer habitats for bay anchovy), whereas for other species, extents varied annually with no clear trend. None of the species examined were at the southern limit of their geographic range, and as waters of the Chesapeake Bay continue to warm (Hinson et al., 2021), we expect that suitable habitat extent may increase for species with broad thermal tolerances such as spot, weakfish, and bay anchovy. Other climate-related effects predicted for the region include increased precipitation and sea-level rise, both of which may alter salinity and salinity stratification. Salinity mediates the thermal tolerances of some fishes (e.g., Lankford and Targett, 1994; Nepal and Fabrizio, 2020) and could serve to limit suitable habitats in the future. When coupled with field observations, laboratory-based investigations of the interactive effects of salinity on the thermal tolerances for these forage species could be informative (e.g., Lankford and Targett, 1994).

The relationship between the extent of suitable habitat and relative abundance of forage species was species-dependent and when present, varied seasonally. Such relationships have not been widely explored for aquatic species. Although manipulative field experiments permit exploration of these relationships on small spatial scales (e.g., Parsons et al., 2014), only a few

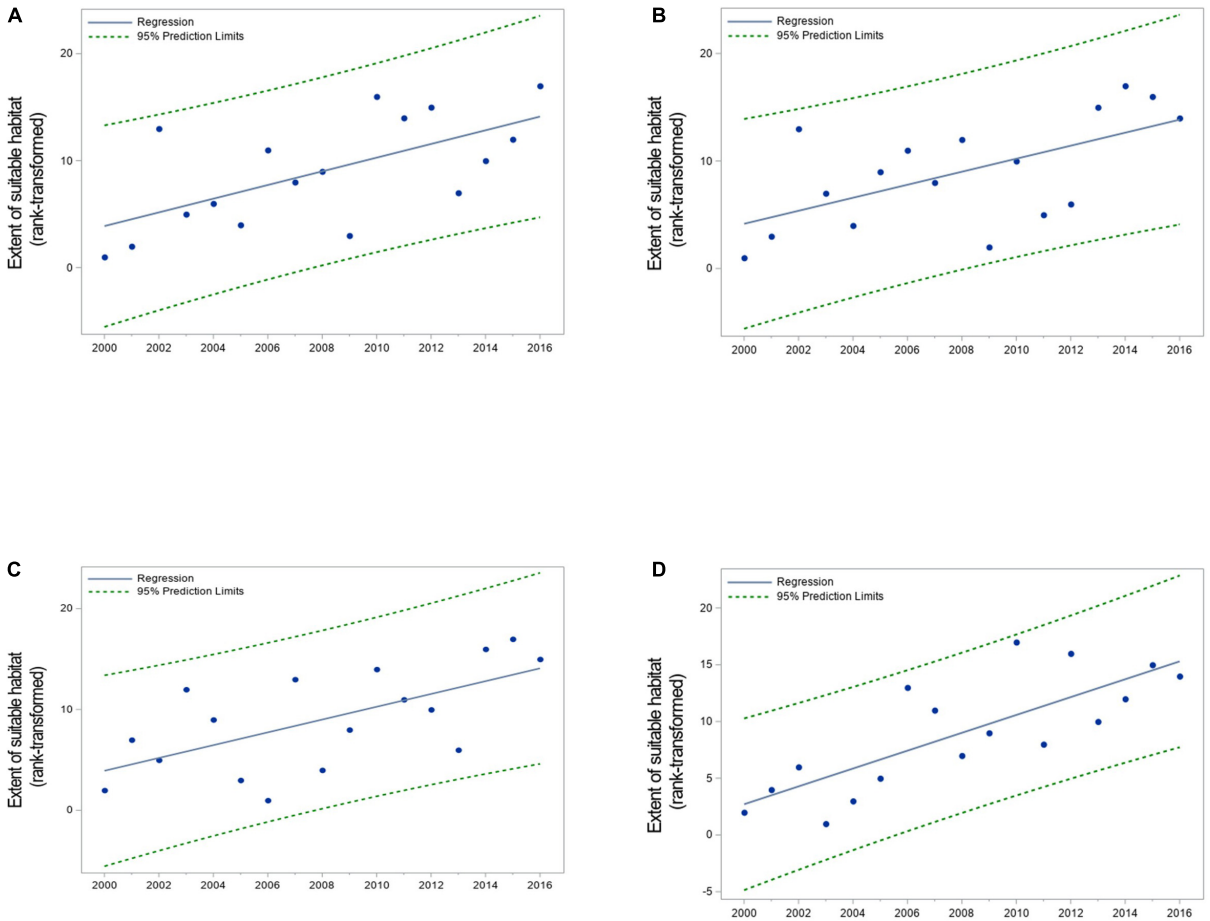

FIGURE 10 | Pattern of change in the extent of suitable habitat for **(A)** juvenile weakfish in summer, **(B)** juvenile weakfish in fall, **(C)** bay anchovy in spring, and **(D)** bay anchovy in summer in Chesapeake Bay, 2000–2016. The extent of suitable habitat was rank-transformed due to the low sample size (n = 17), and the regression line estimates the nonparametric fit to the data. Regression slopes were positive and significantly different from 0 at the $\alpha = 0.05$ level (**Table 3**). Values of $HSI_{am} \geq 0.5$ were considered suitable habitat.

studies attempted to relate extent of suitable habitat and relative abundance of aquatic species using large-scale, field-based observations (e.g., Holbrook et al., 2000; Le Pape et al., 2003; Sundblad et al., 2014; French et al., 2018; Yu et al., 2019). For example, Yu et al. (2019) used a graphical assessment to note the consistency between declines in the relative abundance of neon flying squid *Ommastrephes bartramii* and the spatial shrinkage of suitable habitats in the northwest Pacific Ocean. Sundblad et al. (2014) used linear regression to examine relationships between relative adult abundance and the percent of the study area that was suitable nursery habitat for 12 populations of pikeperch *Sander lucioperca* and European perch *Perca fluviatilis*. Here, we used non-parametric regression to statistically evaluate the strength of such relationships for forage fishes in Chesapeake Bay during 17 years and found a positive relationship between suitable habitat extent and baywide relative abundance of juvenile spot in summer and bay anchovy in winter, indicating that environmental and geophysical conditions affect the carrying capacity of the Chesapeake Bay for these species during a portion of the year. As such, our results are consistent with the findings of a meta-analysis of the correlation between abundance and indicators of habitat suitability (Weber et al., 2017).

We found no evidence that suitable habitat extents were limiting in Chesapeake Bay for juvenile spotted hake in spring, juvenile spot in fall, juvenile weakfish in summer and fall, and bay anchovy in spring and summer. These results suggest that seasonal suitable habitat extent exceeded that necessary to support these populations, and that other factors such as predation mortality (e.g., Minello et al., 1989), food availability (e.g., Tableau et al., 2016), or degradation of egg and larval habitats (e.g., Sundblad et al., 2014) may have contributed to changes in relative abundances. For example, suitable habitat extent for juvenile weakfish in summer and fall increased significantly since 2000, but was not significantly related to changes in relative abundance of juvenile weakfish, suggesting that factors other than those considered in the suitability model exerted a greater role in driving annual fluctuations in abundance. Indeed, abundance and growth of juvenile weakfish are greatest in habitats that contain abundant mysid resources and that are found adjacent to large expanses of

salt marsh (Boutin and Targett, 2019). Consideration of the spatial distribution and annual variation in mysid abundance may be required to further delineate suitable habitats for juvenile weakfish in Chesapeake Bay. Additionally, the lack of a significant relationship between annual relative abundance of juvenile weakfish and suitable habitat extent may have resulted from changes in predation mortality of juvenile weakfish. This hypothesis is consistent with the observed increase in natural mortality rates for this species in the 2000s (Atlantic States Marine Fisheries Commission [ASMFC], 2019; Krause et al., 2020); although the sources of increased natural mortality in weakfish remain unclear, increased predation is believed to have played a role (Atlantic States Marine Fisheries Commission [ASMFC], 2019). Abundances of other forage fishes may be limited by similar trophic interactions with their predators or prey, or by unsuitable egg and larval habitats in Chesapeake Bay.

Although habitat extents limited the relative abundance of juvenile spot in summer, no such limitation occurred in fall. We hypothesize that the summer-to-fall decoupling of the relationship between habitat extent and relative abundance of juvenile spot may be partly explained by temperature. Mean water temperatures in Chesapeake Bay are greatest in late August-early September and the increased metabolic rates and energy demands of predators during this time may increase predation mortality on juvenile spot. Rising water temperatures during this time may also affect the phenology of spot emigration resulting in earlier emigration and fewer juvenile spot remaining in the estuary during fall in warmer years. We were unable to detect a change in the extent of suitable fall habitats for juvenile spot between 2000 and 2016, but warming water temperatures in fall associated with directional climate change may alter future habitat conditions in fall for this species. Continued monitoring of fall abundances and a better understanding of the cues that trigger spot emigration in fall will be necessary to address this hypothesis.

Another possible reason for our inability to detect a relationship between suitable habitat area and baywide relative abundance of some forage fishes is that the suite of static and dynamic habitat covariates we considered failed to adequately describe suitable habitat conditions. For example, the abundance of predators, availability of prey, and proximity to biogenic habitats such as seagrass or oyster reefs may help to shape the distribution and abundance of forage fishes. In addition, seasonal relative abundance indices for many of the forage species were imprecise and hence were statistically invariable across time (based on 95% credible intervals of the baywide hierarchical index). Thus, a 'good' year with a relatively high mean index of abundance was not statistically discernible from a 'poor' year with a relatively low index; this lack of contrast may have hampered our ability to uncover a relationship between relative abundance indices and suitable habitat extents. These results suggest that if the abundance of forage fishes in Chesapeake Bay is changing, the temporal and spatial intensity of sampling by current fisheries surveys is insufficient to detect changes. Alternatively, and more likely, abundance may be stable but highly variable across years, as is typically exhibited by forage fishes (Hilborn et al., 2017). Continued monitoring will be critical to detect directional changes in abundance.

Seasonal variation in the geographic location of suitable estuarine habitats may be linked to variation in freshwater input (e.g., Rubec et al., 2019). In Chesapeake Bay, freshwater input influences salinity and salinity stratification; however, we identified additional hydrodynamic covariates such as temperature stratification and current speed that contributed to variation in suitable habitats. For example, for bay anchovy, the suitability of winter habitats was partly determined by bottom DO and the horizontal gradient in tidal-averaged current speeds. Water temperature was not considered in the HSI model for bay anchovy in winter, however, we cannot rule out temperature as an important covariate describing habitat conditions for bay anchovy in winter because bottom temperature and DO were correlated. Interestingly, the HSI model for juvenile spot also included bottom DO and a measure of temperature stratification, suggesting that multiple hydrodynamic factors are required to describe suitable habitats.

In this study, we considered information from two fishery-independent surveys, one of which yielded fewer annual observations but sampled shallow-water habitats across a large portion of the estuarine salinity gradient. Overall, our fisheries observations were collected at a relatively fine spatial resolution (>100 sites sampled/month) and high temporal intensity (monthly), thereby minimizing biases due to seasonal or short-term habitat use (e.g., by sampling only one or two months each year). By using fisheries observations from multiple surveys, we were able to detect the effect of bottom DO on the relative baywide abundance of juvenile spot in summer. Low bottom DO conditions in summer are more prevalent and persistent in Maryland waters of the Chesapeake Bay; in particular, the Chester River, Eastern Bay, Choptank River, Little Choptank River, and Patuxent River exhibited bottom DO levels in summer that were markedly lower than what was observed in Virginia waters. These observations extended the range of bottom DO conditions over which these effects could be investigated. In addition, the Maryland survey sampled sites in shallow habitats close to shore, and these conditions were not well sampled in Virginia waters. Together, the two surveys provided observations from a greater range of environmental and bathymetric conditions commonly encountered in the Chesapeake Bay region.

Hydrodynamic models and other numerical models of environmental conditions can provide information on dynamic habitat features that are not measured at the time of sampling and represent a significant advancement toward refining spatial relationships between fish and their environment (e.g., Crear et al., 2020a). Consideration of such information may yield habitat models with greater predictive accuracy (Scales et al., 2017). In our study, we used tidal-averaged conditions to develop habitat suitability models that reasonably reflected the relationship between (daily) environmental conditions and relative abundance (at the tow level) of forage fishes in Chesapeake Bay. This fine-scale approach is preferable to one that uses seasonal averages of habitat conditions to build habitat suitability models (Scales et al., 2017). Indeed, the temporal resolution of environmental covariates used to

build the suitability model affects the scale of inference. For example, Woodland et al. (2021) described annual patterns in the distribution and abundance of forage fishes and invertebrates relative to patterns of predation and environmental conditions in the Chesapeake Bay and its major tributaries. Seasonal changes, however, could not be addressed because several habitat conditions were represented by annual means (e.g., average discharge from tributaries, and the Atlantic Multidecadal Oscillation [AMO] index). Large-scale climatic changes as indexed by the AMO affect mean abundances of bay anchovy and juvenile spot (Woodland et al., 2021), and our findings for bay anchovy in winter and juvenile spot in summer are consistent with findings in Woodland et al. (2021). Specifically, our results suggested that variations in environmental conditions contributed to the observed variation in relative abundance of these forage species. Although we did not consider large-scale climate indices *per se*, we did examine small-scale environmental indicators of climate change (temperature, salinity) and demonstrated how these changes affect habitat suitability and relative baywide indices of abundance for juvenile spot in summer and bay anchovy in winter. Unlike Woodland et al. (2021) who found greater relative abundance of bay anchovy in the upper bay than in the lower bay, our indices of relative abundance for bay anchovy in summer were an order of magnitude lower in Maryland than those observed in Virginia waters, likely because we lacked samples from deep sites (>2.7 m) in Maryland which were considered in Woodland et al. (2021). Also, the time scale of our studies differed. Similar to Woodland et al. (2021), we found that bay anchovy and juvenile spot exhibited higher relative abundances in southern tributaries than in northern tributaries of the Chesapeake Bay, but our spatial depiction of suitable habitat conditions throughout the system allowed us to examine the fine-scale spatial distribution of suitable habitats. Such depictions are helpful for identifying the geographic scale at which habitat protection or restoration must be implemented.

As the number of habitat descriptors available to researchers increases, variable reduction techniques are critical to the selection of covariates that help explain the variation in observed abundance and distribution of aquatic organisms. For example, satellite imagery, ocean observing systems, and hydrodynamic models yield a multitude of environmental descriptors of habitat and these data are commonly used to inform fish habitat models. Similar to Georgian et al. (2019), we used BRTs to identify influential covariates from a large suite of possible covariates. Because tree complexity can play a large role in improving the outcome of cross-validation and BRT model fitting, BRT parameters should be optimized using the data under consideration. Many researchers either fail to optimize regression trees or when optimization is implemented, only a single parameter is optimized (typically learning rate; e.g., Georgian et al., 2019; Yu et al., 2020) after using the default bag fraction (0.75) and selecting an arbitrary value for tree complexity (typically between 2 and 5; e.g., Georgian et al., 2019; Pennino et al., 2020; Yu et al., 2020). We recommend optimization of BRTs using the approach we applied here or the R package *gbm.auto* (Dedman et al., 2017).

Across the four species we examined, the geometric mean formulation of the HSI was best for only a single species – juvenile spotted hake. Although the HSI_{gm} is widely used, this formulation may penalize the index too harshly for mobile species that can tolerate broad variations in environmental conditions, including sub-optimal conditions for limited periods of time. For instance, in areas where bottom temperature exceeds a given species' thermal tolerance, the SI for bottom temperature may be close to 0; in this case, the value of the HSI_{gm} will also be close to 0, but other environmental conditions in these areas may be suitable, even optimal, and thus, the overall suitability may not be well indexed by the HSI_{gm}. Because the appropriate HSI formulation depends on species, we recommend use of data-driven analyses and assessment of model performance to inform selection (Chang et al., 2012; Tanaka and Chen, 2015; Yu et al., 2019; this study).

Our fisheries observations from two surveys that sampled across a broad geographic area of the largest estuary in the United States represented 17 years of monthly sampling and reflected the breadth of habitat conditions that fishes were likely to encounter in Chesapeake Bay. Although survey designs differed (stratified random design in Virginia, and fixed-site design in Maryland), the integration of information from such surveys can provide models with good predictive performance as long as observations are spatially extensive (Soranno et al., 2020). Furthermore, when the number of observations used to fit the model is sufficiently large, then projections for unsampled areas within the same time frame are considered valid interpolations (Elith and Leathwick, 2009; Soranno et al., 2020). Our projections of HSIs in areas not sampled by the Maryland or Virginia surveys (e.g., Mobjack Bay, Potomac River) were based on 25,333 observations, and as such, were valid interpolations for assessment of habitat conditions in non-sampled areas during the timeframe of the study (2000–2016). Our interpolations did not 'extend beyond the conditions represented by the data used to fit the model,' and thus we avoided extrapolations to areas where novel combinations of predictors occur (Conn et al., 2015).

Finally, we note that the uncertainties associated with habitat suitability modeling and resulting projections are not typically assessed (Elith and Leathwick, 2009), although such uncertainties are useful for understanding the limitations of model-based results for conservation and fisheries management. Uncertainties may arise from model specification (e.g., the type of model used, or covariates omitted from the model) and from the observations used to fit the model (e.g., samples may not represent the population of interest, or sample size may be inadequate). Ensemble approaches have been used to partially mitigate model uncertainty, but ensemble models do not fully overcome the limitations of the individual component models (Elith et al., 2010). Model and observational uncertainties may affect habitat suitability projections in different ways, and uncertainty analysis for HSI models warrants further research (Zajac et al., 2015) and engagement with environmental statisticians.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: William & Mary ScholarWorks (https://scholarworks.wm.edu/) provides open public access to the monthly abiotic data collected as part of the VIMS trawl survey (https://doi.org/10.25773/97cg-7w42), the seasonal habitat suitability maps (https://doi.org/10.25773/d37g-1h89), and the estimated areas of suitable habitat (https://doi.org/10.25773/d37g-1h89) for 2000 to 2016. Maps and data are under a CC-NC-SA 4.0 license. Maryland trawl survey data should be requested from the original data source.

AUTHOR CONTRIBUTIONS

All the authors conceptualized the research, performed the analyses, interpreted the results, read and approved the published version of the manuscript. MF prepared the original draft of the manuscript. TT, AB, and MM contributed to writing and editing of the manuscript. MF and AB contributed to the project administration and visualization. MF and TT contributed to the funding acquisition.

FUNDING

This study was funded by the NOAA Chesapeake Bay Office, grant NA17NMF4570156 to the authors.

ACKNOWLEDGMENTS

The UnTRIM code used in the hydrodynamic modeling was developed and provided by Professor Vincenzo Casulli (University of Trento, Italy). We thank J. Shen (Virginia Institute of Marine Science, VIMS) and B. Marcek (US Fish & Wildlife Service) for extending the model of bottom DO conditions. We are grateful to J. Buchanan (VIMS) and V. Nepal (VIMS) for application and verification of the DO model, to W. Lowery (VIMS) for preparing **Figure 1**, and to R. Dixon (VIMS) and S. Smith (VIMS) for assistance with verification of the HSI modeling approach. We thank C. Walstrum (MD DNR) for providing fisheries data from the Maryland trawl survey. We appreciate the steadfast sampling in Virginia tidal waters by Captains H. Brooks, V. Hogge, and W. Lowery, and the scientific crew of the VIMS Juvenile Fish Trawl Survey that participated in trawl sampling between 2000 and 2016. Finally, we thank the reviewers for comments that helped to improve the manuscript. This manuscript is Contribution No. 4051 of the Virginia Institute of Marine Science, William & Mary.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.706666/full#supplementary-material

REFERENCES

Arbeider, M., Sharpe, C., Carr-Harris, C., and Moore, J. W. (2019). Integrating prey dynamics, diet, and biophysical factors across an estuary seascape for four fish species. *Mari. Ecol. Progr. Ser.* 613, 151–169. doi: 10.3354/meps12896

Atlantic States Marine Fisheries Commission [ASMFC] (2018). *Five-Year Strategic Plan 2019-2023*. Available online at: http://www.asmfc.org/fisheries-management/program-overview (accessed September 23, 2021).

Atlantic States Marine Fisheries Commission [ASMFC] (2019). *Weakfish Stock Assessment Update Report*. Available online at: http://asmfc.org/species/weakfish (accessed September 23, 2021).

Bever, A. J., MacWilliams, M. L., Herbold, B., Brown, L. R., and Feyrer, F. V. (2016). Linking hydrodynamic complexity to delta smelt (*Hypomesus transpacificus*) distribution in San Francisco Estuary, USA. *San Francisco Estuary Watershed Sci.* 14, 1–27. doi: 10.15447/sfews.2016v14iss1art3

Birkmanis, C. A., Freer, J. J., Simmons, L. W., Partridge, J. C., and Sequeira, A. M. M. (2020). Future distribution of suitable habitat for pelagic sharks in Australia under climate change models. *Front. Mari. Sci.* 7:570. doi: 10.3389/fmars.2020.00570

Boutin, B. P., and Targett, T. E. (2019). Density, growth, production, and feeding dynamics of juvenile weakfish (*Cynoscion regalis*) in Delaware Bay and salt marsh tributaries: spatiotemporal comparison of nursery habitat quality. *Estuaries Coasts* 42, 274–291. doi: 10.1007/s12237-018-0457-9

Brady, D. C., and Targett, T. E. (2013). Movement of juvenile weakfish Cynoscion regalis and spot Leiostomus xanthurus in relation to diel-cycling hypoxia in an estuarine tidal tributary. *Mari. Ecol. Progr. Seri.* 491, 199–219. doi: 10.3354/meps10466

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. New York: Chapman and Hall.

Brown, S. K., Buja, K. R., Jury, S. H., and Monaco, M. E. (2000). Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays, Maine. *North Am. J. Fish. Manag.* 20, 408–435. doi: 10.1577/1548-8675(2000)020<0408:hsimfe>2.3.co;2

Buchheister, A., Bonzek, C. F., Gartland, J., and Latour, R. J. (2013). Patterns and drivers of the demersal fish community of Chesapeake Bay. *Mari. Ecol. Progr. Ser.* 481, 161–180. doi: 10.3354/meps10253

Buchheister, A., and Latour, R. J. (2011). Trophic ecology of summer flounder in lower Chesapeake Bay inferred from stomach content and stable isotope analyses. *Trans. Am. Fish. Soc.* 140, 1240–1254. doi: 10.1080/00028487.2011.618364

Buchheister, A., and Latour, R. J. (2015). Diets and trophic-guild structure of a diverse fish assemblage in Chesapeake Bay, U.S.A. *J. Fish Biol.* 86, 967–992. doi: 10.1111/jfb.12621

Byrne, R. J., Hobbs, C. H. III, and Carron, M. J. (1983). *Baseline Sediment Studies To Determine Distribution, Physical Properties, Sedimentation Budgets, And Rates in The Virginia Portion Of The Chesapeake Bay. Final Report To The US Environmental Protection Agency Chesapeake Bay Program*. Gloucester Point, VA: Virginia Institute of Marine Science.

Cameron, M. J., Lucieer, V., Barrett, N. S., Johnson, C. R., and Edgar, G. J. (2014). Understanding community-habitat associations of temperate reef fishes using fine-resolution bathymetric measures of physical structure. *Mari. Ecol. Progr. Ser.* 506, 213–229. doi: 10.3354/meps10788

Casulli, V., and Zanolli, P. (2002). Semi-implicit numerical modelling of non-hydrostatic free-surface flows for environmental problems. *Math. Comput. Mod.* 36, 1131–1149. doi: 10.1016/s0895-7177(02)00264-9

Casulli, V., and Zanolli, P. (2005). High-resolution methods for multidimensional advection-diffusion problems in free-surface hydrodynamics. *Ocean Mod.* 10, 137–151. doi: 10.1016/j.ocemod.2004.06.007

Chang, Y.-J., Sun, C.-L., Chen, Y., Yeh, S.-Z., and Dinardo, G. (2012). Habitat suitability analysis and identification of potential fishing grounds for swordfish,

Xiphias gladius, in the South Atlantic Ocean. *Int. J. Remote Sensing* 33, 7523–7541. doi: 10.1080/01431161.2012.685980

Conn, P. B. (2010). Hierarchical analysis of multiple noisy abundance indices. *Can. J. Fish. Aquatic Sci.* 67, 108–120. doi: 10.1139/f09-175

Conn, P. B., Johnson, D. S., and Boveng, P. L. (2015). On extrapolating past the range of observed data when making statistical predictions in ecology. *PLoS One* 10:e0141416. doi: 10.1371/journal.pone.0141416

Craig, J. K., and Crowder, L. B. (2005). Hypoxia-induced habitat shifts and energetic consequences in Atlantic croaker and brown shrimp on the Gulf of Mexico shelf. *Mari. Ecol. Progr. Ser.* 294, 79–94. doi: 10.3354/meps294079

Crear, D. P., Latour, R. J., Friedrichs, M. A. M., St-Laurent, P., and Weng, K. C. (2020a). Sensitivity of a shark nursery habitat to a changing climate. *Mari. Ecol. Progr. Ser.* 652, 123–136. doi: 10.3354/meps13483

Crear, D. P., Watkins, B. E., Saba, V. S., Graves, J. E., Jensen, D. R., Hobday, A. J., et al. (2020b). Contemporary and future distributions of cobia, Rachycentron canadum. *Biodiversity Res.* 26, 1002–1015. doi: 10.1111/ddi.13079

Day, J. W. Jr., Hall, C. A. S., Kemp, W. M., and Yáñez-Arancibia, A. (1989). *Estuarine Ecology*. New York: John Wiley & Sons.

Dedman, S., Officer, R., Clarke, M., Reid, D. G., and Brophy, D. (2017). Gbm.auto: a software tool to simplify spatial modelling and marine protected area planning. *PLoS One* 12:e0188955. doi: 10.1371/journal.pone.0188955

Du, J., and Shen, J. (2014). Decoupling the influence of biological and physical processes on the dissolved oxygen in the Chesapeake Bay. *J. Geophys. Res. Oceans* 120, 78–93. doi: 10.1002/2014jc010422

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* 1, 54–77.

Elith, J., Kearney, M., and Phillips, S. (2010). The art of modelling range-shifting species. *Methods Ecol. Evol.* 1, 330–342. doi: 10.1111/j.2041-210x.2010.00036.x

Elith, J., and Leathwick, J. (2017). *Boosted Regression Trees For Ecological Modeling*. Available online at: https://rspatial.org/raster/sdm/9_sdm_brt.html (accessed September 23, 2021).

Elith, J., and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Ann. Rev. Ecol. Evol. Syst.* 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Fabrizio, M. C., Manderson, J. P., and Pessutti, J. P. (2013). Habitat associations and dispersal of black sea bass from a mid-Atlantic Bight reef. *Mari. Ecol. Progr. Ser.* 482, 241–253. doi: 10.3354/meps10302

França, S., Costa, M. J., and Cabral, H. N. (2009). Assessing habitat specific fish assemblages in estuaries along the Portuguese coast. *Estuarine Coastal Shelf Sci.* 83, 1–12. doi: 10.1016/j.ecss.2009.03.013

Freitag, A., Vogt, B., and Hartley, T. (2018). Ecosystem-based fisheries management in the Chesapeake Bay: developing functional indicators. *Coastal Manag.* 46, 127–147. doi: 10.1080/08920753.2018.1451729

French, K. J., Shackell, N. L., and den Heyer, C. E. (2018). Strong relationship between commercial catch of adult Atlantic halibut (*Hippoglossus hippoglossus*) and availability of suitable habitat for juveniles in the Northwest Atlantic Ocean. *Fish. Bull.* 116, 107–121.

Fulford, R. S., Peterson, M. S., Wu, W., and Grammer, P. O. (2014). An ecological model of the habitat mosaic in estuarine nursery areas: part II – projecting effects of sea level rise on fish production. *Ecol. Mod.* 273, 96–108. doi: 10.1016/j.ecolmodel.2013.10.032

Georgian, S. E., Anderson, O. F., and Rowden, A. A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fish. Res.* 211, 256–274. doi: 10.1016/j.fishres.2018.11.020

Glaspie, C. N., Clouse, M., Huebert, K., Ludsin, S. A., Mason, D. M., Pierson, J. J., et al. (2019). Fish diet shifts associated with the northern Gulf of Mexico hypoxic zone. *Estuaries Coasts* 42, 2170–2183. doi: 10.1007/s12237-019-00626-x

Gray, C. A., Rotherham, D., and Johnson, D. D. (2011). Consistency of temporal and habitat-related differences among assemblages of fish in coastal lagoons. *Estuarine Coastal Shelf Sci.* 95, 401–414. doi: 10.1016/j.ecss.2011.10.010

Guan, L., Chen, Y., and Wilson, J. A. (2016). Evaluating spatio-temporal variability in the habitat quality of Atlantic cod (*Gadus morhua*) in the Gulf of Maine. *Fish. Oceanogr.* 26, 83–96. doi: 10.1111/fog.12188

Hess, G. R., and Bay, J. M. (2000). A regional assessment of windbreak habitat suitability. *Environ. Monit. Assess.* 61, 237–254.

Hilborn, R., Amoroso, R. O., Bogazzi, E., Jensen, O. P., Parma, A. M., Szuwalski, C., et al. (2017). When does fishing forage species affect their predators? *Fishe. Res.* 191, 211–221. doi: 10.1016/j.fishres.2017.01.008

Hinson, K., Friedrichs, M. A. M., St-Laurent, P., Da, F., and Najjar, R. G. (2021). Extent and causes of Chesapeake Bay warming. *J. Am. Water Res. Assoc.* 1–21. doi: 10.1111/1752-1688.12916

Holbrook, S. J., Forrester, G. E., and Schmitt, R. J. (2000). Spatial patterns in abundance of a damselfish reflect availability of suitable habitat. *Oecologia* 122, 109–120. doi: 10.1007/pl00008826

Irby, I. D., Friedrichs, M. A. M., Friedrichs, C. T., Bever, A. J., Hood, R. R., Lanerolle, L. W. J., et al. (2016). Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison. *Biogeosciences* 13, 2011–2028. doi: 10.5194/bg-13-2011-2016

Jaonalison, H., Durand, J.-D., Mahafina, J., Demarcq, H., Teichert, N., and Ponton, D. (2020). Predicting species richness and abundance of tropical post-larval fish using machine learning. *Mari. Ecol. Progr. Ser.* 645, 125–139. doi: 10.3354/meps13385

Jenkins, G. P., Spooner, D., Conron, S., and Morrongiello, J. R. (2015). Differing importance of salinity stratification and freshwater flow for the recruitment of apex species of estuarine fish. *Mari. Ecol. Progr. Ser.* 523, 125–144. doi: 10.3354/meps11147

Kerhin, R. T., Halka, J. P., Wells, D. V., Hennessee, E. L., Blakeslee, P. J., Zoltan, N., et al. (1988). *Chesapeake Bay Earth Science Study (CBESS): Physical Properties of Surficial Sediments, Chesapeake Bay, Maryland (Tabular Data). Baltimore, Maryland: Maryland Geological Survey*. Available online at: http://www.mgs.md.gov/coastal_geology/baysedata.html (accessed September 23, 2021).

Knudby, A., Brenning, A., and LeDrew, E. (2010). New approaches to modelling fish-habitat relationships. *Ecol. Mod.* 221, 503–511. doi: 10.1016/j.ecolmodel.2009.11.008

Kostecki, C., Le Loc'h, F., Roussel, J.-M., Desroy, N., Huteau, D., Riera, P., et al. (2010). Dynamics of an estuarine nursery ground: the spatio-temporal relationship between the river flow and the food web of the juvenile common sole as revealed by stable isotope analysis. *J. Sea Res.* 64, 54–60. doi: 10.1016/j.seares.2009.07.006

Krause, J. R., Hightower, J. E., Buckel, J. A., Turnure, J. T., Grothues, T. M., Manderson, J. P., et al. (2020). Using acoustic telemetry to estimate weakfish survival rates along the U.S. east coast. *Mari. Coastal Fish.* 12, 241–257. doi: 10.1002/mcf2.10095

Kritzer, J. P., Delucia, M., Greene, E., Shumway, C., Topolski, M. F., Thomas-Blate, J., et al. (2016). The importance of benthic habitats for coastal fisheries. *BioScience* 66, 274–284.

Lankford, T. E. Jr., and Targett, T. E. (1994). Suitability of estuarine nursery zones for juvenile weakfish (*Cynoscion regalis*): effects of temperature and salinity on feeding, growth and survival. *Mari. Biol.* 119, 611–620. doi: 10.1007/bf00354325

Latour, R. J., Gartland, J., Bonzek, C. F., and Johnson, R. (2008). The trophic dynamics of summer flounder (*Paralichthys dentatus*) in Chesapeake Bay. *Fish. Bull.* 106, 47–57.

Lauver, C. L., Busby, W. H., and Whistler, J. L. (2002). Testing a GIS model of habitat suitability for a declining grassland bird. *Environ. Manag.* 30, 88–97. doi: 10.1007/s00267-001-2609-z

Layher, W. G., and Maughan, O. E. (1985). Spotted bass habitat evaluation using an unweighted geometric mean to determine HSI values. *Proc. Oklahoma Acad. Sci.* 65, 11–17.

Le Pape, O., Chauvet, F., Mahevas, S., Lazure, P., Guerault, D., and Desaunay, Y. (2003). Quantitative description of habitat suitability for the juvenile common sole in the Bay of Biscay (France) and the contribution of different habitats to the adult population. *J. Sea Res.* 50, 139–149. doi: 10.1016/s1385-1101(03)00059-5

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., and Edinger, E. N. (2015). Spatial scale and geographic context in benthic habitat mapping: review and future directions. *Mari. Ecol. Progr. Ser.* 535, 259–284. doi: 10.3354/meps11378

Leslie, H. M. (2018). Value of ecosystem-based management. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3518–3520.

Little, A. G., Loughland, I., and Seebacher, F. (2020). What do warming waters mean for fish physiology and fisheries? *J. Fish Biol.* 97, 328–340. doi: 10.1111/jfb.14402

MacWilliams, M. L., Bever, A. J., and Foresman, E. (2016). . 3-D simulations of the San Francisco Estuary with subgrid bathymetry to explore long-term trends in salinity distribution and fish abundance. *San Francisco Estuary Watershed Sci.* 14, 1–24.

Manderson, J. P., Palamara, L., Kohut, J., and Oliver, M. J. (2011). Ocean observatory data are useful for regional habitat modeling of species with different vertical habitat preferences. *Mari. Ecol. Progr. Ser.* 438, 1–17. doi: 10.3354/meps09308

Manderson, J. P., Pessutti, J., Hilbert, J. G., and Juanes, F. (2004). Shallow water predation risk for a juvenile flatfish (winter flounder; Pseudopleuronectes americanus, Walbaum) in a northwest Atlantic estuary. *J. Exp. Mari. Biol. Ecol.* 304, 137–157. doi: 10.1016/j.jembe.2003.12.004

Maryland Geological Survey (1996). *Baltimore Harbor Surficial Sediments: Texture and Chemistry Baltimore, Maryland (Tabular Data). Maryland Department of Natural Resources, Maryland Geological Survey*. Available online at: http://www.mgs.md.gov/publications/data_pages/baltoharbdata.html (accessed September 23, 2021).

Miller, P. J., Lubke, G. H., McArtor, D. B., and Bergeman, C. S. (2016). Finding structure in data using multivariate tree boosting. *Psychol. Methods* 21, 583–602. doi: 10.1037/met0000087

Minello, T. J., Zimmerman, R. J., and Martinez, E. X. (1989). Mortality of young brown shrimp Penaeus aztecus in estuarine nurseries. *Trans. Am. Fish. Soc.* 118, 693–708. doi: 10.1577/1548-8659(1989)118<0693:moybsp>2.3.co;2

Moncure, R., and Nichols, M. (1968). *Characteristics of Sediments in the James River Estuary, Virginia. Special Scientific Report No. 53*.

Nepal, V., and Fabrizio, M. C. (2020). Sublethal effects of salinity and temperature on non-native blue catfish: Implications for establishment in Atlantic slope drainages. *PLoS One* 15:e0244392. doi: 10.1371/journal.pone.0244392

Overton, A. S., Manooch, C. S. III, Smith, J. W., and Brennan, K. (2008). Interactions between adult migratory striped bass (*Morone saxatilis*) and their prey during winter off the Virginia and North Carolina Atlantic coast from 1994 - 2007. *Fish. Bull.* 106, 174–182.

Parsons, D. M., Middleton, C., Smith, M. D., and Cole, R. G. (2014). The influence of habitat availability on juvenile fish abundance in a northeastern New Zealand estuary. *New Zealand J. Mari. Fresh. Res.* 48, 216–228. doi: 10.1080/00288330.2013.875927

Pennino, M. G., Coll, M., Albo-Puigserver, M., Fernández-Corredor, E., Steenbeek, J., Giráldez, A., et al. (2020). Current and future influence of environmental factors on small pelagic fish distributions in the northwestern Mediterranean Sea. *Front. Mari. Sci.* 7:article622. doi: 10.3389/fmars.2020.00622

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reid, J. M., Reid, J. A., Jenkins, C. J., Hastings, M. E., Williams, S. J., and Poppe, L. J. (2005). *usSEABED: Atlantic Coast Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 118, version 1.0*.

Reuchlin-Hugenholtz, E., Shackell, N. L., and Hutchings, J. A. (2016). Spatial reference points for groundfish. *ICES J. Mari. Sci.* 73, 2468–2478. doi: 10.1093/icesjms/fsw123

Rubec, P. J., Santi, C., Ghile, Y., and Chen, X. (2019). Modeling and mapping to assess spatial distributions and population numbers of fish and invertebrate species in the lower Peace River and Charlotte Harbor, Florida. *Mari. Coastal Fish.* 11, 328–350. doi: 10.1002/mcf2.10086

Scales, K. L., Hazen, E. L., Jacox, M. G., Edwards, C. A., Boustany, A. M., Oliver, M. J., et al. (2017). Scale of inference: on the sensitivity of habitat models for wide ranging marine predators to the resolution of environmental data. *Ecography* 40, 210–220. doi: 10.1111/ecog.02272

Scharf, F. S., Manderson, J. P., Fabrizio, M. C., Pessutti, J. P., Rosendale, J. E., Chant, R. J., et al. (2004). Seasonal and interannual patterns of distribution and diet of bluefish within a middle Atlantic Bight estuary in relation to abiotic and biotic factors. *Estuaries* 27, 426–436. doi: 10.1007/bf02803535

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* 1, 103–113. doi: 10.1111/j.2041-210x.2010.00012.x

Schlenger, A. J., North, E. W., Schlag, Z., Li, Y., Secor, D. H., Smith, K. A., et al. (2013). Modeling the influence of hypoxia on the potential habitat of Atlantic sturgeon: a comparison of two methods. *Mari. Ecol. Progr. Ser.* 483, 257–272. doi: 10.3354/meps10248

Skern-Mauritzen, M., Ottersen, G., Handegard, N. O., Huse, G., Dingsør, G. E., Stenseth, N. C., et al. (2016). Ecosystem processes are rarely included in tactical fisheries management. *Fish Fish.* 17, 165–175. doi: 10.1111/faf.12111

Soranno, P. A., Cheruvelil, K. S., Liu, B., Wang, Q., Tan, P.-N., Zhou, J., et al. (2020). Ecological prediction at macroscales using big data: does sampling design matter? *Ecol. Appl.* 30:e02123. doi: 10.1002/eap.2123

Sundblad, G., Bergström, U., Sandström, A., and Eklöv, P. (2014). Nursery habitat availability limits adult stock sizes of predatory coastal fish. *ICES J. Mari. Sci.* 71, 672–680. doi: 10.1093/icesjms/fst056

Tableau, A., Brind'Amour, A., Woillez, M., and Le Bris, H. (2016). Influence of food availability on the spatial distribution of juvenile fish within soft sediment nursery habitats. *J. Sea Res.* 111, 76–87. doi: 10.1016/j.seares.2015.12.004

Tanaka, K., and Chen, Y. (2015). Spatiotemporal variability of suitable habitat for American lobster in Long Island Sound. *J. Shellfish Res.* 34, 531–543. doi: 10.2983/035.034.0238

Theuerkauf, S. J., and Lipcius, R. N. (2016). Quantitative validation of a habitat suitability index for oyster restoration. *Front. Mari. Sci.* 3:64. doi: 10.3389/fmars.2016.00064

Tian, S., Chen, X., Chen, Y., Xu, L., and Dai, X. (2009). Evaluating habitat suitability indices derived from CPUE and fishing effort data for *Ommatrephes bratramii* in the northwestern Pacific Ocean. *Fish. Res.* 95, 181–188. doi: 10.1016/j.fishres.2008.08.012

Tuckey, T. D., and Fabrizio, M. C. (2016). Variability in fish tissue proximate composition is consistent with indirect effects of hypoxia in Chesapeake Bay tributaries. *Mari. Coastal Fish.* 8, 1–15. doi: 10.1080/19425120.2015.1103824

Tuckey, T. D., and Fabrizio, M. C. (2020). Estimating Relative Juvenile Abundance of Ecologically Important Finfish in the Virginia Portion of Chesapeake Bay. Final Report Submitted to Virginia Marine Resources Commission. Gloucester Point, VI: Virginia Institute of Marine Science, William & Mary.

Tyler, R. M., and Targett, T. E. (2007). Juvenile weakfish Cynoscion regalis distribution in relation to diel-cycling dissolved oxygen in an estuarine tributary. *Mari. Ecol. Progr. Ser.* 333, 257–269. doi: 10.3354/meps333257

Velinsky, D., Wade, T. L., Schlekat, C. E., McGee, B. L., and Presley, B. J. (1994). Tidal river sediments in the Washington DC area I. Distribution and sources of trace metals. *Estuaries* 17, 305–320. doi: 10.2307/1352665

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., and Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? *Meta Anal. Ecogr.* 40, 817–828. doi: 10.1111/ecog.02125

Windle, M. J. S., Rose, G. A., Devillers, R., and Fortin, M.-J. (2012). Spatio-temporal variations in invertebrate-cod-environment relationships on the Newfoundland–Labrador Shelf, 1995-2009. *Mari. Ecol. Progr. Ser.* 469, 263–278. doi: 10.3354/meps10026

Woodland, R., Buchheister, A., Latour, R. J., Lozano, C., Houde, E., Sweetman, C. J., et al. (2021). Environmental drivers of forage fishes and benthic invertebrates at multiple spatial scales in a large temperate estuary. *Estuaries Coasts* 44, 921–938. doi: 10.1007/s12237-020-00835-9

Yu, H., Cooper, A. R., and Infante, D. M. (2020). Improving species distribution model predictive accuracy using species abundance: application with boosted regression trees. *Ecol. Mod.* 432:109202. doi: 10.1016/j.ecolmodel.2020.109202

Yu, W., Chen, X., Zhang, Y., and Yi, Q. (2019). Habitat suitability modelling revealing environmental-driven abundance variability and geographical distribution shift of winter-spring cohort of neon flying squid *Ommastrephes bartramii* in the northwest Pacific Ocean. *ICES J. Mari. Sci.* 76, 1722–1735. doi: 10.1093/icesjms/fsz051

Zajac, Z., Stith, B., Bowling, A. C., Lantimm, C. A., and Swain, E. D. (2015). Evaluation of habitat suitability index models by global sensitivity and uncertainty analyses: a case study for submerged aquatic vegetation. *Ecol. Evol.* 5, 2503–2517. doi: 10.1002/ece3.1520

Zeng, X., Tanaka, K. R., Chen, Y., Wang, K., and Zhang, S. (2018). Gillnet data enhance performance of rockfishes habitat suitability index model derived from bottom-trawl survey data: a case study with *Sebasticus marmoratus*. *Fish. Res.* 204, 189–196. doi: 10.1016/j.fishres.2018.02.009

Zhang, H., Ludsin, S. A., Mason, D. M., Adamack, A. T., Brandt, S. B., Zhang, X., et al. (2009). Hypoxia-driven changes in the behavior and spatial distribution of pelagic fish and mesozooplankton in the northern Gulf of Mexico. *J. Exp. Mari. Biol. Ecol.* 381, S80–S91.

Conflict of Interest: AB and MM are employed by Anchor QEA, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fmars.2021.734213

What Have We Lost? Modeling Dam Impacts on American Shad Populations Through Their Native Range

Joseph Zydlewski[1], Daniel S. Stich[2], Samuel Roy[3], Michael Bailey[4], Timothy Sheehan[5] and Kenneth Sprankle[6]*

[1] United States Geological Survey, Maine Cooperative Fish and Wildlife Research Unit, Department of Wildlife, Fisheries, and Conservation, University of Maine, Orono, ME, United States, [2] Biology Department, SUNY College at Oneonta, Oneonta, NY, United States, [3] Senator George J. Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME, United States, [4] U.S. Fish and Wildlife Service, Falls Church, VA, United States, [5] National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, MA, United States, [6] Connecticut River Fish and Wildlife Conservation Office, U.S. Fish and Wildlife Service, Sunderland, MA, United States

American shad (*Alosa sapidissima*) are native to the east coast of North America from the St. Johns River, Florida, to the St. Lawrence River region in Canada. Since the 1800s, dams have reduced access to spawning habitat. To assess the impact of dams, we estimated the historically accessed spawning habitat in coastal rivers (485,618 river segments with 21,113 current dams) based on (i) width, (ii) distance from seawater, and (iii) slope (to exclude natural barriers to migration) combined with local knowledge. Estimated habitat available prior to dam construction (2,752 km^2) was 41% greater than current fully accessible habitat (1,639 km^2). River-specific population models were developed using habitat estimates and latitudinally appropriate life history parameters (e.g., size at age, maturity, iteroparity). Estimated coast-wide annual production potential was 69.1 million spawners compared with a dammed scenario (41.8 million spawners). Even with optimistic fish passage performance assumed for all dams (even if passage is completely absent), the dam-imposed deficit was alleviated by fewer than 3 million spawners. We estimate that in rivers modeled without dams, 98,000 metric tons of marine sourced biomass and nutrients were annually delivered, 60% of which was retained through carcasses, gametes and metabolic waste. Damming is estimated to have reduced this by more than one third. Based on our results, dams represent a significant and acute constraint to the population and, with other human impacts, reduce the fishery potential and ecological services attributed to the species.

Keywords: American shad, *Alosa sapidissima*, diadromous fish, migration, dam, fish passage, marine derived nutrients

OPEN ACCESS

Edited by:
Mark J. Henderson, United States Geological Survey, United States

Reviewed by:
Joshua Raabe, University of Wisconsin–Stevens Point, United States
Andrew M. Fischer, University of Tasmania, Australia

****Correspondence:***
Joseph Zydlewski
josephz@maine.edu

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: *30 June 2021*
Accepted: *20 September 2021*
Published: *25 October 2021*

Citation:
Zydlewski J, Stich DS, Roy S, Bailey M, Sheehan T and Sprankle K (2021) What Have We Lost? Modeling Dam Impacts on American Shad Populations Through Their Native Range. Front. Mar. Sci. 8:734213. doi: 10.3389/fmars.2021.734213

INTRODUCTION

The migration of animals remains one of the most recognizable and ecologically spectacular occurrences in nature. Animals from diverse evolutionary lineages share a behavioral solution to the seasonal and ephemeral nature of habitat suitability for different phases of their life histories (Dingle and Drake, 2007). These may occur across wide expanses of aerial, terrestrial

and aquatic habitat, and the flux of these organisms spans and inextricably links disparate and sometimes distant ecosystems (Weaver et al., 2018). Organisms transport biomass and nutrients (Kelt and Van Vuren, 2001), sometimes over great distance and large geographic scales. The timing of movements, the food-web function and life-history strategies may all influence ecological significance (Rosenzweig, 1971). These phenomena fundamentally shift interlinked systems from marine to inland ecosystems with annual regularity (Doughty et al., 2016).

Migration in sea-run or diadromous fishes is illustrative of seasonal movements that links the ocean with fresh water habitat. The global patterns of diadromy have been well described (McDowall, 1987) and the general trends have been linked to productivity differences between inland and marine habitats (Gross, 1987) in complex ways (Dodson et al., 2009). These fish species have been important to humans prior to colonization of North America, through to the present day.

Wilcove (2010) describes the four great threats to migratory animals, each of which is a result of human population growth: habitat destruction, overexploitation, climate change and barriers to migration. In coastal regions, all of these threats are evident, but impoundments represent a conservation challenge that results from both complementary and contradictory socioeconomic tradeoffs that directly influence fish populations (Roy et al., 2018; Song et al., 2019; Roy et al., 2020). Migratory fish have declined, somewhat predictably, through the loss of connectivity to habitat critical for the expression of their life history. Damming fundamentally alters the longitudinal connectivity of freshwater ecosystems, particularly for anadromous fish (Hall et al., 2011; Liermann et al., 2012), a fact that has been effectively revealed through the removal of dams in coastal systems (e.g., Watson et al., 2018; Wippelhauser, 2021).

American shad (*Alosa sapidissima*) are native to the east coast of North America and demonstrate the human toll on migratory organisms. This anadromous clupeid has an extensive native range, from the St. Johns River, Florida, to the St. Lawrence River region in Canada. Adults must enter freshwater to spawn as a critical part of their life history (Zydlewski and McCormick, 1997b; Zydlewski and Wilkie, 2012). While some populations have flourished outside of their native range (e.g., the Columbia River; Petersen et al., 2003), native populations have been, and remain, depleted (Atlantic States Marine Fisheries Commission (ASMFC), 2020). Since the late 1800s, the four great threats to migratory animals have driven precipitous population declines range wide (Bilkovic et al., 2002; Limburg and Waldman, 2009; Hasselman and Limburg, 2012) and remain continuing and persistent threats to this fish (Burdick, 1954; Talbot, 1954; Bradley, 1959; Chittenden, 1969).

The development of American shad fisheries in the 1700s grew to a pattern of over exploitation by the late 19th century that, combined with habitat loss into the 20th century, resulted in a functional collapse of fisheries and fishery closures in the late 20th and early 21st centuries. Historically abundant, this species supported commercial fisheries with coast-wide landings that exceeded 20,000 metric tons (MT) in the late 1890s (Walburg and Nichols, 1967; Hightower et al., 1996; Limburg et al., 2003, Atlantic States Marine Fisheries Commission (ASMFC), 2007). The ability to access upriver habitat permitted great population potential in river systems and provided sustainable human value (Limburg et al., 2003). Early settlers of Cooperstown, New York, were said to have avoided starvation by American shad reaching Otsego Lake, the Susquehanna River's source, more than 1,000 km from the coast (Taylor, 1995). In the Gulf of Saint Lawrence, American shad reached the Ottawa River, a distance of over 1,100 km (Provost, 1987). Modern fishery impacts are mollified by stricter harvest regulations such as moratoria (Olney and Hoenig, 2001; Atlantic States Marine Fisheries Commission (ASMFC), 2007) but much of the historic human value (economic, recreational and cultural) for this species has been lost with its absence or reduction in coastal rivers due to persistent anthropogenic influences.

Most notably, dam construction greatly restricted access of American shad to spawning and rearing habitat, directly limiting the scope for population and spawner abundance (Rulifson, 1994; Limburg et al., 2003). Dams may provide critical societal functions to meet needs for electricity, water supply, and flood control (Roy et al., 2018) but these dams, their operations, and impoundments may also conflict with fish conservation goals (Song et al., 2019). Fish passage is often a requested and implemented strategy through the Federal Energy Regulatory Commission (FERC) in the United States, but implementation levels can vary widely for these multi decade permits (Vogel and Jansujwicz, 2021).

Latitudinal differences in the life histories of American shad have shaped the influence imposed by dams through their range. Early and influential assessments of these clinal variations in life history traits (i.e., Carscadden and Leggett, 1975; Leggett and Carscadden, 1978; Glebe and Leggett, 1981a) remain heavily relied upon by both researchers and managers. American shad are entirely semelparous (and more fecund) in the southern rivers with a lower length at age (Leggett and Carscadden, 1978; Gilligan-Lunda et al., 2021) than in northerly populations. North of the Cape Fear River in North Carolina, some level of repeat spawning is observed, and the proportion increases in rivers to the north. Northern American shad are larger and are observed to spawn in as many as five (Grote et al., 2014; McBride et al., 2016) and even seven seasons (Provost, 1987). The degree to which this clinal life history variation reflects phenotypic plasticity (versus genotypic differences) remains an important management interest.

American shad exhibit spawning fidelity to their natal rivers (Talbot, 1954; Hill, 1959; Nichols, 1966; Carscadden and Leggett, 1975) with moderate levels of straying (Mansueti and Kolb, 1953; Williams and Daborn, 1984; Melvin et al., 1986). While divergence among rivers in Canada is notable (Hasselman et al., 2010), human mediated transfers among rivers are likely causal to diminished apparent divergence among some populations within the United States (Hasselman et al., 2013). Thus, while genetic discrimination at a river level may be most appropriate, categorization of three main eco-regions (Northern Iteroparous [NI], Southern Iteroparous [SI] and Semelparous [SM]) provides a convenient and logistically functional division

of American shad rivers (**Figure 1**; Atlantic States Marine Fisheries Commission (ASMFC), 2020).

Habitat obstruction or inadequate dam passage may influence populations in these three eco-regions differently. For all eco-regions, access to spawning habitat represents a significant outcome of effective upstream fish passage through dams. The survival of both juveniles and adult migrants during downstream migration is important for population dynamics (Castro-Santos and Letcher, 2010; Stich et al., 2019), albeit to varying degrees for each eco-region. In the SM eco-region, downstream adult passage is not required as they will have made their terminal migration. For the SI and NI eco-regions, the importance of downstream passage for adults is of obvious importance. Failure to provide downstream passage results in an ecological trap that truncates the age distribution of the population (Stich et al., 2019). This is important as repeat spawners have increased reproductive potential (higher fecundity) and provide a buffer against poor recruitment years (Carscadden and Leggett, 1975). The progeny of adult spawners that successfully ascend a dam in all eco-regions, also must all pass downstream as juveniles to reach the ocean.

Together, differences in the life histories among eco-regions may differentially influence the delivery of biomass and nutrients to a river. Semelparous Pacific Salmon (*Oncorhynchus* spp.; Cederholm et al., 1999; Naiman et al., 2002; and Schindler et al., 2003) and sea lamprey (*Petromyzon marinus;* Weaver et al., 2016) deliver significant nutrients subsidies through the decay of carcasses. With northern populations displaying lower spawning mortality rates and a greater instance of repeat spawning than southern populations, nutrient delivery through excretion and spawning (gametes) may be most important. American shad generally do not feed upon entering rivers, but their mass loss during the run can be considerable (Glebe and Leggett, 1981a,b; Leonard and McCormick, 1999; Walter and Olney, 2003). Regardless of the source, the level to which the historical delivery of marine nutrients, and the seaward return of nutrients by juveniles, has shifted because of dams.

Conservation efforts on the Gulf of Maine's Penobscot River resulted in the removal of two main-stem dams (Opperman et al., 2011) and spurred interest in understanding the restoration potential of American shad for this river which historically had supported an abundant population. Annual landings of over 2 million adults were reported in the 1860s (Foster and Atkins, 1867) and American shad had been a critical fishery for the Penobscot Nation. Trinko Lake et al. (2012) defined the spatial extent of American shad habitat in this system and characterized connectivity. The development of a simple population model provided a means of assessing the influence of dam removals on the production potential of a newly opened river (Bailey and Zydlewski, 2013) but fell short in providing a tool that incorporated fish passage. Stich et al. (2019) developed a generalizable model platform that could be applied to any river system based on management needs. This approach allowed for the assessment of upstream and downstream passage on river-specific productivity. The importance of understanding fish passage in light of both upstream and downstream efficiencies and survival was made evident. This approach was then applied to a series of rivers, range wide, based on informed habitat assessments (Gilligan-Lunda et al., 2021). Similar work by Barber et al. (2018) used this population modeling approach to characterize the delivery of nutrient by another alosine, the alewife (*Alosa pseudoharengus*), by employing estimates of nutrient loss, delivery, and exodus.

In this paper we aspire to implement these complementary approaches of spatial analysis, river specific population modeling and assessment of nutrient dynamics to address the lament of Limburg et al. (2003):

> *We can only imagine today what full Atlantic coastal ecosystems (rivers, estuaries, and coastal marine areas) looked like, but one thing is clear, shad played a far more important role then than they do today. To plan for a sustainable future for American shad, we should reenvision those systems and strive to balance fisheries demands with their ecological function. To do this will require better modeling, better data, and above all, renewed commitment.*

To estimate the opportunity cost realized by this species through dam construction throughout its range, we sought to first estimate habitat that was historically and currently exploited by American shad for spawning in Atlantic coastal rivers. This was accomplished by characterizing 485,618 river reaches and 21,113 dams. We identified potential spawning habitat based on criteria of (i) river width, (ii) distance from seawater intrusion and (iii) slope (to exclude natural barriers to migration) combined with local knowledge. The areas of potential spawning habitat were aggregated to estimate historic habitat available prior to the construction of dams and impoundments for each coastal river. Each river was then assessed using a life history-based population model incorporating latitudinally appropriate life history parameters (e.g., clines in size at age, maturity rates, iteroparity, and maximum age). In aggregate this approach allowed a direct assessment of the theoretical spawning potential lost coast-wide to the construction of dams and allowed us to estimate the historic and current capacities for biomass and nutrient delivery.

METHODS

Characterization of American Shad Habitat Through Their Native Range

We estimated American shad habitat based on available knowledge of habitat extent, area, and accessibility. These data were collected for the entire historic geographic extent of American shad, spanning eastern United States and Canada using a two-step approach. First, we used the United States National Hydrography Dataset (USGS, 2019) and Canadian National Hydrographic Network (Natural Resources Canada, 2019) to determine the potential freshwater networks available for migration, spawning, and rearing. The data were organized as a series of flowline segments, representing interconnected stream and river reach segments (**Figure 2**). To simplify this analysis, we assumed initially that American shad would not migrate to reach segments with a mean channel width of less than 15 m, in accordance with pre-existing habitat suitability

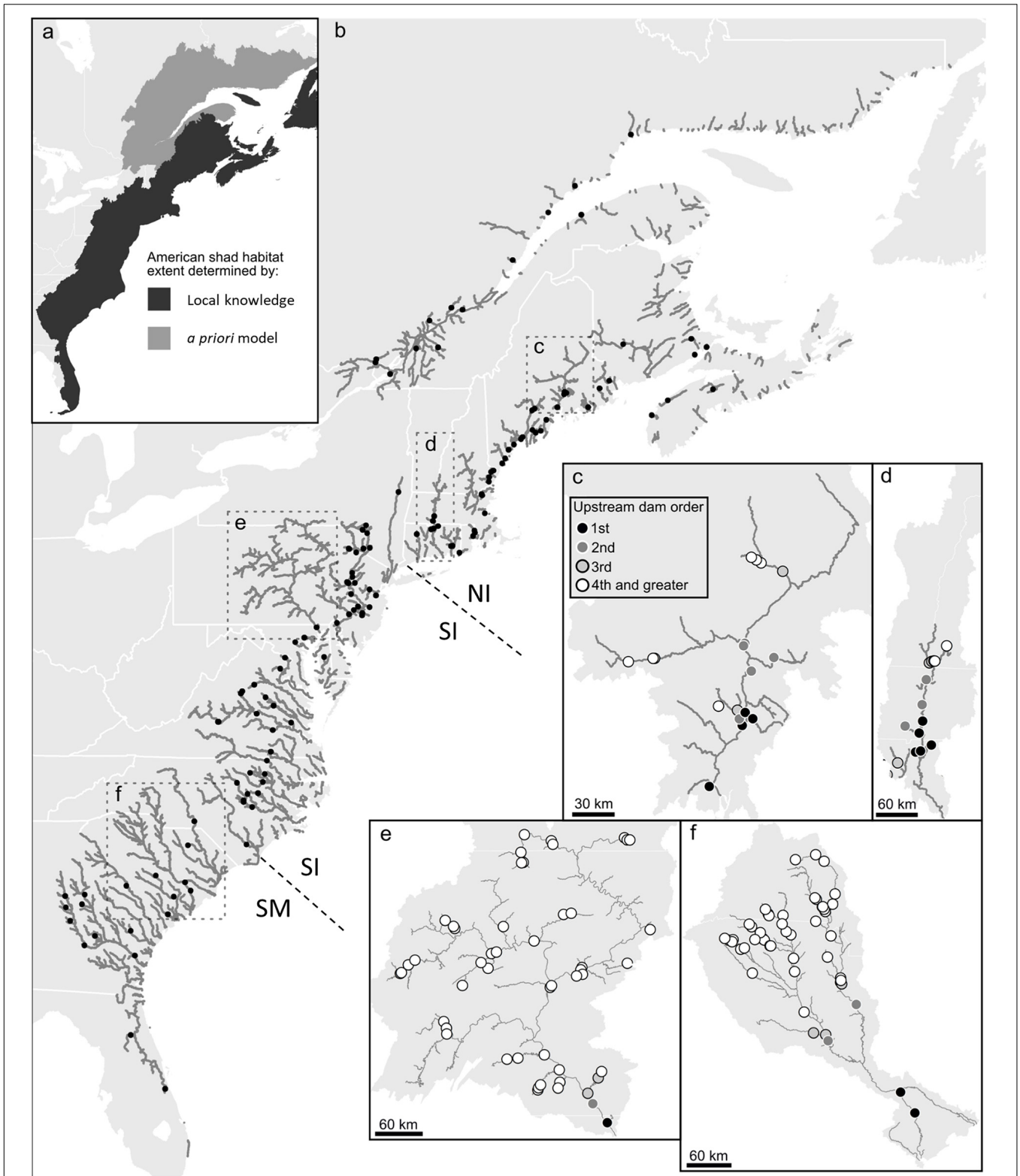

FIGURE 1 | Estimated American shad habitat through a combination of local knowledge and *a priori* modeling along the eastern coast of North America **(a)**, the full extent of their historic range. In the full panel **(b)**, gray lines indicate suitable riverine habitat. Only the first dam on each river or tributary is shown (black circles) for clarity. Dashed lines indicate divisions between northern iteroparous (NI), southern iteroparous (SI) and semelparous SM) eco-regions Insets provide greater detail for four rivers; Penobscot River **(c)**, Connecticut River **(d)**, Susquehanna River **(e)**, Cooper River and **(f)** the Santee River, indicating the numerous impoundments that constrain the migration of American shad in these watersheds.

models (Stier and Crance, 1985; Harris and Hightower, 2012). Second, we validated our *a priori* assessment of historic habitat extent of American shad with the help of local experts from each state or province in our study region (see section "ACKNOWLEDGMENTS"). The experts modified historic extents based on the presence of natural barriers (e.g., steep rapids, waterfalls) or environmental conditions (e.g., temperature, salinity, and reach segment width) that are unsuitable for American shad spawning.

Empirically determined stream discharge-width relationships were used to calculate potential habitat area from the flowline data (e.g., Leopold and Maddock, 1953). This approach allowed for the estimation of horizontal surface area at any reach segment based on drainage area, geographic location, and nearby stream gage data. We first used the enhanced unit runoff method (*EROM*; McKay et al., 2012) to estimate mean annual discharge, then estimated mean reach segment width (w) using the power law equation:

$$w = kQ^b \quad (1)$$

where Q is discharge, k was a derived width coefficient, and b was a derived exponent. Values for k and b vary by region but are typically close to 10 and 0.5, respectively (Bray, 1982; Sweet and Geratz, 2003; Dudley, 2004; Mohamoud and Parmar, 2006; Bent and Waite, 2013). Horizontal surface area was then calculated as:

$$A = 0.8 \times wl \quad (2)$$

where l is segment reach length. We assumed that fluctuations in discharge cause 20% of this area to be periodically dry along the shorelines and therefore inadequate for migration and spawning (*sensu* NOAA, 2009). This approach allowed us to specifically exclude dam impoundments from any analyses. Additionally we excluded lake and pond areas in the watershed assessments as American shad avoid lacustrine habitat (Stier and Crance, 1985).

We then calculated cumulative habitat areas segmented by the presence of dams by combining our habitat flowline data with a congruent dam geodatabase compiled from multiple sources (Martin and Apse, 2011; Martin, 2013, 2019; Natural Resources Canada, 2019). We summed habitat area for all reach segments upstream of each dam point, in addition to each coastal outlet point of streams and rivers (**Figure 1**). These point data were also vetted by local experts. Sums were taken iteratively by starting at the reach segment where the point was located, then adding upstream neighbor reaches until upstream dam points or headwater points were found. This search was recursive to avoid summing overestimations due to downstream flow bifurcations (**Figure 2**). Finally, we tagged individual streams and rivers with common names used by the Atlantic States Marine Fisheries Commission (Atlantic States Marine Fisheries Commission (ASMFC), 2007) for United States rivers and best available names for Canadian Rivers (Natural Resources Canada, 2019). In the end, this analysis produced a range wide assessment of putative American shad habitat as influenced by each impoundment found within the identified area. Data are compiled for the three eco-regions based on regional spawning strategies (Northern Iteroparous [NI], Southern Iteroparous [SI], and Semelparous [SM]) as delineated by Hasselman et al. (2013) and adopted by Atlantic States Marine Fisheries Commission (ASMFC) (2020).

Population Model

In order to assess the theoretical impact of dams on American shad range wide, we simulated the population potential for each of 164 identified systems. This approach allowed us to compare three broad scale scenarios (i) historical or intact "*Undammed*" rivers, (ii) "*Dammed*", with current dams in place with no fish passage, and (iii) dams with favorable upstream and downstream passage to reflect the "*Current*" condition (**Figure 3**). All modeling routines were implemented in the '*anadrofish*' package (Stich et al., 2020) for R (R Core Team, 2019) to provide a user interface for reproducibility and further exploration.

The river specific population modeling effort was based on Stich et al. (2019), who applied a stochastic life-history based simulation model to assess the theoretical effects of dam passage and migratory delay on abundance, population demographics, and spatial distribution of spawning adults over time. However, we followed Harris and Hightower (2012) and Bailey and Zydlewski (2013) in adopting an age-structured (rather than individual) approach to migration dynamics to generalize the modeling framework across the known range for American shad. This general model structure facilitated the incorporation of geographically appropriate life history parameters. Because these fish are iteroparous in the northern extent of their range and semelparous in the southern extent, downstream migration of both juveniles and adults was considered important for coast-wide population dynamics. Each river was identified as being in the NI, SI or SM region and assigned life history parameters estimated for each eco-region (Atlantic States Marine Fisheries Commission (ASMFC), 2020). As this population model has been described previously, the description and equations are included as **Supplementary Material 1** and available as open source code[1].

Dam passage probabilities were assigned based on each of the passage scenarios. For the Undammed scenario, the dams exerted no influence on passage (upstream and downstream passage probabilities were 1.00), whereas for the Dammed scenario, upstream passage probability was set to zero. For the Current scenario, upstream passage was assigned based on available estimates (**Table 1**), assuming a passage probability of 0.40 (the unweighted mean of reported values). Separate downstream survival probabilities were used for adults (S_{DA} = 0.80) and juveniles (S_{DJ} = 0.95) and were based on available estimates or recent modeling efforts (e.g., USFWS, 2019). These values reflect reportedly "excellent" upstream passage rates for American shad (Haro and Castro-Santos, 2012) and are intended to represent the most optimistic assessment of American shad passage at all dams (even when there are fish ways). Downstream values are averages of reported values from sources in **Table 1**.

For each scenario, upstream and downstream passage probabilities were fixed across all dams. In all cases, upstream passage probabilities restricted the number of adults reaching spawning areas while downstream passage rates were applied as cumulative, catchment-wide mortality risks for juveniles and

[1]https://github.com/danStich/anadrofish

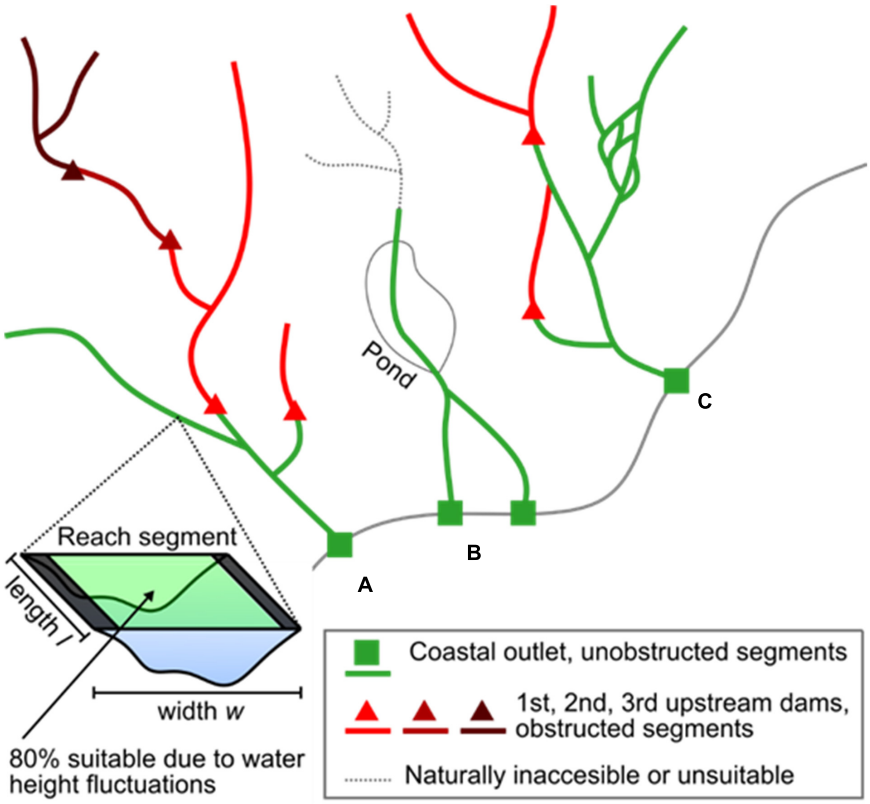

FIGURE 2 | Example of our approach to habitat area calculation. **(A)** Habitat was calculated for all reach segments. We then summed habitat groups that occurred between point features (e.g., habitat between a coastal outlet point and the first upstream dams in that river, habitat between a dam and other dams immediately upstream). **(B)** Rivers that bifurcated along the coast were summed as a single habitat group due to their upstream connection. Reach segments deemed inaccessible or unsuitable by experts were excluded from the habitat calculation, in addition to pond and lake areas. **(C)** We included downstream bifurcations in our habitat group calculations. Dams located on a bifurcation obstructed habitat only to the end of the bifurcation when other dominant flow paths were unobstructed.

adults based on their locations above dams. The probability of fish reaching each habitat segment (P_{Access}) was calculated as the cumulative product of P_{UP} through the number of dams ($ORDER$) downstream of that segment ($P_{UP}^{ORDER_i}$). The amount of available habitat (km^2) in each i^{th} habitat segment (H_S) was pro-rated based on this product, and the resulting "functional" habitat (H_F) was summed throughout the catchment to yield over n habitat units an estimate of total habitat (H_T):

$$H_T = \sum_{i=1}^{n} H_{F_i}, H_{F_i} = H_{S_i} \times P_{UP}^{ORDER_i} \quad (3)$$

The available habitat for returning adults and the simulated population of spawning fish were used to develop a river- and scenario-specific life history-based population model. Survival of adult fish to spawn was randomly drawn for each simulation. We used a Beta distribution (a = 90 and b = 10) to achieve a distribution of pre-spawn survival (mean of 0.90, standard deviation of 0.025, constrained on the interval [0, 1]). Based on the scenario-specific habitat (area) available in each river, we applied carrying capacity (k) to catchment-wide larval production (r) from a vector of age-structured spawners (s) using a Beverton and Holt (1957) recruitment curve with density dependence and a multiplicative error structure:

$$\log_e r = \log_e \left(\frac{a \times s}{1 + b' \times s} \right) \quad (4)$$

The density dependent parameter (b') was tuned to impose a k of 24,711 adult fish per square kilometer (100 adult fish/acre), an often used production potential for stock assessment (Atlantic States Marine Fisheries Commission (ASMFC), 2020) based on Stevenson (1899). This resulted in b = 0.340297 when mean fecundity was used for the value of a (density-independent parameter, specified as F_t):

$$b' = \frac{b}{H_T \times \frac{S}{\sum_{t=1}^{t_{\max}} N_{S_t}}} \quad (5)$$

Because H_T was summed across all units, this approach includes the implicit assumption that spawning fish distribute in the river according to proportional availability of habitat at the reach scale. Likewise, this assumption is also implicit in density-dependent recruitment due to the inclusion of H_T in the denominator of the function describing b' (density dependence in larval recruitment at the catchment level).

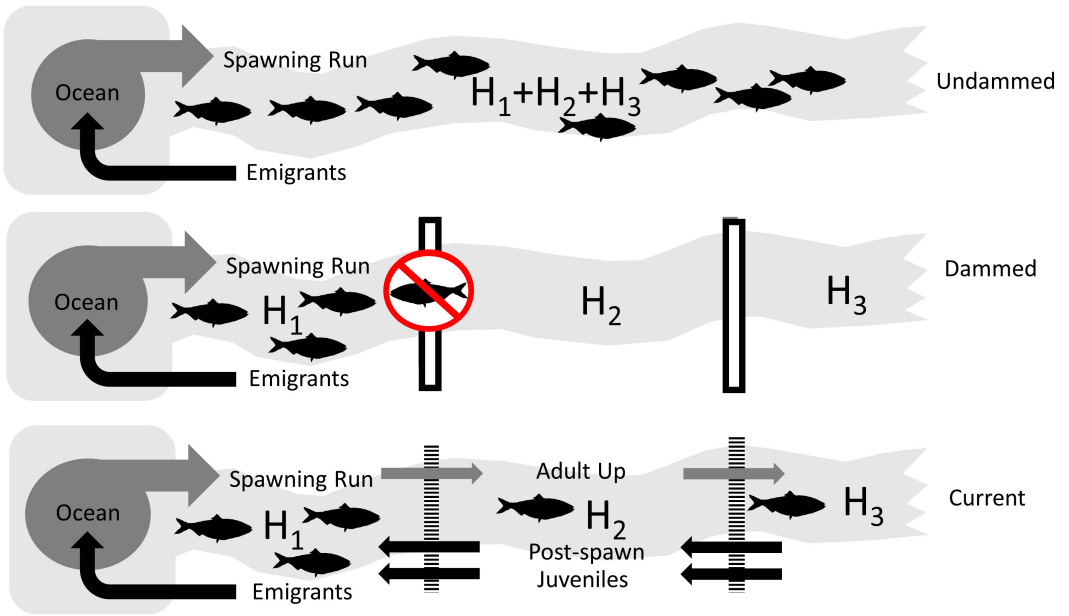

FIGURE 3 | Conceptual diagrams of the three scenarios used to assess the impact of lost habitat and reduced connectivity on the coast-wide production potential of American shad due to dams. The Undammed scenario represents habitat accessibility in the historically undammed rivers. The Dammed scenario depicts the access to habitat that is not impacted by extant impoundments (assuming worst case, non-existing passage). The Current scenario depicts the *status quo*, where the current array of dams is in place, but upstream and downstream passage is applied at levels that represent the "best case" for American shad.

We assigned the larval-to-outmigrant survival (S_J) as a random draw from a normal distribution with a mean of 0.007535 and a standard deviation of 0.04 [70-d survival based on daily mortality rates reported by Crecco et al. (1983)] to estimate juvenile migrants from each river (N_{juv}) for each habitat unit. This quantity was simulated on the $\log_e$-scale and back-transformed to avoid negative values. Adults incurred post-spawn mortality based on the difference between natural mortality (M) and the projected degree of iteroparity as derived from Leggett and Carscadden (1978) and applied by Bailey and Zydlewski (2013). Iteroparity (I) was simulated from river-specific latitude using the equation developed by Leggett and Carscadden (1978):

$$I = \frac{5.08 \times Latitude - 165}{100} \quad (6)$$

We assumed that natural annual survival ($S = 1 - A$) was the joint product of surviving the spawning period (S_{post}) and surviving the duration of the year. We assumed there was no fishing mortality in all systems. Therefore, we simulated post-spawn survival as the quotient of I and S:

$$S_{post} = \frac{I}{S} \quad (7)$$

For populations in which I was predicted to be negative based on the equation from Leggett and Carscadden (1978) we set $I = 0$, and therefore S_{post} was likewise zero (semelparous populations in extreme southern extent). Likewise, if the value of S_{post} was greater than one, we set $S_{post} = 1$ (iteroparous populations in extreme northern extent).

For an assumed 2-month residence time in freshwater (based on Leggett, 1972 [40-100 days] and Chittenden, 1976 [> 60 days]; **Table 1**), survival in freshwater during the spawning run (S_{spawn}) was calculated as the product of two components:

$$S_{spawn} = S_{post} \times S_{2month} \quad (8)$$

Where S_{2month} is 2-month interval survival (S) applied as an instantaneous rate of mortality over 2 months and calculated by using the relationship between annual interval mortality (A) and instantaneous mortality Z_{12} (for a 12-month period; Miranda et al., 2007) to calculate Z for a 2-month interval (Z_2) as:

$$Z_2 = -0.167 \times \ln(1 - A) \quad (9)$$

So that monthly interval survival is calculated as:

$$S_{2month} = e^{-Z_2} \quad (10)$$

For semelparous populations, S_{spawn} is determined by S_{post} (which is set to zero), while for the northern iteroparous populations, in river mortality is determined by natural annual survival (S_{2month}) that is the same survival as for fish remaining in the ocean. Mortalities were imposed at the end of this period and therefore did not influence spawning.

Downstream mortality of juveniles or adults through dams was based on proportional distribution of available habitat as appropriate to each scenario. The total mortality incurred through downstream dam passage was estimated as a function of the imposed downstream survival probability through dams (for both juvenile [S_{DJ}] and adults [S_{DA}]) and the proportional distribution of available habitat with respect to dam order. First,

TABLE 1 | Parameter values for shad population model used to estimate the potential spawning run size for each river in the natural range of American shad.

Parameter name	Value	Description	References
N_{01}	N. Binomial ($\mu = 4 \times 10^5$, $\theta = 10$)	Initial number of age-1 fish	Venables and Ripley (2002)
N_{0t}	$N_{01}l_x$	Age-structured starting population	Derived
N_{St}	Binomial ($p = R_t$, $n = N_{0t}$)	Age-structured spawning adults	Atlantic States Marine Fisheries Commission (ASMFC) (2020)
l_x	$(1 - A)^t$	Lifetime survivorship to age t	Derived
t_{max}	NI = 13 SI = 13 SM = 11	Maximum age by region	Atlantic States Marine Fisheries Commission (ASMFC) (2020)
M	$4.899 \times t_{max}^{-0.916}$	Instantaneous natural mortality	Then et al. (2015)
A	$1 - e^{(-M)}$	Annual mortality	Derived
R_t	NI = (0, 0, 0, 0.04, 0.69, 0.69, 0.9, 1, 1, 1, 1, 1, 1) SI = (0, 0, 0, 0.04, 0.27, 0.64, 0.81, 0.9, 1, 1, 1, 1, 1, 1) SM = (0, 0, 0.01, 0.09, 0.33, 0.63, 0.92, 1, 1)	Recruitment to first spawn by region	Atlantic States Marine Fisheries Commission (ASMFC), 2020; Zydlewski, unpublished
L_t	$L_t = L_\infty \left(1 - e^{-K \cdot (t - t_0)}\right)$	Fork length (mm)	von Bertalanffy (1938)
L_∞, K, t_0	Drawn from correlated region-specific posterior estimates	von Bertalanffy growth parameters	Gilligan-Lunda et al. (2021)
W_t	aL_t^b	Female fish mass (g)	Derived
a	NI = 0.00017929 SI = 0.0000357 SM = 0.00000015600	Intercept of $\log_{10}$length-$\log_{10}$weight relationship for females	Atlantic States Marine Fisheries Commission (ASMFC) (2020)
b	NI = 2.591912 SI = 2.872063 SM = 3.761322	Region-specific slope of $\log_{10}$length-$\log_{10}$weight relationship for females	Atlantic States Marine Fisheries Commission (ASMFC) (2020)
BF_t	$10^{\alpha + \beta \times Wt}$	Batch fecundity	Olney and McBride (2003)
α	NI = 0.239 SI = -0.540 SM = −1.450	Region-specific intercepts for $\log_{10}$mass-$\log_{10}$fecundity relationships	Olney and McBride (2003)
β	NI = 1.39 SI = 1.64 SM = 1.96	Region-specific slopes for $\log_{10}$mass-$\log_{10}$fecundity relationships	Olney and McBride (2003)
PAF	$BF_t \times B_S$	Potential annual fecundity	Hyle et al. (2014); McBride et al. (2016)
B_S	Normal($\mu = 6.1$, $\sigma = 1$)	Number of batches spawned	McBride et al. (2016)
P_{UP}	Dammed = 0.00 Current = 0.40 Undammed = 1.00	Upstream passage probabilities through dams. Dammed and Undammed are defined, Current based on citations.	Weaver et al., 1972; Larinier and Travade, 2002; Sullivan, 2004; Haro and Castro-Santos, 2012; Groux et al., 2015; Castro-Santos et al., 2016
S_{DA}	Dammed = 0.00 Current = 0.80 Undammed = 1.00	Adult downstream survival probabilities through dams. Dammed and Undammed are defined, Current based on citations.	Bell and Kynard, 1985; Hogans and Melvin, 1985; Heisey et al., 1992; Exelon, 2012; Dadswell et al., 2018; USFWS, 2019
S_{DJ}	Dammed = 0.00 Current = 0.95 Undammed = 1.00	Juvenile downstream survival probabilities through dams. Dammed and Undammed are defined, Current based on citations.	Stokesbury and Dadswell, 1991; Heisey et al., 1992; Mathur et al., 1994; Gibson and Meyers, 2002; FirstLight, 2016; TransCanada, 2016
P_{Access}	$P_{UP}^{ORDER_i}$	Cumulative upstream passage probability to habitat segment *i*	Derived
H_{S_i}	km^2	Total habitat in segment *i* of selected river	Derived, see methods
H_{F_i}	$H_{F_i} = H_{S_i} \times P_{UP}^{ORDER_i}$	Functional in habitat segment *i*, pro-rated for cumulative upstream passage probability.	Derived
H_T	$H_T = \sum_{i=1}^{n} H_{F_i}$	Total habitat accessible in selected river	Derived
r	$\log_e r = \log_e \left(\frac{a \times s}{1 + b' \times s}\right)$	Catchment-wide larval recruitment	Beverton and Holt (1957)
a	B_S	Density-independent parameter of Beverton-Holt recruitment curve	Beverton and Holt (1957)
b	0.340297	Density-dependent parameter of Beverton-Holt recruitment curve	Beverton and Holt (1957)

(Continued)

TABLE 1 | (Continued)

Parameter name	Value	Description	Reference
b'	$b' = \frac{b}{H_T \times \frac{S}{\sum_{t=1}^{Tmax} N_{S_t}}}$	Density-dependent parameter of Beverton-Holt recruitment curve tuned to adult k of about 100 fish per acre	Beverton and Holt (1957)
I	$I = \frac{5.08 \times Latitude - 165}{100}$	Iteroparity	Leggett and Carscadden (1978)
S_{post}	$S_{post} = \frac{I}{S}$	Post-spawn survival	Bailey and Zydlewski (2013)
S_{DC_i}	Juveniles = $P_{UP}^{ORDER_i}$ Adults = $P_{UP}^{ORDER_i}$	Cumulative downstream survival probability through dams from the ith habitat segment	Derived
P_{H_i}	$P_{H_i} = \frac{H_{F_i}}{\sum_{i=1}^{n} H_{F_i}}$	Proportion of functional habitat catchment-wide that occurs in the ith habitat segment	Derived
S_{DT}	$S_{DT} = \sum_{i=1}^{n} P_{H_i} \times S_{DC_i}$	Catchment-wide downstream survival for juveniles or adults (calculated separately)	Derived

See section "METHODS" and ***Supplementary Materials (1)*** *for a description of the population model application.*

downstream survival from each i^{th} river segment to the ocean was calculated as the cumulative probability of downstream dam survival (S_{DC_i}) separately for juveniles (S_{DC_J}) and adults (S_{DC_A}) based on the number of dams downstream of each segment. Next, the total proportion of accessible habitat (P_{H_i}) in each i^{th} river segment was calculated as the quotient of prorated habitat in each segment (H_{S-P}) and the sum of all available habitat:

$$P_{H_i} = \frac{H_{S-P_i}}{\sum_{i=1}^{n} H_{S-P_i}} \tag{11}$$

We then calculated the catchment-wide survival during downstream migration for adults,

$$S_{DT_A} = \sum_{i=1}^{n} P_{H_i} \times S_{DC_{A,i}} \tag{12}$$

or juveniles as the weighted sum of S_{DC_i} (sum of the products of P_{H_i} and S_{DC_i}):

$$S_{DT_J} = \sum_{i=1}^{n} P_{H_i} \times S_{DC_{J,i}} \tag{13}$$

Age-structured out-migrants were then added back into the non-spawning (ocean) population, the population was projected one time-step based on instantaneous mortality (M in the absence of fishery impacts or Z otherwise) and the simulation continued for a total of 50 years to ensure stabilization of abundance estimates within rivers. We repeated this simulation for each river and each passage scenario by randomly sampling the river, passage scenario (Dammed, Undammed, Current) and each of the stochastic parameters during each simulation. This resulted in approximately 2,000 simulations per passage scenario per river. For each of the 164 river and scenario combinations, the output for the last year (50) was collected for the number of age specific spawners and juvenile emigrants.

Biomass and Nutrient Delivery

To estimate changes in biomass, total N and total P delivery to the freshwater environment (by adults) and to the marine environment (by juveniles), we used the average outputs from the simulations from each river. The average age specific numbers of spawners returning to each river and passage scenario combination were used to estimate the total spawner biomass seasonally entering each river:

$$Spawner\ Biomass_{River} = \sum_{t=1}^{n} P_M \times N_t \times W_{t,M} + \sum_{t=1}^{n} P_F \times N_t \times W_{t,F} \tag{14}$$

where P is the proportion of male (M) or females (F), N is the average number of spawners of age t and $W_{t,F}$ or $W_{t,M}$ is the sex specific mass at age t. Sex-specific mass for an individual fish was calculated by calculating average length at age and mass as a function of length described in **Supplementary Equations 4, 5**, using region-specific parameters.

Juvenile biomass leaving each river was calculated as the summation of the number of juveniles produced from all habitat units in each river (N_{JUV}) multiplied by the average, weighted survival downstream to the ocean (S_{DT_J}, Equation 13).

$$Juvenile\ Biomass_{River} = N_{JUV} \times S_{DT_J} \times W_{JUV} \tag{15}$$

The total river output ($JuvenileBiomass_{River}$) was simply the sum of juvenile biomass reaching the ocean from each habitat unit. The estimated mass of a juvenile emigrant, W_{JUV}, was based on reported juvenile masses ranging from 2.0 to 4.5 g (Zydlewski and McCormick, 1997a; Haskell, 2018); we applied an average W_{JUV} of 3.36 g (**Table 2**).

The delivery (and retention) of marine derived biomass ($Biomass_{delivered}$) and nutrients into each river during the spawning migration was estimated as a function of: (i) the initial biomass of adults arriving in freshwater ($SpawnerBiomass_{River}$), (ii) the probability of death during spawning, and (iii) if surviving and returning to the ocean (for rivers in the NI and SI regions), the estimated biomass lost during the spawning migration through spawning and metabolic loss.

$$Biomass_{delivered} = (Biomass_{carcass} + Biomass_{metabolic}) \tag{16}$$

TABLE 2 | Parameter used to estimate annual, river specific nutrient input (via adult carcasses, spawning, and metabolism) and nutrient output (juvenile emigration).

Parameter	Unit	Value Used	Reported values and references
Phosphorus (P) male	% (g · g wet mass^{-1})	0.666	Haskell, 2018
Phosphorus (P) female	% (g · g wet mass^{-1})	0.673	Haskell, 2018
Phosphorus (P) juvenile	% (g · g wet mass^{-1})	0.887	Haskell, 2018
Nitrogen (N) male adult pre spawn	% (g · g wet mass^{-1})	2.941	Haskell, 2018
Nitrogen (N) female adult pre spawn	% (g · g wet mass^{-1})	2.958	Haskell, 2018
Nitrogen (N) juvenile	% (g · g wet mass^{-1})	2.803	Haskell, 2018
Juvenile mass	mass (g)	3.36	2.03, 3.54 (Haskell, 2018) 4.5 (Zydlewski and McCormick, 1997a)
Proportional loss in mass during migration (male)	–	0.3584	0.18, 0.44 (Haskell, 2018), 0.09-0.26 (Raabe and Hightower, 2014), 0.45 (Chittenden, 1976), 0.51 (Leggett, 1972)
Proportional loss in mass during migration (female)	–	0.456	0.33, 0.44 (Haskell, 2018), 0.38-0.48 (Raabe and Hightower, 2014), 0.57 (Chittenden, 1976), 0.53 (Leggett, 1972)

$Biomass_{carcass}$ represents the overall biomass of presumptive spawners and successful spawners that die in the river (this includes the contribution from fish carcasses, gametes and metabolism of fish that ultimately die in fresh water). Biomass carcass is calculated as:

$$Biomass_{carcass} = \sum_{t=1}^{n} \left(1 - S_{spawn}\right) \times P_M \times N_t \times W_{M,t} + \sum_{t=1}^{n} \left(1 - S_{spawn}\right) \times P_F \times N_t \times W_{F,t} \quad (17)$$

Where S_{spawn} is probability of survival through the spawning migration (Equation 8), N is the number of spawners at age t, P is the proportion of males (M) or females (F) in a given river, and W is the mass of an individual returning male (or female) adult at age t.

$Biomass_{metabolic}$ is the sum of metabolism expenditures in freshwater and shed biomass in the form of gametes for spawners who successfully enter a river as a presumptive spawner and return to the ocean. This biomass lost in river by surviving fish (gonadal and metabolic) is calculated as:

$$Biomass_{metabolic} = \sum_{t=1}^{n} \left(S_{spawn}\right) \times P_M \times N_t \times W_{M,t} \times \Delta_M + \sum_{t=1}^{n} \left(S_{spawn}\right) \times P_F \times N_t \times W_{F,t} \times \Delta_F \quad (18)$$

All parameters are defined as described in Equation 17. Biomass lost during the spawning run for surviving migrants (Δ) is the proportional sex specific loss of mass due to metabolic expenditure and gametic release. These values are based on empirical data reported by Leggett (1972), Chittenden (1976), Raabe and Hightower (2014), and Haskell (2018) that range 18–51% observed mass loss for males and 33–57% for females. We applied average values of 36 % and 46% mass loss for males (Δ_M) and females (Δ_F) respectively (**Table 2**). We acknowledge this as a simplifying assumption as it is well-known that distance traveled (Leonard and McCormick, 1999), temperature experience (Glebe and Leggett, 1981b; Raabe and Hightower, 2014) and residence time (Raabe and Hightower, 2014) all influence the extent of individual mass lost in freshwater.

Marine derived nutrient transport (phosphorus or nitrogen) was calculated using estimated nutrient density (grams of nutrient per gram of fish wet mass) applied to estimated of $Biomass_{delivered}$. Several simplifying assumptions were applied. We used nutrient density values 0.670 g/g wet mass for phosphorus and 2.950 g/g wet mass for nitrogen (based on averaged male and female data from Haskell, 2018). We assumed nutrient density was the same for males and females, and across tissues (i.e., nutrient density for gonads, metabolic tissue loss and whole-body densities did not differ). Thus, pre-spawn estimates of nutrient density are assumed to be representative of the nutrients shed through gonadal and metabolic loss. This is generally the case as whole body nutrient densities differ by < 1% from male to female (Haskell, 2018). For the gonads, however, we note that the testes are enriched in both N and P (∼10 % over other tissues) while the ovaries are enriched in N (∼10%) but depleted in P (∼50%), resulting in an over estimate for P. However, because female gonadal mass represents approximately 10 % of the mass of a pre-spawn female (Leonard and McCormick, 1999), and the applied mass loss based on observations is more than four-fold greater (46% mass loss during spawning), the influence on phosphorus estimates is minimal.

RESULTS

Habitat

Historic American shad habitat was estimated to be 2,287 km^2 for the entire range and comprised 164 rivers that were identified to have suitable habitat based on the identified criteria. This assessment was based on a combination of *a priori* assumptions and informed by local knowledge through most of the United States and part of Canada. For some of the assessment in Canada, estimates are based solely on the *a priori* model and our inability to locate appropriate experts. The habitat was divided fairly evenly among eco-regions with 30, 38, and 32% identified in the NI, SI and SM eco-regions, respectively (**Figure 4**). In the NI eco-region, 18% of the habitat was found in

Canada while 12% was found in the United States (see also Table in **Supplementary Material 2**).

The presence of dams on 68 rivers throughout the range of American shad fully or partially blocks access to an estimated 41% of their coast-wide habitat (**Figure 4**). The constraint of access to habitat found upstream of at least one barrier is comparably high for all eco-regions, with 37, 44, and 39% loss of habitat due to first main-stem dams for NI, SI and SM, respectively. It is notable that habitat in the Canadian jurisdiction of the NI eco-region has remained comparably intact with an estimated 19% loss. In contrast, the United States NI region is heavily impounded, with a loss of connectivity to 65% of the habitat.

Throughout their range, ten rivers account for more than half (52%) of riverine habitat of American shad (Susquehanna [11.0%], St. Lawrence [8.3%], Altamaha [6.5%], Delaware [4.7%], Savannah [4.3%], James [4.3%], Roanoke [3.4%], Cape Fear [3.0%], Connecticut [3.0%], Congaree [2.9%]; **Figure 5**; Table in **Supplementary Material 2**). Dams on these rivers have resulted in the reduction of habitat access from a minimum of 14% loss in the Delaware River to complete habitat blockage in the Congaree River. Dam blockage at these ten rivers accounts for the loss of 575 km^2 of habitat, nearly 62% of the coast-wide loss in habitat for this species. On the Susquehanna River alone, dams block 243 km^2 of habitat, 10.6% of total historically accessible habitat coast-wide.

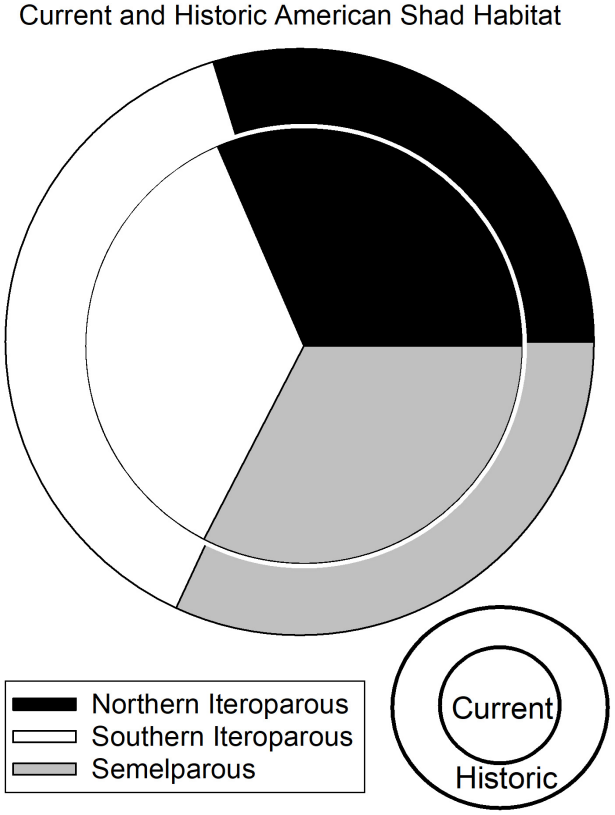

FIGURE 4 | Graphic representation of the Undammed (outer) and Dammed (inner) scenario freshwater spawning habitat available to American shad throughout their native range. The size of the circles is proportional to the area of historic unimpounded habitat (2,752 km^2) and current habitat access (1,639 km^2), indicating that American shad are impeded from reaching nearly 41% of their historic habitat.

Population

Historic American shad population potential was estimated to be 69.1 million adults for the entire range. While historic habitat area is roughly equal for each of the eco-regions, the modeled spawner density is heavily skewed to the NI and SI eco-regions. We calculated an estimated historic potential of 37.5 million (54%) and 27.2 million (39%) spawners for the NI and SI regions, respectively (**Figure 6**). Historical abundance in the Canadian NI ecoregion was estimated at 24.0 million fish (35% of the coastal run) while the United States portion of the NI eco-region was 13.6 million (20% of the coastal run; See also **Supplementary Material 2**). The entire SM was estimated at only 4.3 million spawning fish for the entire region, accounting for only 6% of the spawner potential. When considering the potential for all 164 rivers to support spawning runs of adult American shad, estimates of spawner potential are directly linked to estimates of available habitat (**Figure 7**). However, latitudinal variation in growth, maturation, and post-spawn survival (Leggett and Carscadden, 1978; Gilligan-Lunda et al., 2021) all influence the relative differences observed in population potential. On average, a square km of habitat results in 55,200 spawners in the NI eco-region but only 31,000 in the SI eco-region. Because of the lack of additive age classes in the SM eco-region (i.e., no repeat spawners), a square kilometer of area is modeled to support only 5,900 spawners.

Coast-wide, there is an estimated 39% loss in population potential, directly attributable to lost habitat between the Undammed and Dammed scenarios (**Figure 8**). Regional population potential is diminished by 35, 46, and 40% reductions in abundance for NI, SI and SM regions, respectively. Population potential loss in the Canadian jurisdiction of the NI eco-region is 18% while the United States NI region is reduced by 65%.

Under the Current scenario, where we applied upstream passage probabilities that represent some of the most favorable conditions reported in the literature, there was limited increase in spawning potential of 9% coast-wide from the Dammed scenario (from 41.8 to 44.6 million). Thus, the most optimistic application of dam passage parameters to each of the hundreds of dams that block access to American spawning habitat provides a theoretical benefit of restoring only 4% of the historic spawning potential that is diminished by 39% due to dams. This limited alleviation of population potential influence is lowest in the SI eco-region (2.7% increase), and is comparable in the NI (4.9%) and SM (5.4%). Passage mediated alleviation of population potential in the Canadian jurisdiction of the NI eco-region is 4.1% while the United States NI region is 6.2%, suggesting a greater potential for recovery.

Biomass and Nutrients

Historic American shad populations supported an estimated average of 69.1 million spawners coast-wide, delivering 98,000

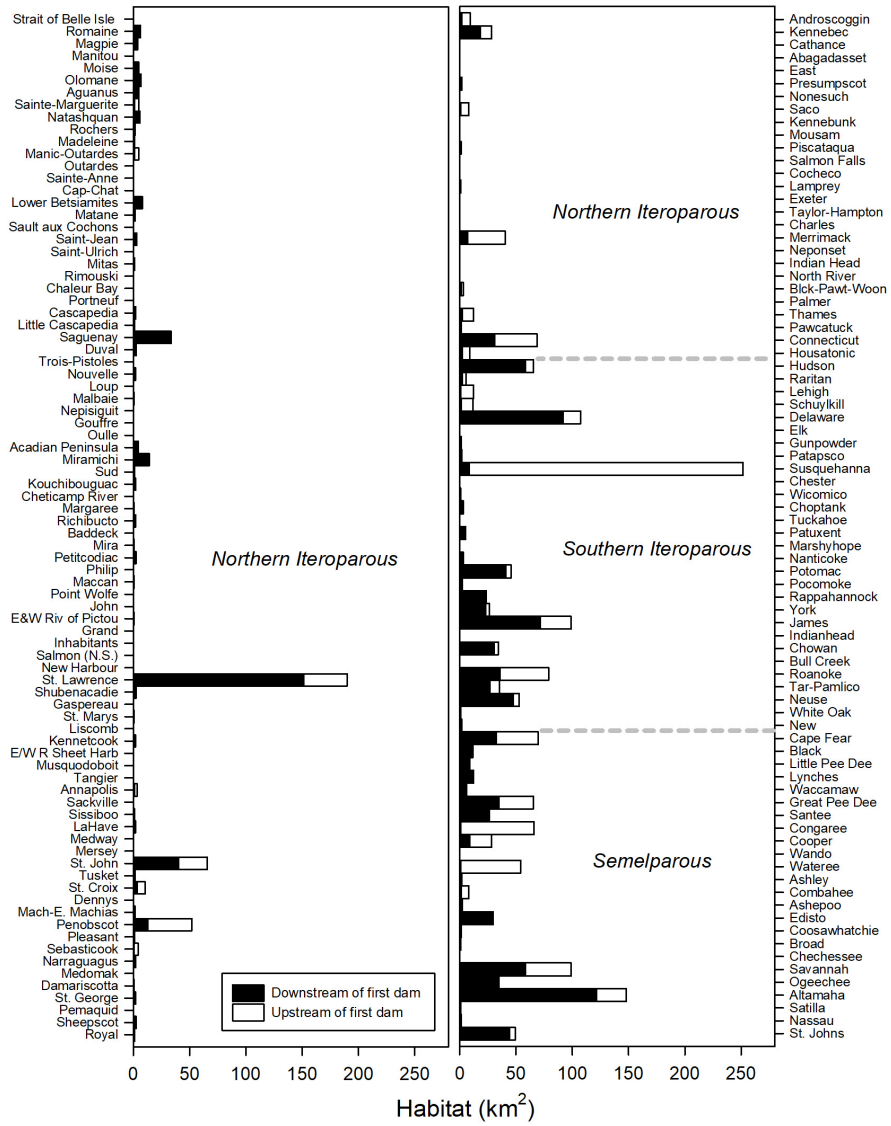

FIGURE 5 | Total habitat available to American shad upstream and downstream of the first dam by river throughout their native range. Dotted lines indicate transitions between eco-regions. Rivers (*y*-axis) are ordered in descending latitude.

MT of biomass to undammed freshwater coastal systems on an annual basis. We estimate that 59,100 MT (more than 60%) of the delivered biomass remained in the river systems in the form of carcasses, tissues shed during spawning, and metabolic loss. Historic biomass delivered to the NI was 51,500 MT (with 25,800 MT retained) and 41,200 MT delivered to the SI eco-region (with 35,500 MT retained; **Figure 9**). For the SM eco-region, a much smaller 5,200 MT biomass was delivered, but all would be retained as part of the semelparous life history in the region. Differences among eco-regions are driven by competing trends of lower potential spawner numbers in the southern latitudes, greater spawner number due to iteroparity in the north and reduced individual spawner sizes (due to a truncated age distribution as repeat spawning declines in the SI and the SM eco-regions).

Under an Undammed scenario, juvenile migration from fresh water transports an estimated 6,600 MT of biomass to the ocean coast-wide, with comparable transport from each of the eco-regions (1,900, 2,900 and 1,800 MT respectively for NI, SI and SM eco-regions). Estimated biomass export to the ocean is considerably lower than the biomass import into freshwater. We estimate 13-fold greater import in the NI eco-region, 12-fold in the SI, but only 3-fold greater in the SM (**Figure 9**).

For the Dammed and the Current scenarios, the relative relations between import and export remain similar among eco-regions and coast-wide, however the magnitudes of both import and export reflect the reduced population under habitat access restriction. Coast-wide, retained biomass is reduced from 59,100 MT for the No Dam scenario to 38,900 MT for the No Passage scenario. Under the Current Scenario, biomass delivered

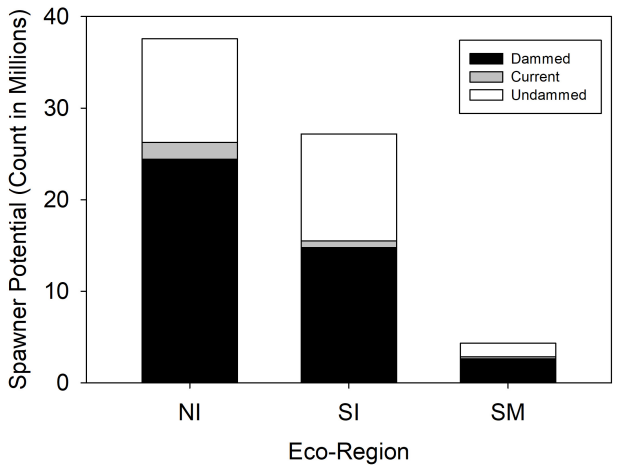

FIGURE 6 | Modeled spawner abundance potential of American shad under the no passage ("Dammed"), *status quo* ("Current"), and no dam ("Undammed") passage scenarios by eco-region through their native range from Canada to Florida.

is slightly increased to 41,400 MT. These represent a 34% and 30% reduction in biomass delivered for the No Passage and Current scenarios respectively, compared to the No Dam scenario. Similar reductions in retained biomass import for the Dammed (37, 45, and 40%) and Current scenarios (32, 42, and 34%) were observed in each of the three eco-regions (NI, SI, and SM). Juvenile nutrient export was reduced from 6,600 to 3,500 MT for the Dammed scenario, and modestly recovered to 3,900 MT under the Current scenario (37 and 42% reductions, respectively).

Because the inputs we used for nitrogen and phosphorus densities are not markedly different between males, females and juveniles (**Table 2**) patterns of nutrient delivery match the trends observed for biomass by eco-region. Under an Undammed scenario, coast-wide retained nitrogen is estimated at 1,963 MT (761, 1,048 and 154 MT for NI, SI and SM) and retained phosphorus is estimated as 446 MT (173, 238 and 35 MT for NI, SI and SM). Under an Undammed scenario, juvenile migration from fresh water exports an estimated 186 MT of nitrogen (54, 82, and 49 MT for NI, SI and SM) and 59 MT of phosphorus (17, 26 and 16 MT for NI, SI and SM). The patterns of reduced nutrient delivery under No Passage and Current Conditions match the trends observed for biomass by eco-region.

Coast-wide, retained N delivery is reduced from 1,963 MT to 1,147 for the Dammed scenario (478, 576, and 93 MT for NI, SI and SM), and modestly recovered to 1,221 MT under the Current scenarios (515, 605, and 102 MT for NI, SI and SM; **Figure 9**). Parallel reduction in retained P delivery is reduced from 446 to 260 MT for the Dammed scenario (108, 131, and 21 MT for NI, SI and SM), and boosted to 277 MT under the Current scenarios (117, 137, and 23 MT for NI, SI and SM; **Figure 9**). Juvenile nitrogen export was reduced from 186 to 110 for the Dammed scenario (34, 46, and 30 for NI, SI and SM), and grew to 119 MT under the Current scenario (38, 48, and 33 for NI, SI and SM). Juvenile phosphorus export was reduced from 59 to 35 MT for the Dammed scenario (11, 14, and 9 for NI, SI and SM), and grew to 37 MT under the Current scenario (12, 15, 10 for NI, SI and SM).

DISCUSSION

Based on *a priori* modeling and local knowledge, the historic habitat availability for American shad spanning the east coast of North America approached 2,300 km^2. While our estimation was derived from 164 coastal rivers, it is likely that American shad may spawn in additional smaller systems, though these would contribute minimally to the overall coastal population. The estimate of 41% loss of habitat connectivity through damming of coastal systems corroborates assertions that loss of spawning habitat access is likely the major cause of population declines in this species (Limburg et al., 2003; Atlantic States Marine Fisheries Commission (ASMFC), 2020). The proportional distribution of accessible habitat is relatively constant among the three eco-regions (when governmental borders are not considered). This indicates that, while impacts are regionally variable, dam construction has negatively influenced habitat access by American shad, similarly, throughout their range.

Because the recruitment function we used incorporates habitat area directly (see Stich et al., 2019), it is unsurprising that theoretical capacity to support runs of adult American shad into rivers mirrors this assumption. Despite the fact that total available habitat was divided fairly evenly among eco-regions, the relative scope for population size is substantially greater in the NI, decreased in the SI and lowest in the SM eco-region due to differences in life history. While this is intuitive based on the degree of iteroparity, growth rates, and maximum sizes (Gilligan-Lunda et al., 2021), it highlights the difference between relative and absolute losses of spawner potential due to dams. The SM eco-region's lost scope of 1.74 million fish is dwarfed by the loss of more than 13 million in the NI eco-region, suggesting that parallel differences in fisheries potential also exist. While the historic spawner production potential is higher in the north and lower to the south, it is notable that each of the regions generally lost comparably high proportions of production habitat. Thus, the fish in southern rivers may be viewed as having the highest conservation vulnerability, particularly at the southern end of the range where climate change may impose additional ecological constraints. Based on this modeling exercise, coast-wide production potential is more than 69.1 million spawners per year without dams compared with the Dammed scenario of just under 41.8 million spawners, a reduction of 39 %.

The intent is of this modeling exercise has been to provide a realistic quantification of the lost potential for American shad that has altered coastal ecosystems and reduced commercial and recreational fisheries opportunities over the species' native range. We paint a clear picture of the relative loss of potential coast-wide, but a cautionary note is warranted. Population estimation is a challenge even when the input parameters are well characterized, and that is not the case here. We are aware that American shad select areas for spawning based on temperature, water velocity, depth and substrate

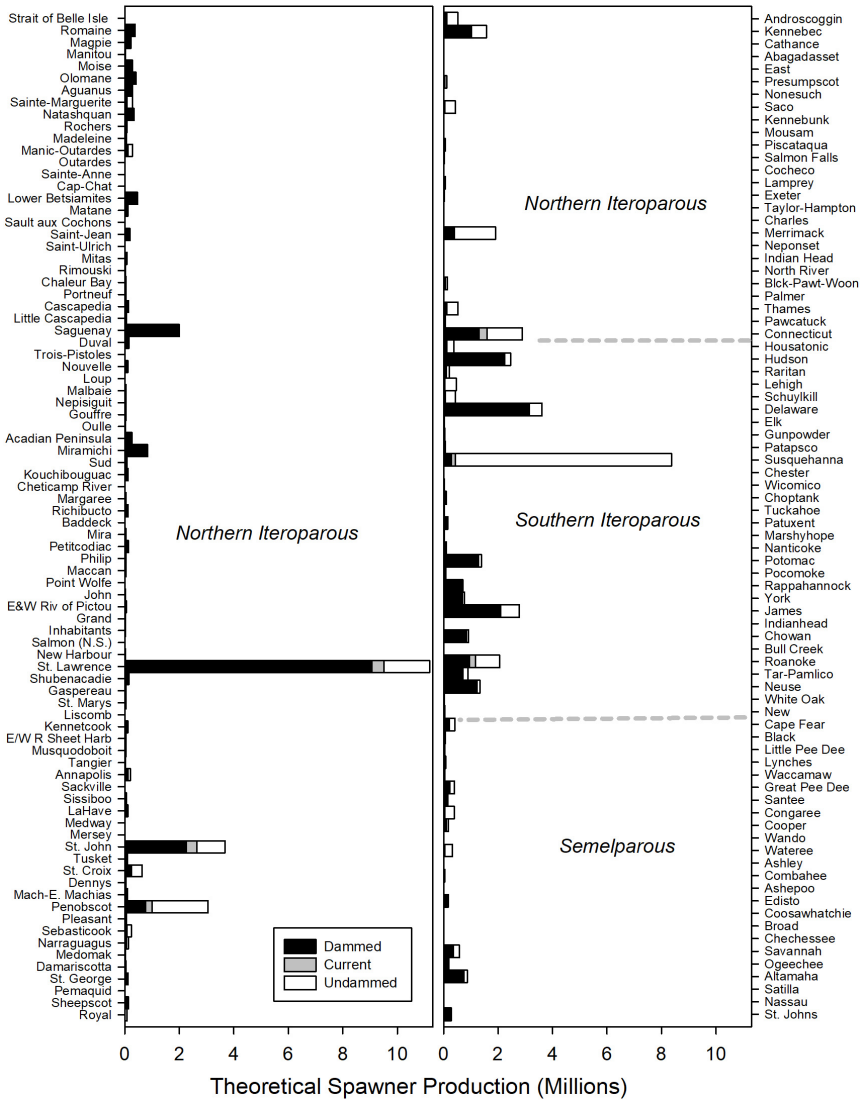

FIGURE 7 | Estimated spawning potential of American shad for each river through their native range from Canada to Florida. Dotted lines indicate transitions between Eco-regions regions. Rivers (*y*-axes) are ordered in descending latitude. The stacked bars depict estimated potential for the No Passage, Current and Undammed scenarios.

(Hightower et al., 2012), evidence that some habitats are more suitable than others. Additionally, juvenile alosine recruitment from reaches of a watershed likely differs based on migratory distance and conditions (Tommasi et al., 2015), yet we have made the implicit assumption of equal habitat quality. Models such as the ones presented here are built with considerable uncertainty, and assumptions in the model parameters. This has been well highlighted here, and elsewhere (Castro-Santos and Letcher, 2010; Bailey and Zydlewski, 2013; Stich et al., 2019). For American shad, there remain many fundamental relations that are poorly characterized (e.g., passage, reproduction, survival and juvenile recruitment), a condition common in such model efforts (Goethel et al., 2011).

It is self-evident that the parameterization of the system we imposed defined the boundaries of our outcomes, and we anticipate that continued research efforts will aid in refining our estimates. However, our estimates are likely to be conservatively low based on model construction, parameter values applied, and assumptions made. Specifically, the Beverton-Holt recruitment curve we have employed is derived from Stevenson (1899), an approach that may grossly under estimate capacity. These cautions withstanding, the values of spawners we generated are realistic in magnitude and represent the best available information. Our estimate of 98,000 MT of spawners returning to all rivers is congruent with historic commercial fisheries landings that exceeded 20,000 MT annually (Walburg and Nichols, 1967; Limburg et al., 2003). Undammed potentials for specific rivers (**Supplementary Material 2**) are generally consistent with historic annual landings harvest from Walburg and Nichols (1967), the data suggest a high fisheries exploitation rate of

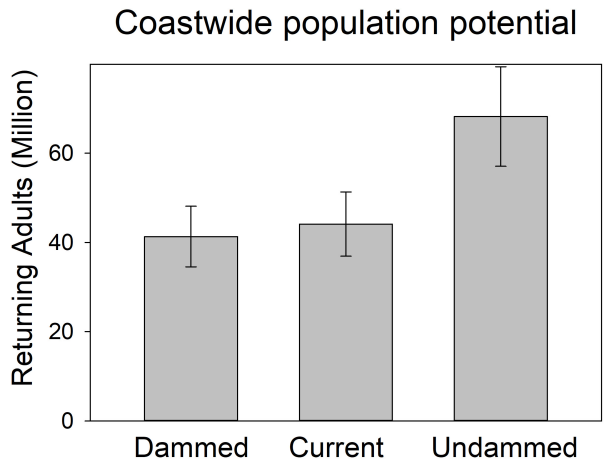

FIGURE 8 | Coast-wide spawning potential of American shad under the no passage ("Dammed"), *status quo* ("Current"), and no dam ("Undammed") passage scenarios. Boxes indicate 25th to 75th interquartile range and whiskers indicate 95% CI.

nearly 20% of the average entire run in the Neuse, James, and Delaware Rivers, 50% in the Potomac and 70% in the York River. A reported harvest of over 2 million adults in the Penobscot River in the 1860s (Foster and Atkins, 1867) would represent a harvest of 65% (of our average estimated run of 3.1 million fish). Harvest data from other rivers indicate that our model may significantly undervalue spawner potential. Reported annual harvest in the Nanticote and Choptank Rivers (Walburg and Nichols, 1967) were two- and three-fold greater than our average undammed population projections. Such difference are undoubtedly due to both population stochasticity and model underestimation. Thus, the most robust value of this exercise is the relative influence of dams which provides an unambiguous relative index of the impact of impaired habitat access.

The inability for current realistic fish passage measures to restore any more than 9% of the estimated spawning potential coast-wide is troubling news for the restoration of these fish. Our results indicate that the application of current upstream and downstream passage rates at all dams affords a remarkably small increase in the theoretical production potential relative to rivers that are wholly inaccessible upstream of the first dam. We estimate that fishway passage coast-wide at dams represents a fixed constraint of about 35% on the spawning run potential of American shad. It is possible that as advances in fish passage engineering, other protective measures, and understanding of fish behavior continue to evolve, passage efficacy may improve beyond our optimistic estimates. The use of fish passage performance criteria (e.g., Stich et al., 2019; CRASC, 2020) may also facilitate fish passage improvements by providing biologically relevant targets. These potential advances withstanding, the low theoretical return on investment of fishways is heavily influenced by the presence of multiple dams on rivers, resulting in a compounding influence on passage and survival.

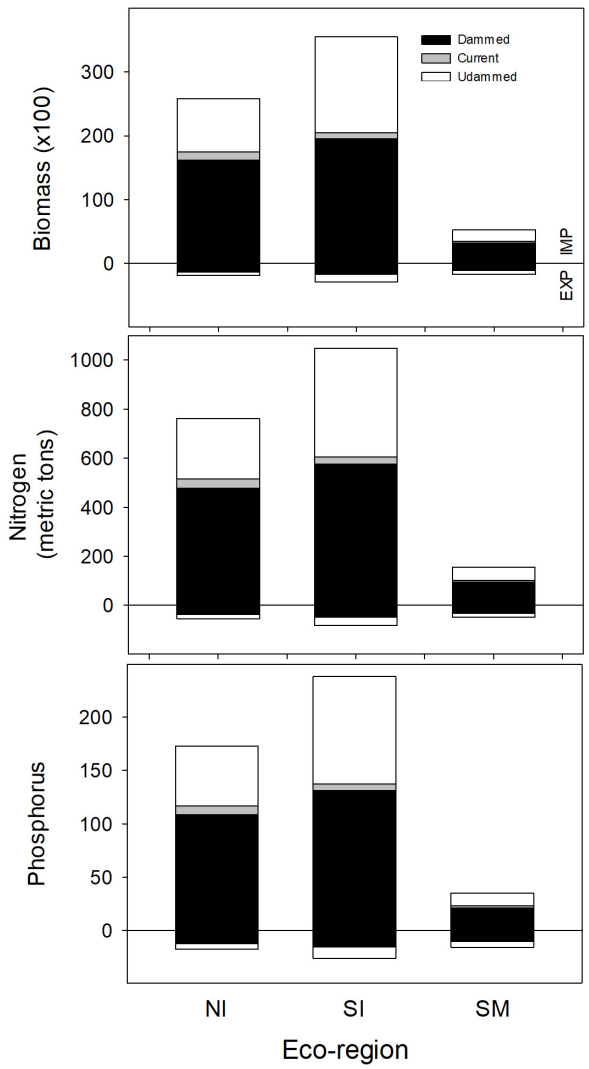

FIGURE 9 | Estimates of annual import (IMP: positive from the zero line) and export (EXP: negative from the zero line) of biomass, nitrogen and phosphorus delivered to each eco-region by American shad through their native range from Canada to Florida under the no passage ("Dammed"), *status quo* ("Current"), and no dam ("Undammed") passage scenarios.

It is notable that this significant imposition on theoretical spawning potential results from a model that estimates contemporary passage values considered to be "excellent" compared to those typically achieved (Haro and Castro-Santos, 2012). True impacts of dams in systems with poor fish passage may be – and are likely to be – greater. Even when significant engineering efforts are made to construct fishways to allow upstream migration, effective connection to upstream habitat can be illusory (Noonan et al., 2012; Bunt et al., 2012, 2016). It is notoriously challenging to provide safe, effective and timely upstream passage past dams for American shad, and the poor passage performance of this species has long been noted (Stevenson, 1899). Challenges associated with dams (beyond passage and survival) have not been incorporated in our

model. Flow disruptions, temperature, and changed community composition could further increase the negative impacts of dams on American shad populations. Loss through coastal commercial and recreational fisheries continues in states where Fisheries Management Plans (or alternatives) are in place (Atlantic States Marine Fisheries Commission (ASMFC), 2020). Lastly, other factors (e.g., habitat degradation) may impose further loss of production potential on river- and region-specific levels.

Fishway design continues to be a critical interface of biology and engineering with significant advances in concepts and application (Silva et al., 2018). However, though fishways have been constructed at some of the numerous dams that many American shad encounter in their migration, many, if not most, are largely or wholly ineffective (Haro and Castro-Santos, 2012). Other dams simply have no specific passage accommodations (e.g., Roanoke River; Hightower and Sparks, 2003), yet for this exercise we attributed optimistic values of fishway performance. Because the time course for hydropower relicensing through the FERC process is 40 years or more (Vogel and Jansujwicz, 2021) the current state of fish passage has a long sustained effect. Managers may only experience one opportunity to shape passage at a dam once during their professional careers. For these reasons, poor utilization of fishways is likely to remain a substantial, if not dominant, culprit for diminishing American shad spawning runs (Atlantic States Marine Fisheries Commission (ASMFC), 2020).

Fish passage is often assessed as a binary event, but even successful dam passage may still exact a toll that is not revealed through modeling. Even when fish passage opportunities are available, dams restrict the timing of access to spawning habitat (e.g., Grote et al., 2014). Dams may therefore increase a migrant's vulnerability to predation (Andrews et al., 2018) or a fishery (Atlantic States Marine Fisheries Commission (ASMFC), 2020), as well as depleting energy stores by imposing delay. The sustained delay below a dam may, however, may preclude successfully passing fish from reaching their reproductive potential (Leonard and McCormick, 1999). Among teleosts, American shad display an elevated basal metabolic rate that is likely causal to their upstream migration being more energetically costly than anadromous salmon (Leonard and McCormick, 1999). Metabolism is determined by water temperature through its effects on enzymatic, metabolic, and cardiac processes (Lennox et al., 2018) so that accelerated energetic depletion occurs in the warmed water near dams (Martin et al., 2015).

Reduction in energy reserves may result in reduced individual fitness (Nadeau et al., 2010) by impairing spawning behaviors and gamete production (Brett, 1962; Rand and Hinch, 1998; Martin et al., 2015). Migrating adults passing a dam do experience an advantage in gaining access to habitat with fewer conspecific competitors. This may afford an advantage in cases where the juvenile rearing carrying capacity is reached, resulting in greater per capita recruitment to the juvenile stage (Walters and Korman, 1999). This advantage, if realized, comes at a significant cost for American shad in the iteroparous eco-regions. In order for spawning adults to spawn again, they must retain enough energy to reach the ocean. Fish must not only reach their spawning habitat in time to spawn, but also must maintain energy stores sufficient to return to the marine environment where they will recommence feeding and growth (Doucett et al., 1999). After spawning, adults leave spawning areas and move into the estuary (Grote et al., 2014) to feed and recondition (Walter and Olney, 2003). Delays at dams are commonly observed in downstream migrating adult shad (Weaver et al., 2019) and undoubtedly exacerbate energy loss at a time when stores are at their lowest. Impediments to the resumption of feeding may therefore increase mortality through prolonged energy loss (Castro-Santos and Letcher, 2010) or reduced performance (Leggett, 1972; Chittenden, 1976; Raabe and Hightower, 2014). Based on comparisons of iteroparous and semelparous populations, it has been suggested that a depletion of greater than 60% of energy reserves may serve as a constraint to an iteroparous life history (Glebe and Leggett, 1981b), while the threshold for post-spawn survival may be as low as 30-40% (Leonard and McCormick, 1999). In addition to a physiological constraint, these fish face high risks of mortality passing dams, particularly if passing through a turbine or other unsafe route (Bell and Kynard, 1985; Hogans and Melvin, 1985; Haro and Castro-Santos, 2012).

The population impact of an impoundment may be amplified by a sequence of dams that must be navigated twice by iteroparous adults to reach spawning habitat, thereby imposing compounding mortality risk on spawners. It is useful to differentiate between the latitudinal cline that is observed in American shad (semelparity in the south to iteroparity in the north) from reduced post-spawn adult survival in the southern and northern iteroparous eco-regions. The degree to which these life history differences reflect population level characteristics versus phenotypic plasticity is poorly characterized. Populations accessing suitable spawning habitat through anthropogenic facilities are subject to an adult mortality akin to an intercept fishery on size and age distributions and results in the systematic loss of "*big old fat female fish*" (Hixon et al., 2014). This has been well characterized in the Connecticut River where the proportion of repeat spawners declined from 49% in the late 1950s (Walburg and Nichols, 1967; Carscadden and Leggett, 1975; Limburg et al., 2003), to a mean of 5 % for the period 2006-2015 (Atlantic States Marine Fisheries Commission (ASMFC), 2020). Failure to recognize the important role of passage has led to the unlikely assertion that upstream passage may be detrimental to a population because of energetic constraints (Leggett et al., 2004) rather than poor downstream passage. In order to repeat spawn, adults in the SI and NI eco-regions need both access to habitat and an effective exit strategy or "*forced semelparity*" may result.

Whether parents are semelparous or iteroparous, the progeny of American shad spawned upstream of dams suffer the risk of mortality when moving downstream. Young American shad generally remain in freshwater until migrating downstream in the fall associated with declining river temperature (Leggett and Whitney, 1972; O'Leary and Kynard, 1986). Downstream passage at dams may entail delay, confinement and turbulence during a period of time when these fish exhibit a heightened sensitivity to stress (Shrimpton et al., 2001) and an environmentally influenced loss of ion regulatory ability in freshwater (Zydlewski and McCormick, 1997b). Juvenile shad, being far smaller than adults, are generally not effectively excluded from turbines. These fish

have less individual risk than entrained adults (Heisey et al., 1992; Mathur et al., 1994) though it is biologically significant. The impacts of dams through injury or delay may shape the disposition of migrants. Their success (or failure) depends on phenology of their physiological development (Zydlewski et al., 2003) in conjunction with biotic and abiotic environmental factors (McCormick et al., 1998; Limburg, 2001).

Prior to the construction of dams range wide, American shad delivered considerable biomass and nutrient loads to freshwater systems (Limburg et al., 2003; **Figure 9**). It is difficult to fully appreciate the ecological significance or magnitude of delivery of 1,963 MT of nitrogen and 446 MT of phosphorus on an annual basis. The seasonal influx driven by the carcasses, gametic release and metabolism would undoubtedly change the freshwater system as has been observed for other species (Weaver et al., 2018). The estimated 34% reduction in nutrient delivery associated with dams is significant. As Haskell (2018) asserts, for American shad, spatial context is of particular importance because shad are broadcast spawners using open water rather than being associated with substrate as is the case for sea lamprey (Weaver et al., 2018) and salmon (Rex and Petticrew, 2008). Spawner distribution in a natural river results in distribution that is punctuated over time and space. Therefore the reduction in spawner potential associated with dams is not only a reduction in the magnitude of nutrient delivered, but a curtailment of the spatial distribution of marine derived nutrient incursions.

In the context of nutrient spiraling theory, in streams and rivers, downstream nutrient changes are driven by delivery, uptake and flushing (Newbold et al., 1982). As such, American shad represent only one broken link in the chain that has historically connected the marine and freshwater ecosystems through the seasons (Limburg et al., 2003). The interruption of ecological connectivity is common for many coastal ecosystems; in the Pacific Northwest nutrient delivery by Pacific salmon is estimated to be only 6-7 % of historical levels (Gresh et al., 2000). The nature of these connections likely changes over the range of the American shad, as anadromous species have a greater scope for influence in freshwater systems with nutrient limitation (nitrogen, phosphorus or co-limited) that influences the degree of nutrient incorporation (Cederholm et al., 1999; Chaloner et al., 2002; Bellmore et al., 2014; Samways and Cunjak, 2015). Increased N and P availability leads to increased primary production, but these effects may be swamped by nutrients generated from anthropogenic land use practices (Twining et al., 2013). The nutrient dynamics of American shad are less comprehensively explored than other alosines (Durbin et al., 1979; Post and Walters, 2009; Walters et al., 2009; West et al., 2010) offering an important direction of future research.

The presence of dams greatly curtails estimated nutrient transport (both import and export) and, in parallel with spawner estimates, fish passage provides a low level of restoration of this ecological function. The dynamics and magnitude of nutrient exchange are strongly influenced by population size and life history differences among the three eco-regions in the natural range. Our models project that in the NI and SI ranges, where iteroparity results in greater spawner potential, delivery is tilted towards import. Haskell (2018) also assessed American shad to be net importers of N and P in the Columbia River where iteroparity levels are 32% (Petersen et al., 2003). In the SM eco-region, despite the higher proportion of carcasses delivered during migration, we estimate that the net flux of nutrients in the southern part of their range to be significantly lower than the northern regions. The nature of delivery among regions differs as well. Carcasses of migrating fish can break down over days and weeks (Garman, 1992; Weaver et al., 2015) so nutrients are not immediately liberated whereas nutrients released via gametes and metabolic processes are more readily available.

Our presentation of average spawner escapement and juvenile transport likely obscures a more nuanced story of nutrient delivery that has been altered by the imposition of dams on the landscape. The complex role of nutrients delivered by alosine species is largely driven by levels of escapement (Barber et al., 2018), allowing locally disparate net nutrient balance. At high levels of spawner escapement, as occurs in our stabilized population models, biomass delivery (and resulting nitrogen and phosphorus delivery) are markedly skewed to import, particularly in the NI and SI regions (**Figure 9**) where spawner numbers are higher (**Figure 6**). In cases where adult dam passage is poor, the juvenile recruits produced per spawner can be expected to be higher than average, resulting in a more equitable exchange, or even a net export (Barber et al., 2018). In such cases, with reduced competition, larger migrating juveniles may further shift the net balance in favor of export (Moore and Schindler, 2004; Moore et al., 2011). Such a localized patchwork of nutrient balance may also occur in natural systems with impediments to migration (e.g., natural falls) or simply due to the attrition of spawners over the length of a river (Meixler et al., 2009; Hall et al., 2011; Pess et al., 2014). Even when a river is producing spawners, failure to connect these fish to their habitat can result in a lower import or net export in nutrient limited systems.

As juvenile American shad develop and grow, they are directly embedded into pelagic, littoral, benthic, and terrestrial systems (Limburg et al., 2003). Feeding – or being preyed upon – directly links these fish to food webs. As is the case for many anadromous species, juvenile American shad (and adults) may be directly consumed through aquatic (Willson and Halupka, 1995; Jaecks and Quinn, 2014), terrestrial and avian predators (Dalton et al., 2009; DeBruyne et al., 2012). Aquatic macroinvertebrates also actively feed on carcasses during their freshwater residency, serving as further conduit of nutrients between the terrestrial and aquatic environments (Polis et al., 1997; Vanni, 2002; Hocking and Reimchen, 2009). Both bottom-up and top-down pathways of nutrient incorporation may result from the influence that American shad likely have in the freshwater ecosystem.

The restriction of American shad to lower reaches of the coastal rivers is part of a fundamental shift in the riverine communities due to impoundment (Kiraly et al., 2015; Watson et al., 2018). Fish communities in these altered systems have both winners and losers, often favoring native and non-native "invaders" (*sensu* Carey et al., 2012) such as black bass (*Micropterus* spp.). By presenting an obstacle to migrations, dams may make American shad and other alosines increasingly vulnerable to novel predators (such blue catfish and

[*Ictalurus furcatus*] and flathead catfish [*Pylodictis olivarisast*] in the Chesapeake Bay region; Schmitt et al., 2017).

In addition to the impacts on the freshwater environment, the dam mediated loss of connectivity also affects estuarine and coastal systems. The presence of adult and juvenile American shad in coastal systems benefit estuarine and marine organisms. The exodus of juvenile shad from freshwater is protracted from the summer into the fall (Williams and Bruger, 1972; O'Leary and Kynard, 1986; Zydlewski and McCormick, 1997b), overlapping with the migration of other alosines. Such migrations attract predators (Davis et al., 2012; McDermott et al., 2015), but at the same time reduce individual risk. Many predatory species depend on the seasonal pulses of prey species (Willson and Womble, 2006; Richardson et al., 2014; Furey and Hinch, 2017). As such, patterns in alosine migration may shape both current and historic distributions of marine fish species (Baird, 1883; Ames, 2004). Specialized detection of relatively high frequency sounds by American shad may also indicate evolutionary predatory pressures, suggesting that changes in abundance may influence the foraging success of echolocating marine mammals (Mann et al., 1998).

We have considered the influence of dams on American shad through their coastal range and used the best available data to quantify the dam-mediated impact on habitat loss, spawner production potential and nutrient transport range wide. The ecological potential and human value of these fish has been markedly reduced through the partial and complete occlusion of access to spawning and juvenile rearing habitat. Historically these fish have linked freshwater systems to the marine environment, from the mangrove estuaries of Florida to the boreal forests of Canada. More than 100 years ago Stevenson (1899) asserted that "*There is no species of fish more important to residents of the Atlantic seaboard than the shad.*" The fisheries, and the human connections to the fish, have diminished in spite of fisheries closures and extensive passage efforts (Atlantic States Marine Fisheries Commission (ASMFC), 2020). The data suggest that dams remain the most significant impediment to restoration of the "founding fish" (McPhee, 2003). While contemporary passage rates fail to achieve even modest population recovery, dam removal appears to remain a viable solution (Raabe and Hightower, 2014; Izzo et al., 2016; Moser and Paradis, 2017; Watson et al., 2018). The recolonization of newly accessible habitat by migratory fish is well-documented (e.g., Burdick and Hightower, 2006; Hogg et al., 2013). While decision-making at dams involves a wide range of stakeholders with diverse and sometimes conflicting objectives (Roy et al., 2018), "active restoration" has been framed as a balanced approach that integrates both values and science (Hart et al., 2002). There is a growing appreciation for the biological and economic benefits of restoring coastal connectivity (Dias et al., 2019), thereby regaining that which we have lost.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JZ led in the framing of the manuscript and compilation of data and led the writing. DS led the population modeling and was a major contributor to writing. SR led the GIS analysis and was a major contributor to writing. All authors contributed to the framing of the manuscript, interpretation and editorial review.

ACKNOWLEDGMENTS

Logistical support was provided by the United States Geological Survey Maine Cooperative Fish and Wildlife Research Unit. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the United States Government. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service. All data generated or analyzed during this study are included in the main text or as **Supplementary Information** for this publication. We thank the Atlantic States Marine Fisheries Commission (ASMFC) and individuals from the ASMFC American Shad Technical Committee and Stock Assessment Subcommittee (SAS) members who contributed data and knowledge that has informed our efforts. We are exceptionally grateful to the dozens of biologists from both the United States and Canada who generously shared their local knowledge of the landscape, habitat use and history of American shad in their region. Only fear of accidental exclusion prevents the authors' attempt at listing the names of those who spoke for the American shad.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.734213/full#supplementary-material

REFERENCES

Ames, E. P. (2004). Atlantic cod stock structure in the Gulf of Maine. *Fisheries* 29, 10–28. doi: 10.1577/1548-8446(2004)29[10:acssit]2.0.co;2

Andrews, S. N., Zelman, K., Ellis, T., Linnansaari, T., and Curry, R. A. (2018). Diet of striped bass and muskellunge downstream of a large hydroelectric dam: a preliminary investigation into suspected Atlantic salmon smolt predation. *North Am. J. Fish. Manag.* 38, 734–746. doi: 10.1002/nafm.10074

Atlantic States Marine Fisheries Commission (ASMFC) (2007). *American Shad Stock Assessment Report for Peer Review: Volume I - Stock Assessment Overview*. Arlington, VA: ASMFC.

Atlantic States Marine Fisheries Commission (ASMFC) (2020). *American shad Benchmark Stock Assessment and Peer Review*. Arlington, VA: ASMFC.

Bailey, M. M., and Zydlewski, J. D. (2013). To stock or not to stock? Assessing the restoration potential of a remnant American Shad spawning run with hatchery supplementation. *North Am. J. Fish. Manag.* 33, 459–467. doi: 10.1080/02755947.2013.763874

Baird, S. (1883). *U.S. Commissioner of Fish and Fisheries Report of 1883*. Washington, DC: NOAA, xi–xiv.

Barber, B. L., Gibson, A. J., O'Malley, A. J., and Zydlewski, J. (2018). Does what goes up also come down? Using a recruitment model to balance alewife nutrient import and export. *Mar. Coast. Fish.* 10, 236–254. doi: 10.1002/mcf2.10021

Bell, C. E., and Kynard, B. (1985). Mortality of adult American shad passing through a 17 megawatt Kaplan turbine at a low-head hydroelectric dam. *North Am. J. Fish. Manag.* 5, 33–38. doi: 10.1577/1548-8659(1985)5<33:moaasp>2.0.co;2

Bellmore, J. R., Fremier, A. K., Mejia, F., and Newsom, M. (2014). The response of stream periphyton to Pacific salmon: using a model to understand the role of environmental context. *Freshw. Biol.* 59, 1437–1451. doi: 10.1111/fwb.12356

Bent, G. C., and Waite, A. M. (2013). *Equations for Estimating Bankfull Channel Geometry and Discharge for Streams in Massachusetts. U.S. Geological Survey Scientific Investigations Report 2013–5155*. Reston, VA: U.S. Geological Survey, doi: 10.3133/ sir20135155

Beverton, R. J. J., and Holt, S. J. (1957). *On the dynamics of exploited fish populations. United Kingdom Ministry of Agriculture and Fisheries. Fisheries Investigations*, Vol. 19. London: H.M. Stationery Office.

Bilkovic, D. M., Olney, J. E., and Hershner, C. H. (2002). Spawning of American shad (*Alosa sapidissima*) and striped bass (Morone saxatilis) in the Mattaponi and Pamunkey Rivers, Virginia. *Fish. Bull.* 100:632.

Bradley, M. C. (1959). *An Ecological Survey Of The Potomac And Anacostia Rivers With Special Emphasis On Pollution*. Washington, DC: Catholic University of America Press.

Bray, D. I. (1982). "Regime equations for gravel-bed rivers," in *Gravel-Bed Rivers*, eds R. D. Hey, J. C. Bathurst and C. R. Thome (Chichester: Wiley), 517–542.

Brett, J. R. (1962). Some considerations in the study of respiratory metabolism in fish, particularly salmon. *J. Fish. Board Can.* 19, 1025–1038. doi: 10.1139/f62-067

Bunt, C. M., Castro-Santos, T., and Haro, A. (2012). Performance of fish passage structures at upstream barriers to migration. *River Res. Applic.* 28, 457–478. doi: 10.1002/rra.1565

Bunt, C. M., Castro-Santos, T., and Haro, A. (2016). Reinforcement and validation of the analyses and conclusions related to fishway evaluation data from Bunt 'performance of fish passage structures at upstream barriers to migration. *River Res. Applic.* 32, 2125–2137. doi: 10.1002/rra.3095

Burdick, G. E. (1954). An analysis of the factors, including pollution, having possible influence on the abundance of shad in the Hudson River. *N. Y. Fish Game J.* 1, 188–205.

Burdick, S. M., and Hightower, J. E. (2006). Distribution of spawning activity by anadromous fishes in an Atlantic slope drainage after removal of a low-head dam. *Trans. Am. Fish. Soc.* 135, 1290–1300. doi: 10.1577/t05-190.1

Carey, M. P., Sanderson, B. L., Barnas, K. A., and Olden, J. D. (2012). Native invaders–challenges for science, management, policy, and society. *Front. Ecol. Environ.* 10:373–381. doi: 10.1890/110060

Carscadden, J. E., and Leggett, W. C. (1975). Meristic differences in spawning populations of American shad, *Alosa sapidissima*: evidence for homing to tributaries in the St. John River, New Brunswick. *J. Fish. Board Can.* 32, 653–660. doi: 10.1139/f75-084

Castro-Santos, T., and Letcher, B. H. (2010). Modeling migratory energetics of Connecticut River American shad (*Alosa sapidissima*): implications for the conservation of an iteroparous anadromous fish. *Can. J. Fish. Aquat. Sci.* 67, 806–830. doi: 10.1139/f10-026

Castro-Santos, T., Sprankle, K., and Perry, R. (2016). "Passage performance and migration delay of American Shad and the Holyoke Fishlifts. Fish Passage 2016," in *Proceedings of the International Conference on River Connectivity, University of Massachusetts, Amherst, MA. June 20-22, 2016*, (Amherst, MA),

Cederholm, C. J., Kunze, M. D., Murota, T., and Sibatani, A. (1999). Pacific salmon carcasses: essential contributions of nutrients and energy for aquatic and terrestrial ecosystems. *Fisheries* 24, 6–15. doi: 10.1577/1548-8446(1999)024<0006:psc>2.0.co;2

Chaloner, D. T., Martin, K. M., Wipfli, M. S., Ostrom, P. H., and Lamberti, G. A. (2002). Marine carbon and nitrogen in southeastern Alaska stream food webs: evidence from artificial and natural streams. *Can. J. Fish. Aquat. Sci.* 59, 1257–1265.

Chittenden, M. E. Jr. (1969). *Life History and Ecology of the American shad, Alosa sapidissima, in the Delaware River*. Doctoral Dissertation. New Brunswick, NJ: Rutgers University.

Chittenden, M. E. (1976). Weight loss, mortality, feeding, and duration of residence of adult American shad, *Alosa sapidissima*, in fresh water. *Fish. Bull.* 74, 151–157.

CRASC (2020). *Addendum on American Shad Passage Performance Criteria, for the Connecticut River American Shad Management Plan*. Sunderland, MA: CRASC.

Crecco, V., Savoy, T., and Gunn, L. (1983). Daily mortality rates of larval and juvenile American shad (*Alosa sapidissima*) in the Connecticut River with changes in year-class strength. *Can. J. Fish. Aquat. Sci.* 40, 1719–1728.

Dadswell, M. J., Spares, A. D., Mclean, M. F., Harris, P. J., and Rulifson, R. A. (2018). Long-term effects of tidal hydroelectric propeller turbine on the populations of three anadromous fish species. *J. Fish. Biol.* 93, 192–206. doi: 10.1111/jfb.13755

Dalton, C. M., Ellis, D., and Post, D. M. (2009). The impact of double-crested cormorant (*Phalacrocorax auritus*) predation on anadromous alewife (*Alosa pseudoharengus*) in south-central Connecticut, USA. *Can. J. Fish. Aquat. Sci.* 66, 177–186. doi: 10.1139/f08-198

Davis, J. P., Schultz, E. T., and Vokoun, J. C. (2012). Striped Bass consumption of Blueback Herring during vernal riverine migrations: does relaxing harvest restrictions on a predator help conserve a prey species of concern? *Mar. Coast. Fish.* 4, 239–251. doi: 10.1080/19425120.2012.675972

DeBruyne, R. L., DeVault, T. L., and Duerr, A. E. (2012). Spatial and temporal comparisons of double-crested cormorant diets following the establishment of Alewife in Lake Champlain, USA. *J. Great Lakes Res.* 38, 123–130. doi: 10.1016/j.jglr.2011.05.001

Dias, B. S., Frisk, M. G., and Jordaan, A. (2019). Opening the tap: Increased riverine connectivity strengthens marine food web pathways. *PLoS One* 14:e0217008. doi: 10.1371/journal.pone.0217008

Dingle, H., and Drake, V. A. (2007). What is migration? *Bioscience* 57, 113–121.

Dodson, J. J., Laroche, J., and Lecomte, F. (2009). Contrasting evolutionary pathways of anadromy in euteleostean fishes. *Am. Fish. Soc. Symp.* 69, 63–77.

Doucett, R. R., Hooper, W., and Power, G. (1999). Identification of anadromous and nonanadromous adult brook trout and their progeny in the Tabusintac River, New Brunswick, by means of multiple-stable-isotope analysis. *Trans. Am. Fish. Soc.* 128, 278–288. doi: 10.1577/1548-8659(1999)128<0278:ioaana>2.0.co;2

Doughty, C. E., Roman, J., Faurby, S., Wolf, A., Haque, A., Bakker, E. S., et al. (2016). Global nutrient transport in a world of giants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 868–873. doi: 10.1073/pnas.1502549112

Dudley, R. W. (2004). *Hydraulic-Geometry Relations for Rivers in Coastal and Central Maine. Scientific Investigations Report 2004-5042*. Reston, VA: U.S. Geological Survey.

Durbin, A. G., Nixon, S. W., and Oviatt, C. A. (1979). Effects of the spawning migration of the Alewife, *Alosa pseudoharengus*, on freshwater ecosystems. *Ecology* 60, 8–17. doi: 10.2307/1936461

Exelon. (2012). Estimation of Survival of Adult American Shad Passed Through Francis and Kaplan Turbines. RSP 3.2 Conowingo Hydroelectric Project. FERC # 405. Prepared by Normandeau Associates and Gomez and Sullivan Engingeers. Utica, NY: New Hampshire and Gomez and Sullivan Engineers, P.C. Available online at: https://mde.state.md.us/programs/Water/WetlandsandWaterways/Documents/ExelonMD/FERC/Conowingo-FRSP-3.02.pdf (accessed October 5, 2021).

FirstLight (2016). *Evaluate Downstream Passage of Juvenile American Shad, Interim Study Report*. Washington, DC: FERC Relicensing.

Foster, N. W., and Atkins, C. G. (1867). *Report of Commission on Fisheries. In Twelfth annual report of the Secretary of the Maine Board of Agriculture. Stevens and Sayward Printers to the State*. Augusta, ME: Secretary of the Maine Board of Agriculture.

Furey, N. B., and Hinch, S. G. (2017). Bull trout movements match the life history of sockeye salmon: consumers can exploit seasonally distinct resource pulses. 2017. *Trans. Am. Fish. Soc.* 146, 450–461. doi: 10.1080/00028487.2017.1285353

Garman, G. C. (1992). Fate and potential significance of post spawning anadromous fish carcasses in an Atlantic coastal river. *Trans. Am. Fish. Soc.* 121, 390–394. doi: 10.1577/1548-8659(1992)121<0390:fapsop>2.3.co;2

Gibson, A. J. F., and Meyers, R. A. (2002). A logistic regression model for estimating turbine mortality at hydroelectric generating stations. *Trans. Am. Fish. Soc.* 131, 623–633. doi: 10.1577/1548-8659(2002)131<0623:alrmfe>2.0.co;2

Gilligan-Lunda, E. K., Stich, D. S., Mills, K. E., Bailey, M. M., and Zydlewski, J. D. (2021). Climate change may cause shifts in growth and instantaneous natural mortality of American Shad throughout their native range. *Trans. Am. Fish. Soc.* 150, 407–421. doi: 10.1002/tafs.10299

Glebe, B. D., and Leggett, W. C. (1981a). Latitudinal differences in energy allocation and use during the freshwater migrations of American shad (*Alosa sapidissima*) and their life history consequences. *Can. J. Fish. Aquat. Sci.* 38, 806–820. doi: 10.1139/f81-109

Glebe, B. D., and Leggett, W. C. (1981b). Temporal, intra-population differences in energy allocation and use by American shad (*Alosa sapidissima*) during the spawning migration. *Can. J. Fish. Aquat. Sci.* 38, 795–805. doi: 10.1139/f81-108

Goethel, D. R., Quinn, T. J., and Cadrin, S. X. (2011). Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. *Rev. Fish. Sci.* 19, 119–136. doi: 10.1080/10641262.2011.557451

Gresh, T., Lichatowich, J., and Schoonmaker, P. (2000). An estimation of historic and current levels of salmon production in the Northeast Pacific ecosystem: evidence of a nutrient deficit in the freshwater systems of the Pacific Northwest. *Fisheries* 25, 15–21. doi: 10.1577/1548-8446(2000)025<0015:aeohac>2.0.co;2

Gross, M. R. (1987). Evolution of diadromy in fishes. *Am. Fish. Soc. Symp.* 1, 14–25.

Grote, A. B., Bailey, M. M., and Zydlewski, J. D. (2014). Movements and demography of spawning American shad in the Penobscot River, Maine, prior to dam removal. *Trans. Am. Fish. Soc.* 143, 552–563. doi: 10.1080/00028487.2013.864705

Groux, F., Therrien, J., Chanseau, M., Courret, D., and Tetard, S. (2015). *Knowledge Update on Shad Upstream Migration Fishway Design and Efficiency – Project LIFE09 NAT/DE/000008 – Conservation and Restoration of the Allis Shad in the Gironde and Rhine watersheds – Action A1. Report from WSP to ONEMA*. 81.

Hall, C. J., Jordaan, A., and Frisk, M. G. (2011). The historic influence of dams on diadromous fish habitat with a focus on river herring and hydrologic longitudinal connectivity. *Landsc. Ecol.* 26, 95–107. doi: 10.1007/s10980-010-9539-1

Haro, A., and Castro-Santos, T. (2012). Passage of American Shad: paradigms and realities. *Mar. Coast. Fish.* 4, 252–261. doi: 10.1080/19425120.2012.675975

Harris, J. E., and Hightower, J. E. (2012). Demographic population model for american shad: will access to additional habitat upstream of dams increase population sizes? *Mar. Coast. Fish.* 4, 262–283. doi: 10.1080/19425120.2012.675969

Hart, D. D., Johnson, T. E., Bushaw-Newton, K. L., Horwitz, R. J., Bednarek, A. T., Charles, D. F., et al. (2002). Dam removal: challenges and opportunities for ecological research and river restoration: we develop a risk assessment framework for understanding how potential responses to dam removal vary with dam and watershed characteristics, which can lead to more effective use of this restoration method. *BioScience* 52, 669–682.

Haskell, C. A. (2018). From salmon to shad: Shifting sources of marine-derived nutrients in the Columbia River Basin. *Ecol. Freshw. Fish.* 27, 310–322. doi: 10.1111/eff.12348

Hasselman, D. J., Bradford, R. G., and Bentzen, P. (2010). Taking stock: defining populations of American shad (*Alosa sapidissima*) in Canada using neutral genetic markers. *Can. J. Fish. Aquat. Sci.* 67, 1021–1039. doi: 10.1139/f10-031

Hasselman, D. J., and Limburg, K. E. (2012). Alosine restoration in the 21st century: challenging the status quo. *Mar. Coast. Fish.* 4, 174–187. doi: 10.1080/19425120.2012.675968

Hasselman, D. J., Ricard, D., and Bentzen, P. (2013). Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Mol. Ecol.* 22, 1558–1573. doi: 10.1111/mec.12197

Heisey, P. G., Mathur, D., and Rineer, T. (1992). A reliable tag-recapture technique for estimating turbine passage survival: application to young-of-the-year American Shad (*Alosa sapidissima*). *Can. J. Fish. Aquat. Sci.* 49, 1826–1834. doi: 10.1139/f92-202

Hightower, J. E., Harris, J. E., Raabe, J. K., Brownell, P., and Drew, C. A. (2012). A Bayesian spawning habitat suitability model for American Shad in southeastern United States Rivers. *J. Fish. Wildlife Manag.* 3, 184–198. doi: 10.3996/082011-jfwm-047

Hightower, J. E., and Sparks, K. L. (2003). "Migration and spawning habitat of American shad in the Roanoke river, North Carolina," in *Biodiversity, Status, and Conservation Of The World's Shads*, eds K. E. Limburg and J. R. Waldman (Bethesda: American Fisheries Society Symposium), 193–199.

Hightower, J. E., Wicker, A. M., and Endres, K. M. (1996). Historical trends in abundance of American shad and river herring in Albemarle Sound, North Carolina. *North Am. J. Fish.Manag.* 16, 257–271. doi: 10.1577/1548-8675(1996)016<0257:htiaoa>2.3.co;2

Hill, D. R. (1959). *Some Uses of Statistical Analysis in Classifying Races of American Shad (Alosa sapidissima)*. Washington, DC: US Government Printing Office.

Hixon, M. A., Johnson, D. W., and Sogard, S. M. (2014). BOFFFs: on the importance of conserving old- growth age structure in fishery populations. *ICES J. Mar. Sci.* 71, 2171–2185. doi: 10.1093/icesjms/fst200

Hocking, M. D., and Reimchen, T. E. (2009). Salmon species, density and watershed size predict magnitude of marine enrichment in riparian food webs. *Oikos* 118, 1307–1318. doi: 10.1111/j.1600-0706.2009.17302.x

Hogans, W. E., and Melvin, G. D. (1985). *Mortality of Adult American Shad (Alosa sapidissima) Passed Through a Stratflo Turbine at the low Head Tidal Power Generating Station at Annapolis Royal, Nova Scotia*. Wolfville, NS: T. P. H. Applied Fisheries Research, Inc.

Hogg, R., Coghlan, S. M. Jr., and Zydlewski, J. (2013). Anadromous sea lampreys recolonize a Maine coastal river tributary after dam removal. *Trans. Am. Fish. Soc.* 142, 1381–1394. doi: 10.1080/00028487.2013.811103

Hyle, A. R., McBride, R. S., and Olney, J. E. (2014). Determinate versus indeterminate fecundity in American shad, an anadromous clupeid. *Trans. Am. Fisher. Soc.* 143, 618–633. doi: 10.1080/00028487.2013.862178

Izzo, L. K., Maynard, G. A., and Zydlewski, J. (2016). Upstream movements of Atlantic salmon in the lower Penobscot River, Maine following two dam removals and fish passage modifications. *Mar. Coast. Fish.* 8, 448–461. doi: 10.1080/19425120.2016.1185063

Jaecks, T., and Quinn, T. P. (2014). Ontogenetic shift to dependence on salmon-derived nutrients in Dolly Varden char from the Iliamna River, Alaska. *Environ. Biol. Fish.* 97, 1323–1333. doi: 10.1007/s10641-014-0221-3

Kelt, D. A., and Van Vuren, D. H. (2001). The ecology and macroecology of mammalian home range area. *Am. Natural.* 157, 637–645. doi: 10.2307/3079304

Kiraly, I. A., Coghlan, S. M. Jr., Zydlewski, J., and Hayes, D. (2015). An assessment of fish assemblage structure in a large river. *River Res. Applic.* 31, 301–312. doi: 10.1002/rra.2738

Larinier, M., and Travade, F. (2002). The design of fishways for shad. *Bull. Francais Peche Piscicul.* 364, 135–146. doi: 10.1051/kmae/2002098

Leggett, W. C. (1972). Weight loss in American shad (*Alosa sapidissima*), Wilson) during the freshwater migration. *Trans. Am. Fish. Soc.* 101, 549–552. doi: 10.1577/1548-8659(1972)101<549:wliasa>2.0.co;2

Leggett, W. C., and Carscadden, J. E. (1978). Latitudinal variation in reproductive characteristics of American Shad (*Alosa sapidissima*): Evidence for population specific life history strategies in fish. *J. Fish. Res. Board Can.* 35, 1469–1478. doi: 10.1139/f78-230

Leggett, W. C., Savoy, T. F., and Tomichek, C. A. (2004). "The impact of enhancement initiatives on the structure and dynamics of the Connecticut River population of American shad," in *The Connecticut River Ecological Study (1965–1973) Revisited: Ecology of the Lower Connecticut River, 1973–2003. American Fisheries Society, Monograph 9*, eds P. M. Jacobson, D. A. Dixon, W. C. Leggett, B. C. Marcy Jr., and R. R. Massengill (Bethesda, MD: American Fisheries Society), 391–405.

Leggett, W. C., and Whitney, R. R. (1972). Water temperature and the migrations of American shad. *Fish. Bull.* 70, 659–670.

Lennox, R. J., Eliason, E. J., Havn, T. B., Johansen, M. R., Thorstad, E. B., Cooke, S. J., et al. (2018). Bioenergetic consequences of warming rivers to adult Atlantic salmon *Salmo salar* during their spawning migration. *Freshw. Biol.* 63, 1381–1393. doi: 10.1111/fwb.13166

Leonard, J. B., and McCormick, S. D. (1999). Effects of migration distance on whole-body and tissue-specific energy use in American shad (*Alosa sapidissima*). *Can. J. Fish. Aquat. Sci.* 56, 1159–1171. doi: 10.1139/f99-041

Leopold, L., and Maddock, T. (1953). The hydraulic geometry of stream channels and some physiographic implications. *U.S. Geol. Survey Prof. Pap.* 252, 57.

Liermann, C. R., Nilsson, C., Robertson, J., and Ng, R. Y. (2012). Implications of dam obstruction for global freshwater fish diversity. *BioScience* 62, 539–548. doi: 10.1525/bio.2012.62.6.5

Limburg, K., Hattala, K., and Kahnle, A. (2003). *American Shad in its Native Range. Biodivers. Status Conserv. Worlds Shads*. 125–140.

Limburg, K. E. (2001). Through the gauntlet again: demographic restructuring of American shad by migration. *Ecology* 82, 1584–1596. doi: 10.1890/0012-9658(2001)082[1584:ttgadr]2.0.co;2

Limburg, K. E., and Waldman, J. R. (2009). Dramatic declines in North Atlantic diadromous fishes. *BioScience* 59, 955–965. doi: 10.1525/bio.2009.59.11.7

Mann, D. A., Lu, Z., Hastings, M. C., and Popper, A. N. (1998). Detection of ultrasonic tones and simulated dolphin echolocation clicks by a teleost fish, the American shad (*Alosa sapidissima*). *J. Acoustical Soc. Am.* 104, 562–568. doi: 10.1121/1.423255

Mansueti, R., and Kolb, H. (1953). *A Historical Review of the Shad Fisheries of North America*. Solomons, MD: Chesapeake Biological Laboratory.

Martin, B. T., Nisbet, R. M., Pike, A., Michel, C. J., and Danner, E. M. (2015). Sport science for salmon and other species: ecological consequences of metabolic power constraints. *Ecol. Lett.* 18, 535–544. doi: 10.1111/ele.12433

Martin, E. (2013). *Chesapeake Fish Passage Prioritization*. Available online at: http://maps.tnc.org/EROF_ChesapeakeFPP/ (accessed October 5, 2021).

Martin, E. H. (2019). Assessing and prioritizing barriers to aquatic connectivity in the Eastern United States. *J. Am. Water Resour. Assoc.* 55, 401–412. doi: 10.1111/1752-1688.12694

Martin, E. H., and Apse, C. D. (2011). *Northeast Aquatic Connectivity: An Assessment of Dams on Northeastern Rivers. The Nature Conservancy, Eastern Freshwater Program*. Arlington, VA: The Nature Conservancy.

Mathur, D., Heisey, P. G., and Robinson, D. A. (1994). Turbine passage mortality of juvenile American shad at a low head hydroelectric dam. *Trans. Am. Fish. Soc.* 123, 108–111. doi: 10.1577/1548-8659(1994)123<0108:ntpmoj>2.3.co;2

McBride, R. S., Ferreri, R., Towle, E. K., Boucher, J. M., and Basilone, G. (2016). Yolked oocyte dynamics support agreement between determinate- and indeterminate-method estimates of annual fecundity for a northeastern United States population of American shad. *PLoS One* 11:e0164203. doi: 10.1371/journal.pone.0164203

McCormick, S. D., Hansen, L. P., Quinn, T. P., and Saunders, R. L. (1998). Movement, migration, and smolting of Atlantic salmon (*Salmo salar*). *Can. J. Fish. Aquat. Sci.* 55, 77–92. doi: 10.1139/d98-011

McDermott, S. P., Bransome, N. C., Sutton, S. E., Smith, B. E., Link, J. S., and Miller, T. J. (2015). Quantifying alosine prey in the diets of marine piscivores in the Gulf of Maine. *J. Fish. Biol.* 86, 1811–1829. doi: 10.1111/jfb.12692

McDowall, R. M. (1987). "The occurrence and distribution of diadromy among fishes, Vol. 1," in *Common Strategies of Anadromous and Catadromous Fishes*, eds M. J. Dadswell, R. J. Klauda, C. M. Moffit, R. L. Saunders, R. A. Rulifson, and J. E. Cooper (Bathesda, MD: American Fisheries Society Symposium), 1–13.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A. (2012). *NHDPlus Version 2: User Guide*. Washington, DC: US Environmental Protection Agency.

McPhee, J. (2003). *The Founding Fish*. New York, NY: Farrar, Straus and Giroux.

Meixler, M. S., Bain, M. B., and Walter, M. T. (2009). Predicting barrier passage and habitat suitability for migratory fish species. *Ecol. Modell.* 220, 2782–2791. doi: 10.1016/j.ecolmodel.2009.07.014

Melvin, G. D., Dadswell, M. J., and Martin, J. D. (1986). Fidelity of American shad, *Alosa sapidissima* (Clupeidae), to its river of previous spawning. *Can. J. Fish. Aquat. Sci.* 43, 640–646. doi: 10.1139/f86-077

Miranda, L. E., Bettoli, P. W., and Brown, M. L. (2007). in *Analysis and Interpretation of Freshwater Fisheries Data*, eds C. S. Guy and M. L. Brown (Bethesda MD: American Fisheries Society), 229–277. doi: 10.47886/9781888569773.ch6

Mohamoud, Y. M., and Parmar, R. S. (2006). Estimating streamflow and associated hydraulic geometry, the mid-atlantic region, USA 1. *JAWRA J. Am. Water Resour. Assoc.* 42, 755–768. doi: 10.1111/j.1752-1688.2006.tb04490.x

Moore, J. W., Hayes, S. A., Duffy, W., Gallagher, S., Michel, C. J., and Wright, D. (2011). Nutrient fluxes and the recent collapse of coastal California salmon populations. *Can. J. Fish. Aquat. Sci.* 68, 1161–1170. doi: 10.1139/f2011-054

Moore, J. W., and Schindler, D. E. (2004). Nutrient export from freshwater ecosystems by anadromous Sockeye Salmon (*Oncorhynchus nerka*). *Can. J. Fish. Aquat. Sci.* 61, 1582–1589. doi: 10.1139/f04-103

Moser, M. L., and Paradis, R. L. (2017). Pacific lamprey restoration in the Elwha River drainage following dam removals. *Am. Curr.* 42, 3–8.

Nadeau, P. S., Hinch, S. G., Hruska, K. A., Pon, L. B., and Patterson, D. A. (2010). The effects of experimental energy depletion on the physiological condition and survival of adult sockeye salmon (*Oncorhynchus nerka*) during spawning migration. *Environ. Biol. Fish.* 88, 241–251. doi: 10.1007/s10641-010-9635-8

Naiman, R. J., Bilby, R. E., Schindler, D. E., and Helfield, J. M. (2002). Pacific salmon, nutrients, and the dynamics of freshwater and riparian ecosystems. *Ecosystems* 5, 399–417. doi: 10.1007/s10021-001-0083-3

Natural Resources Canada. (2019). *National Hydrographic Network (NHN)*. Available online at: http://ftp.maps.canada.ca/pub/nrcan_rncan/vector/geobase_nhn_rhn/ (accessed October 5, 2021).

Newbold, J. D., Mulholland, P. J., Elwood, J. W., and O'neill, R. V. (1982). Organic carbon spiralling in stream ecosystems. *Oikos* 38, 266–272. doi: 10.2307/3544663

Nichols, P. R. (1966). *Comparative Study of Juvenile American Shad Populations by Finray and Scute Counts (No. 525). US Department of the Interior, Bureau of Commercial Fisheries*. Washington, DC: US Department of the Interior.

NOAA (2009). *Biological Valuation of Atlantic Salmon Habitat Within the Gulf of Maine Distinct Population Segment*. Gloucester, MA: NOAA National Marine Fisheries Service.

Noonan, M. J., Grant, J. W., and Jackson, C. D. (2012). A quantitative assessment of fish passage efficiency. *Fish Fish.* 13, 450–464. doi: 10.1111/j.1467-2979.2011.00445.x

O'Leary, J. A., and Kynard, B. (1986). Behavior, length, and sex ratio of seaward-migrating juvenile American shad and blueback herring in the Connecticut River. *Trans. Am. Fish. Soc.* 115, 529–536. doi: 10.1577/1548-8659(1986)115<529:blasro>2.0.co;2

Olney, J. E., and Hoenig, J. M. (2001). Managing a fishery under moratorium: assessment opportunities for Virginia's stocks of American shad. *Fisheries* 26, 6–12. doi: 10.1577/1548-8446(2001)026<0006:mafuma>2.0.co;2

Olney, J. E., and McBride, R. S. (2003). Intraspecific variation in batch fecundity of American shad: revisiting the paradigm of reciprocal latitudinal trends in reproductive traits. *Am. Fish. Soc. Symp.* 35, 185–192.

Opperman, J. J., Royte, J., Banks, J., Day, L. R., and Apse, C. (2011). The Penobscot River, Maine, USA: a basin-scale approach to balancing power generation and ecosystem restoration. *Ecol. Soc.* 16:7.

Pess, G. R., Quinn, T. P., Gephard, S. R., and Saunders, R. (2014). Re-colonization of Atlantic and Pacific rivers by anadromous fishes: linkages between life history and the benefits of barrier removal. *Rev. Fish Biol. Fish.* 24, 881–900. doi: 10.1007/s11160-013-9339-1

Petersen, J. H., Hinrichsen, R. A., Gadomski, D. M., Feil, D. H., and Rondorf, D. W. (2003). "American shad in the Columbia River," in *Biodiversity, Status, and Conservation of the World's Shads*, Vol. 35, eds K. E. Limburg and J. R. Waldman (Bethesda, MD: American Fisheries Society, Symposium), 141–155.

Polis, G. A., Anderson, W. B., and Holt, T. D. (1997). Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs. *Annu. Rev. Ecol. Evol. Syst.* 28, 289–316. doi: 10.1146/annurev.ecolsys.28.1.289

Post, D. M., and Walters, A. W. (2009). Nutrient excretion rates of anadromous Alewives during their spawning migration. *Trans. Am. Fish. Soc.* 138, 264–268. doi: 10.1577/t08-111.1

Provost, J. (1987). *L'alose Savoureuse (Alosa sapidissima,Wilson) du Fleuve Saint-Laurent: Etude Comparative des Phenotypes Morphologiques et de Certains Aspects de la Biologie de Quelques Populations*. Master's Thesis. Quebec: University of Quebec. (In French.).

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raabe, J. K., and Hightower, J. E. (2014). American Shad migratory behavior, weight loss, survival, and abundance in a North Carolina river following dam removals. *Trans. Am. Fish. Soc.* 143, 673–688. doi: 10.1080/00028487.2014.882410

Rand, P. S., and Hinch, S. G. (1998). Swim speeds and energy use of upriver-migrating sockeye salmon (*Oncorhynchus nerka*): simulating metabolic power and assessing risk of energy depletion. *Can. J. Fish. Aquat. Sci.* 55, 1832–1841. doi: 10.1139/f98-068

Rex, J. F., and Petticrew, E. L. (2008). Delivery of marine-derived nutrients to streambeds by Pacific salmon. *Nat. Geosci.* 1, 840–843. doi: 10.1038/ngeo364

Richardson, D. E., Palmer, M. C., and Smith, B. E. (2014). The influence of forage fish abundance on the aggregation of Gulf of Maine Atlantic cod (Gadus morhua) and their catchability in the fishery, Cooper A, editor. *Can. J. Fish. Aquat. Sci.* 71, 1349–1362. doi: 10.1139/cjfas-2013-0489

Rosenzweig, M. L. (1971). Paradox of enrichment: destabilization of exploitation ecosystems in ecological time. *Science* 171, 385–387. doi: 10.1126/science.171.3969.385

Roy, S., Daignault, A., Zydlewski, J., Truhlar, A., Smith, S., Jain, S., et al. (2020). Coordinated river infrastructure decisions enhance social-ecological resilience. *Environ. Res. Lett.* 15:104054. doi: 10.1088/1748-9326/abad58

Roy, S., Uchida, E., Souza, S., Blachly, B., Fox, E., Gardner, K., et al. (2018). Damming decisions: a multi-scale approach to balance trade-offs among dam infrastructure, river restoration, and cost. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12069–12074. doi: 10.1073/pnas.1807437115

Rulifson, R. A. (1994). "Status of anadromous Alosa along the east coast of North American," in *Anadromous Alosa Symposium: Proceedings of a Symposium held at the Seventh Annual meeting of the Tidewater Chapter in Virginia Beach, Virginia, 14-15 January 1993*, eds J. E. Cooper, R. T. Eades, R. J. Klauda, and J. G. Loesch (Bethesda, MD: American Fisheries Society), 134–158.

Samways, K. M., and Cunjak, R. A. (2015). Increases in benthic community production and metabolism in response to marine-derived nutrients from spawning Atlantic Salmon (*Salmo salar*). *Freshw. Biol.* 60, 1647–1658. doi: 10.1111/fwb.12597

Schindler, D. E., Scheuerell, M. D., Moore, J. W., Gende, S. M., Francis, T. B., and Palen, W. J. (2003). Pacific salmon and the ecology of coastal ecosystems. *Front. Ecol. Environ.* 1:31–37.

Schmitt, J. D., Hallerman, E. M., Bunch, A., Moran, Z., Emmel, J. A., and Orth, D. J. (2017). Predation and prey selectivity by nonnative catfish on migrating alosines in an Atlantic slope estuary. *Mar. Coast. Fish.* 9, 108–125. doi: 10.1080/19425120.2016.1271844

Shrimpton, J. M., Zydlewski, J. D., and McCormick, S. D. (2001). The stress response of juvenile American shad to handling and confinement is greater during migration in freshwater than in seawater. *Trans. Am. Fish. Soc.* 130, 1203–1210. doi: 10.1577/1548-8659(2001)130<1203:tsroja>2.0.co;2

Silva, A. T., Lucas, M. C., Castro-Santos, T., Katopodis, C., Baumgartner, L. J., Thiem, J. D., et al. (2018). The future of fish passage science, engineering, and practice. *Fish Fish.* 19, 340–362. doi: 10.1111/faf.12258

Song, C., O'Malley, A., Roy, S., Zydlewski, J., Barber, B., and Mo, W. (2019). Managing dams for energy and fish tradeoffs: what does a win-win solution take? *Sci. Total Environ.* 669, 833–843. doi: 10.1016/j.scitotenv.2019.03.042

Stevenson, C. H. (1899). *The Shad Fisheries of the Atlantic Coast of the United States. Pages 101–269 in U.S. Commission of Fish and Fisheries, part24, Report of the Commissioner for the Year Ending June 30, 1898*. Washington, DC: U.S. Commission of Fish and Fisheries.

Stich, D. S., Roy, S. G., and Zydlewski, J. D. (2020). *Anadrofish: Anadromous Fish Population Responses to Habitat Changes. R Package Version 1.1.0*. Available online at: https://github.com/danStich/anadrofish (accessed October 5, 2021).

Stich, D. S., Sheehan, T. F., and Zydlewski, J. D. (2019). A dam passage performance standard model for American shad. *Can. J. Fish. Aquat. Sci.* 76, 762–779. doi: 10.1139/cjfas-2018-0008

Stier, D. J., and Crance, J. H. (1985). *Habitat Suitability Index Models and Instream Flow Suitability Curves: American Shad*. Washington, DC: U.S. Fish and Wildlife Service.

Stokesbury, K. D., and Dadswell, M. J. (1991). Mortality of juvenile clupeids during passage through a tidal, low head hydroelectric turbine at Annaplis Royal, Nova Scotia. *North Am. J. Fish. Manag.* 11, 149–154. doi: 10.1577/1548-8675(1991)011<0149:mojcdp>2.3.co;2

Sullivan, T. (2004). *Evaluation of the Turners Falls Fishway Complex and Potential Improvements for Passing American Shad.* Master's Thesis. Boston, MA: University of Massachusetts.

Sweet, W. V., and Geratz, J. W. (2003). Bankfull Hydraulic Geometry Relationships and Recurrence Intervals for North Carolina's Coastal Plain. *J. Am. Water Resour. Assoc.* 39, 861–871. doi: 10.1111/j.1752-1688.2003.tb04411.x

Talbot, G. B. (1954). Factors associated with fluctuations in abundance of Hudson River shad. *U.S. Fish Wildlife Serv. Fish. Bull.* 56, 373–413.

Taylor, A. (1995). *William Cooper's Town: Power and Persuasion on the Frontier of the Early American Republic*. New York, NY: A. A. Knopf.

Then, A. Y., Hoenig, J. M., Hall, N. G., and Hewitt, D. A. (2015). Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES J. Mar. Sci.* 72, 82–92. doi: 10.1093/icesjms/fsu136

Tommasi, D., Nye, J., Stock, C., Hare, J. A., Alexander, M., and Drew, K. (2015). Effect of environmental conditions on juvenile recruitment of alewife (*Alosa pseudoharengus*) and blueback herring (*Alosa aestivalis*) in fresh water: a coastwide perspective. *Can. J. Fish. Aquat. Sci.* 72, 1037–1047. doi: 10.1139/cjfas-2014-0259

TransCanada. (2016). *ILP Study 22 Downstream migration of juvenile American shad at Vernon. Study report. Submitted to FERC 2016. Prepared by Normandeau Associates, NH*.

Trinko Lake, T. R., Ravana, K. R., and Saunders, R. (2012). Evaluating changes in diadromous species distributions and habitat accessibility following the Penobscot River Restoration Project. *Mar. Coast. Fish.* 4, 284–293. doi: 10.1080/19425120.2012.675971

Twining, C. W., West, D. C., and Post, D. M. (2013). Historical changes in nutrient inputs from humans and anadromous fishes in New England's coastal watersheds. *Limnol. Oceanogr.* 58, 1286–1300. doi: 10.4319/lo.2013.58.4.1286

USFWS (2019). *Turbine Blade Strike Model*. Washington, DC: USFWS.

USGS (2019). *National Hydrography Dataset (n.d.) Plus, Version 2. U.S. Environmental Protection Agency and U.S. Geological Survey*. Reston, VA: USGS.

Vanni, M. J. (2002). Nutrient cycling by animals in freshwater ecosystems. *Annu. Rev. Ecol. Evol. Syst.* 33, 341–370. doi: 10.1146/annurev.ecolsys.33.010802.150519

Venables, W. N., and Ripley, B. D. (2002). "Random and mixed effects," in *Modern Applied Statistics With S. Statistics and Computing*, (New York, NY: Springer).

Vogel, S., and Jansujwicz, J. (2021). Fish passage is an upstream battle during FERC relicensing in Maine. *Fish. Manag. Ecol.* doi: 10.1111/fme.12513

von Bertalanffy, L. (1938). A quantitative theory of organic growth (inquiries on growth laws II). *Hum. Biol.* 10, 181–213.

Walburg, C. H., and Nichols, P. R. (1967). *Biology and Management of the American Shad and Status of the Fisheries, Atlantic Coast of the United States, 1960. U.S. Fish and Wildlife Service Special Scientific Report–Fisheries 550*. Washington, DC: NOAA.

Walter, J. F. III, and Olney, J. E. (2003). Feeding behavior of American shad during spawning migration in the York River, Virginia. *Am. Fish. Soc. Symp.* 35, 201–209.

Walters, A. W., Barnes, R. T., and Post, D. M. (2009). Anadromous Alewives (*Alosa pseudoharengus*) contribute marine-derived nutrients to coastal stream food webs. *Can. J. Fish. Aquat. Sci.* 66, 439–448. doi: 10.1139/F09-008

Walters, C., and Korman, J. (1999). Linking recruitment to trophic factors: revisiting the Beverton-Holt recruitment model from a life history and multispecies perspective. *Rev. Fish Biol. Fish.* 9, 187–202. doi: 10.1023/A:1008991021305

Watson, J. M., Coghlan, S. M. Jr., Zydlewski, J., Hayes, D. B., and Kiraly, I. A. (2018). Dam removal and fish passage improvement influence fish assemblages in the Penobscot River, Maine. *Trans. Am. Fish. Soc.* 147, 525–540. doi: 10.1002/tafs.10053

Weaver, C. R., Thompson, C. S., and Ossiander, F. J. (1972). *Evaluation of Fish Passage in the Vertical Slot Regulating Section of the South Shore Ladder at the John Day Dam. National Marine Fisheries Service, Final Report to the Army Corps. Of Engingeers*. Portland: Portland District.

Weaver, D. M., Brown, M., and Zydlewski, J. D. (2019). Observations of American Shad Alosa sapidissima approaching and using a vertical slot fishway at the head-of-tide Brunswick Dam on the Androscoggin River, Maine. *North Am. J. Fisher. Manag.* 39, 989–998.

Weaver, D. M., Coghlan, S. M. Jr., Greig, H. S., Klemmer, A. J., Perkins, L. B., and Zydlewski, J. (2018). Subsidies from anadromous sea lamprey (*Petromyzon marinus*) carcasses function as a reciprocal nutrient exchange between marine and freshwaters. *River Res. Applic.* 34, 824–833. doi: 10.1002/rra.3291

Weaver, D. M., Coghlan, S. M. Jr., and Zydlewski, J. (2016). Sea lamprey carcasses exert local and variable food web effects in a nutrient-limited Atlantic coastal stream. *Can. J. Fish. Aquat. Sci.* 73, 1616–1625. doi: 10.1139/cjfas-2015-0506

Weaver, D. M., Coghlan, S. M., Zydlewski, J., Hogg, R. S., and Canton, M. (2015). Decomposition of sea lamprey *Petromyzon marinus* carcasses:

temperature effects, nutrient dynamics, and implications for stream food webs. *Hydrobiologia* 760, 57–67. doi: 10.1007/s10750-015-2302-5

West, D. C., Walters, A. W., Gephard, S., and Post, D. M. (2010). Nutrient loading by anadromous Alewife (*Alosa pseudoharengus*): contemporary patterns and predictions for restoration efforts. *Can. J. Fish. Aquat. Sci.* 67, 1211–1220. doi: 10.1139/F10-059

Wilcove, D. S. (2010). *No Way Home: The Decline of the World's Great Animal Migrations*. Washington, DC: Island Press.

Williams, R. O., and Bruger, G. E. (1972). *Investigations on American Shad in the St. Johns River. Florida Department of Natural Resources, Marine Research Laboratory, Technical Series*, Vol. 66, 1–49.

Williams, R. R., and Daborn, G. R. (1984). Spawning of the American shad (*Alosa sapidissima*) in the Annapolis River, Nova Scotia. *Proc. Nova Scotian Instit. Sci.* 34, 9–14.

Willson, M. F., and Halupka, K. C. (1995). Anadromous fish as keystone species in vertebrate communities. *Conserv. Biol.* 9, 489–497. doi: 10.1046/j.1523-1739.1995.09030489.x

Willson, M. F., and Womble, J. N. (2006). Vertebrate exploitation of pulsed marine prey: a review and the example of spawning herring (2006). *Rev. Fish. Biol. Fish.* 16, 183–200. doi: 10.1007/s11160-006-9009-7

Wippelhauser, G. (2021). Recovery of diadromous fishes: a kennebec river case study. *Trans. Am. Fish. Soc.* 150, 277–290. doi: 10.1002/tafs.10292

Zydlewski, J., and McCormick, S. D. (1997b). The loss of hyperosmoregulatory ability in migrating juvenile American shad, *Alosa sapidissima*. *Can. J. Fish. Aquat. Sci.* 54, 2377–2387. doi: 10.1139/f97-144

Zydlewski, J., and McCormick, S. D. (1997a). The ontogeny of salinity tolerance in the American shad, *Alosa sapidissima*. *Can. J. Fish. Aquat. Sci.* 54, 182–189. doi: 10.1139/f96-251

Zydlewski, J., McCormick, S. D., and Kunkel, J. G. (2003). Late migration and seawater entry is physiologically disadvantageous for American shad juveniles. *J. Fish Biol.* 63, 1521–1537. doi: 10.1111/j.1095-8649.2003.00264.x

Zydlewski, J., and Wilkie, M. P. (2012). Freshwater to seawater transitions in migratory fishes. *Fish Physiol.* 32, 253–326. doi: 10.1016/B978-0-12-396951-4.00006-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fmars.2021.665474

The Role of Climate, Oceanography, and Prey in Driving Decadal Spatio-Temporal Patterns of a Highly Mobile Top Predator

Amaia Astarloa[1], Maite Louzao[1], Joana Andrade[2], Lucy Babey[3], Simon Berrow[4], Oliver Boisseau[5], Tom Brereton[6], Ghislain Dorémus[7], Peter G. H. Evans[8,9], Nicola K. Hodgins[10], Mark Lewis[11], Jose Martinez-Cedeira[12], Malin L. Pinsky[13], Vincent Ridoux[6], Camilo Saavedra[14], M. Begoña Santos[14], James T. Thorson[15], James J. Waggitt[9], Dave Wall[4] and Guillem Chust[1]*

[1] AZTI Marine Research, Basque Research and Technology Alliance (BRTA), Pasaia, Spain, [2] Sociedade Portuguesa para o Estudo das Aves (SPEA), Lisbon, Portugal, [3] ORCA, Portsmouth, United Kingdom, [4] Irish Whale and Dolphin Group (IWDG), Kilrush, Ireland, [5] Marine Conservation Research, Kelvedon, United Kingdom, [6] Marinelife, Bridport, United Kingdom, [7] Observatoire PELAGIS, Université de La Rochelle, La Rochelle, France, [8] Sea Watch Foundation (SWF), Amlwch, United Kingdom, [9] School of Ocean Sciences, Bangor University, Menai Bridge, United Kingdom, [10] Whale and Dolphin Conservation (WDC), Chippenham, United Kingdom, [11] Joint Nature Conservation Committee (JNCC), Peterborough, United Kingdom, [12] Coordinadora para o Estudo dos Mamíferos Mariños (CEMMA), Pontevedra, Spain, [13] Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ, United States, [14] Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Vigo, Vigo, Spain, [15] Habitat and Ecological Processes Research Program (HEPR), Alaska Fisheries Science Center, Seattle, WA, United States

OPEN ACCESS

Edited by:
Mary C. Fabrizio, William & Mary's Virginia Institute of Marine Science, United States

Reviewed by:
Bruno Díaz López, Bottlenose Dolphin Research Institute (BDRI), Spain
Leigh Gabriela Torres, Oregon State University, United States

***Correspondence:**
Amaia Astarloa
aastarloa@azti.es

Specialty section:
This article was submitted to Marine Ecosystem Ecology, a section of the journal Frontiers in Marine Science

Received: *08 February 2021*
Accepted: *17 September 2021*
Published: *18 November 2021*

Citation:
Astarloa A, Louzao M, Andrade J, Babey L, Berrow S, Boisseau O, Brereton T, Dorémus G, Evans PGH, Hodgins NK, Lewis M, Martinez-Cedeira J, Pinsky ML, Ridoux V, Saavedra C, Santos MB, Thorson JT, Waggitt JJ, Wall D and Chust G (2021) The Role of Climate, Oceanography, and Prey in Driving Decadal Spatio-Temporal Patterns of a Highly Mobile Top Predator. Front. Mar. Sci. 8:665474. doi: 10.3389/fmars.2021.665474

Marine mammals have been proposed as ecosystem sentinels due to their conspicuous nature, wide ranging distribution, and capacity to respond to changes in ecosystem structure and functioning. In southern European Atlantic waters, their response to climate variability has been little explored, partly because of the inherent difficulty of investigating higher trophic levels and long lifespan animals. Here, we analyzed spatio-temporal patterns from 1994 to 2018 of one of the most abundant cetaceans in the area, the common dolphin (*Delphinus delphis*), in order to (1) explore changes in its abundance and distribution, and (2) identify the underlying drivers. For that, we estimated the density of the species and the center of gravity of its distribution in the Bay of Biscay (BoB) and tested the effect of three sets of potential drivers (climate indices, oceanographic conditions, and prey biomasses) with a Vector Autoregressive Spatio Temporal (VAST) model that accounts for changes in sampling effort resulting from the combination of multiple datasets. Our results showed that the common dolphin significantly increased in abundance in the BoB during the study period. These changes were best explained by climate indices such as the North Atlantic Oscillation (NAO) and by prey species biomass. Oceanographic variables such as chlorophyll *a* concentration and temperature were less useful or not related. In addition, we found high variability in the geographic center of gravity of the species within the study region, with shifts between the inner (southeast) and the outer (northwest) part of the BoB, although the majority of this variability could not be attributed to the drivers considered in the study.

Overall, these findings indicate that considering temperature alone for projecting spatio-temporal patterns of highly mobile predators is insufficient in this region and suggest important influences from prey and climate indices that integrate multiple ecological influences. Further integration of existing observational datasets to understand the causes of past shifts will be important for making accurate projections into the future.

Keywords: common dolphin, center of gravity, climate indices, predator-prey, environmental variability, time series, Bay of Biscay, VAST

INTRODUCTION

The global mean surface temperature has increased by approximately 1°C from pre-industrial levels (IPCC, 2019), triggering shifts in the abundance, phenology and distribution of organisms worldwide (Parmesan and Yohe, 2003; Poloczanska et al., 2013). Marine ecosystems, despite having experienced a slower warming, show comparable or even greater shift rates and vulnerability than terrestrial systems (Burrows et al., 2011; Poloczanska et al., 2013; Pinsky et al., 2019), with seagrasses, corals, cephalopods and marine mammals exhibiting the most abrupt changes (Trisos et al., 2020).

Marine mammals, as wide ranging top predators, amplify trophic information across multiple spatiotemporal scales and can therefore act as sentinels of ecosystems' responses to climate variability and change (Hazen et al., 2019). However, assessing climate change impacts in higher trophic levels and long lifespan animals such as marine mammals is challenging, as their relationships to climate may be non-linear and affected by time lags (Simmonds and Isaac, 2007; Barlow et al., 2021). In addition, identifying spatio-temporal trends in the context of climate change requires analyzing decadal or longer time series (Thorson et al., 2016), which are rarely available for marine mammal observation data.

Combining data from multiple sampling programs can help overcome this problem (Waggitt et al., 2020; Maureaud et al., 2021), but also increases the intrinsic variability related to observers' skills, sampling design and protocols, which may result in confounding species range shifts with variations in the distribution and intensity of the sampling effort (Thorson et al., 2016). For that reason, separating the observation process from the true underlying spatial distribution is essential to accurately identify range shifts over time (Chust et al., 2014b) and to identify potential drivers (Erauskin-Extramiana et al., 2019b). Recently, a species distribution function (SDF) able to distinguish between sampling variation and true geographic variability has been developed (Thorson et al., 2016). Unlike conventional estimators such as the abundance-weighted average, the SDF is applied through a Vector Autoregressive Spatio Temporal (VAST) model that allows the estimation of species distribution over predicted locations rather than sampled locations, while also estimating a standard error that allows one to distinguish between sampling variation and significant variability (Thorson et al., 2016). Although model-based approaches had been used before to estimate shifts in the distribution of species, VAST typically involves estimating a Gaussian Markov random field (GMRF) representing latent variation in density that is constant over time (a "spatial" term), as well as a GMRF representing latent variation that changes among years (a "spatio-temporal" term), which is expected to improve predictions of species density and distribution compared with using only measured habitat variables (Thorson, 2019).

Until now, this estimator has been mainly applied to commercially important fish stocks (Godefroid et al., 2019; Perretti and Thorson, 2019; Xu et al., 2019), although the fragmented and methodologically variable nature of marine mammal observations suggest the method could be highly useful for analyzing the spatio-temporal patterns of marine megafauna too. Within that context, the Bay of Biscay (BoB hereafter), located in the Northeast Atlantic, off the coasts of France and Spain (**Figure 1**), represents an interesting study area since numerous marine mammal species (e.g., cetaceans) cohabit there, attracted by a highly diverse and abundant community of pelagic fish species (Astarloa et al., 2019; Louzao et al., 2019).

Such productivity and diversity, however, might be altered by climate change in the near future, as rising temperatures (0.26°C per decade; Costoya et al., 2015) are expected to increase ocean stratification and reduce primary production and zooplankton biomass in the area (Chust et al., 2014a). In recent years, losses in fisheries production have already been reported (Free et al., 2019), together with changes in the composition, distribution, and phenology of fish species (Blanchard and Vandermeirsch, 2005; Chust et al., 2019; Baudron et al., 2020). Cetacean spatio-temporal variability, in contrast, has been mainly assessed by exploring changes in their relative abundance (Hemery et al., 2007; Castège et al., 2013; Authier et al., 2018), although both abundance and distribution are considered key criteria by the European Marine Strategy Framework Directive (MSFD; Directive 2008/56/EC) aiming to assess the environmental status of species and ecosystems in European Union waters.

Advancement of both MSFD criteria in this region is therefore necessary, especially when it is known that projections of climate change impacts on cetaceans at large spatial scales (e.g., global; MacLeod, 2009) do not always match with those at regional scales (Hazen et al., 2012). In the Northeast Atlantic, for example, warm-water cetaceans were predicted to expand poleward (MacLeod, 2009; Lambert et al., 2011, 2014), although the south-eastward shift detected for some Northeast Atlantic fish species in the BoB could indicate the opposite pattern in this particular area (Baudron et al., 2020). Indeed, some of the fish species (e.g., horse mackerel *Trachurus trachurus*, anchovy *Engraulis encrasicolus,* and sprat *Sprattus sprattus*) analyzed by Baudron et al. (2020) constitute an important food resource for many cetaceans in the BoB (Meynier et al., 2008;Spitz et al., 2018),

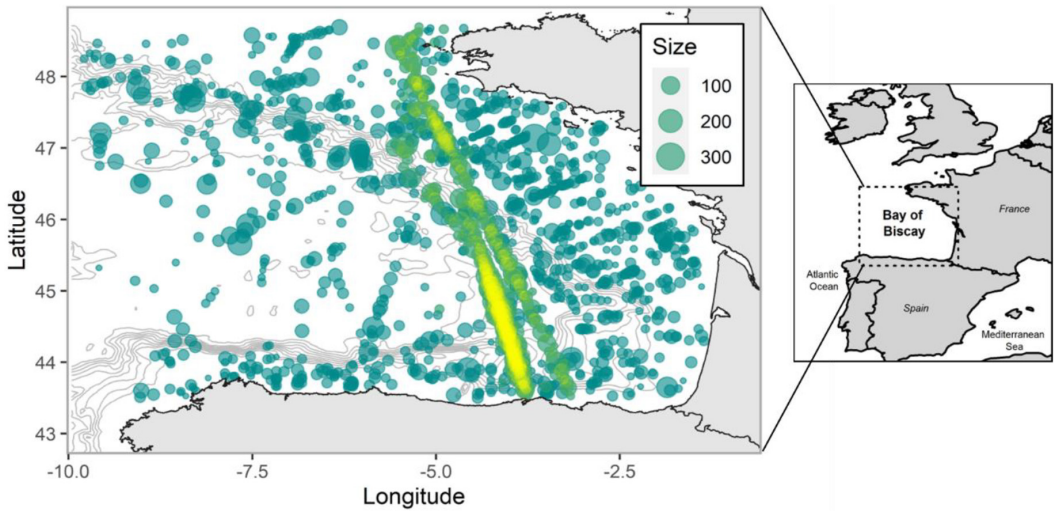

FIGURE 1 | Spatial distribution of common dolphin sightings (displayed in segments of up to 10 km) over the BoB for the 1994–2018 period. Circle sizes are proportional to group size, while solid gray lines indicate the isobaths. Sightings in yellow represent the ferry data used to check model fit.

which can heavily influence the spatial movements of their predators (Díaz López and Methion, 2019; Díaz López et al., 2019; Giralt Paradell et al., 2019).

The hypothesis that climate change may affect top predators through climate influences on their ectothermic prey has been often suggested (Robinson et al., 2005; Simmonds and Isaac, 2007; Evans and Waggitt, 2020). Most studies, however, examine environmental conditions (e.g., temperature) as proxies of prey distribution rather than studying prey data directly (Torres et al., 2008; Díaz López and Methion, 2019; Giralt Paradell et al., 2019) while others focus on exploring the effects of climate indices on the grounds that they act as an integrated measure of multiple variables (Hallett et al., 2004; Hemery et al., 2007). In the Northeast Atlantic, the North Atlantic Oscillation (NAO) is the dominant mode of climate variability, although additional climate indices such as the Atlantic Multidecadal Oscillation (AMO), the East Atlantic pattern (EA), or the South Biscay Climate (SBC) have been also found to exert strong influence, direct or indirectly, on both fish and cetacean species (Guisande et al., 2004; Hemery et al., 2007; Borja et al., 2008; Evans and Waggitt, 2020) through changes in ocean temperature and salinity, vertical mixing and circulation patterns (Drinkwater et al., 2003; Hurrell and Deser, 2009).

Given the multiple drivers potentially influencing cetacean spatio-temporal patterns, understanding the role of each of them is key for a better anticipating of future responses. For that reason, in this study we used a 25-year-long temporal series (1994–2018) to test the effect of prey biomasses, oceanographic conditions and climate indices on the abundance and distribution of the common dolphin (*Delphinus delphis),* one of the most abundant cetaceans inhabiting the BoB waters (Hammond et al., 2017). We used the VAST model (Thorson and Barnett, 2017) and the spatio-temporal species data compiled by Waggitt et al. (2020) to address two main research questions: (1) Has the abundance or the distribution of the common dolphin in the BoB experienced significant changes over the last two decades? (2) If so, are changes best explained by climatic, oceanographic, or prey variables? By answering these questions, this study intends to provide insights that will help understand past and future trends in the distribution and abundance of common dolphin in the BoB while contributing to the management for this species through the development of MSFD criteria in the context of climate change.

MATERIALS AND METHODS

Data Collection and Standardization

Cetacean data analyzed in this study, despite focusing on the BoB, belong to a large compilation made by Waggitt et al. (2020) that included observations collected on aerial and vessel (dedicated and opportunistic) surveys conducted in the Northeast Atlantic between 1980 and 2018. Although the data analyzed here (data providers in **Supplementary Table 1**) is a more updated version that includes higher-resolution tracklines (meaning that fewer data were omitted due to overlap with land-masses and more accurate measurements of distance traveled were obtained), the steps taken in the data processing and standardization stage were the same as in Waggitt et al. (2020), in which they (1) assessed differences in protocols by grouping data according to the (a) survey transect design (line transects, strip transects, and an intermediate method called ESAS, *European Seabirds At Sea*) and (b) the platform-type (vessel vs. aircraft) and (2) fitted detection functions using platform height and Beaufort sea-state as explanatory variables to estimate the proportion of animals missed by the observers (Marques and Buckland, 2004). They also assessed response bias (when animals react to the presence of the platform) through double-platform surveys that enabled the detection of animals before responsive movements. This correction was applicable to vessel surveys and is particularly

relevant to common dolphins, which typically show a positive response to vessels (Cañadas et al., 2004). Finally, they calculated the effective strip half-width (ESW) which considers the decline in the detection probability as a function of distance and covariates and serves to estimate the area effectively covered (Area covered = ESW*s*L) when including the number of observation sides (s) and transect length (L). Full details can be found in Waggitt et al. (2020).

Spatio-Temporal Pattern Detection

Sampling Effort

In order to match with the spatial resolution of the environmental data that we examined in later steps (see "*Identification of Main Drivers*" section), we divided larger transects into 10 km segments (García-Barón et al., 2019). Then, we examined the spatio-temporal coverage of surveys by summing the effort comprised in all segments per month and per year. In addition, we checked whether compiling data had led to a non-uniform distribution of sampling in space and time by exploring the annual latitudinal and longitudinal mean distributions and the corresponding linear regression trends.

Baseline Spatio-Temporal Model

Observations of common dolphin were analyzed by means of a spatio-temporal delta-generalized linear mixed model (delta-GLMM), referred to here as a VAST model (Thorson and Barnett, 2017) and available in R.[1] This model is a flexible variant of the classical delta models that decompose density into two components (Stefánsson, 1996): (1) the probability of encountering the species at a given location and time; and (2) the expected density of the species when encountered. This two-part approach, also known as a hurdle model, helps combat statistical problems with zero-inflation and overdispersion in the original data (Martin et al., 2005) and is therefore suitable for use with cetacean survey data that usually show patchy distributions (Waggitt et al., 2020).

Another feature of the VAST model is that it decomposes spatio-temporal patterns in available point-count data into multiple additive components:

1. A temporal main effect ("intercepts") representing changes in median abundance over time;
2. A spatial main effect ("spatial component") representing the average spatial distribution during the modeled interval;
3. An interaction of space and time ("spatio-temporal component") representing variation in distribution among years;
4. Density covariates, representing the impact of environmental conditions on expected density;
5. Catchability (a.k.a. detectability) covariates, representing the impact of environmental and/or sampling conditions on expected sampling data, but which do not reflect variation in population density and hence are "partialled out" prior to predicting densities.

Each of these components can be included in each of two linear predictors, and these two linear predictors are then transformed *via* inverse-link functions to predict the value of a response variable (in this case, dolphin samples). Spatial and spatio-temporal components are estimated as a Gaussian Markov random field (GMRF) and treated as a random effect. To improve computational speed, the value of these GMRFs is predicted at a fixed set of "knots" that defines a mesh of triangles that covers the entire modeled spatial domain. The value of the GMRF at any location within this domain is then predicted from the value of three knots surrounding that location. We use the stochastic partial different equation (SPDE) approximation to calculate the probability of GMRFs (Lindgren et al., 2011), and the projection from knots to locations is accomplished using bilinear interpolation as computed using R-INLA (Lindgren, 2012). The value of fixed effects are estimated using maximum likelihood techniques while integrating across the probability of random effects (Kristensen et al., 2016), and standard errors are calculated using a generalization of the delta method (Tierney et al., 1989). For further details, please see the VAST user manual.[2]

In our case, we treated year as a fixed effect (default VAST setting), such that there is no shrinkage in overall abundance across years. We modeled spatial and spatio-temporal variation as random effects to help account for multidimensional factors that are not included directly in the model but that can affect the density and distribution of the modeled species (Carroll et al., 2019). In particular, we estimated first-order autocorrelation among years in the spatio-temporal component, such that predicted hotspots in density decay slowly over time; this treatment allows spatio-temporal patterns to be predicted (with associated uncertainty) even in locations with sporadic sampling.

Detectability covariates were not considered here, because Beaufort sea-state and platform height were included in Waggitt et al. (2020). Density covariates were also omitted for our initial investigation of trends (but see "*Identification of Main Drivers*" section). As a response variable, the density of common dolphin was analyzed, after truncating the highest 5% to control outliers (Buckland et al., 2001). The spatio-temporal model was fitted assuming a lognormal error distribution and a Poisson-linked delta model such that the sum of both linear predictors is predicted log-density; this structure, was selected based on the lowest Akaike Information Criterion (AIC) (Sakamoto et al., 1986). Model parameters, as well as spatio-temporal components, were estimated using 200 knots (**Supplementary Figure 1**) based on previous studies that applied this same resolution in bigger areas (Carroll et al., 2019; Thorson, 2019), while confirming that results are qualitatively similar when increasing the number of knots (**Supplementary Table 2**). Species density was predicted at each knot by multiplying the predicted probability of occurrence by the predicted density. Density estimates for each knot were then interpolated to a standard grid of 0.1° spatial resolution (latitudinal range: 43°–49°N; longitudinal range: 1°–10°W) to match with the spatial resolution of the environmental data (see "*Identification of Main Drivers*" section) and multiplied by the

[1] https://github.com/james-thorson/VAST

[2] https://github.com/James-Thorson-NOAA/VAST/blob/main/manual/VAST_model_structure.pdf

area of the grid cell to create annual surfaces of common dolphin abundances across the BoB.

The annual abundances of common dolphin predicted for the study area were then analyzed by means of a linear regression to identify significant temporal trends and compared by means of a correlation test with an observed abundance index to check model fit. The observed abundance index was based on the encounter rate (individuals/km) of common dolphin estimated from monthly at-sea observations taken by a team of experienced observers in a constant effort-based systematic sampling scheme, i.e., the Pride of Bilbao ferry (Louzao et al., 2015; Robbins et al., 2020). This survey consistently crosses the BoB using the same route every year (**Figure 1**), and although it was also used as input for the baseline model, it only forms the 8% of the whole data set. Thus, we believe it can be used to compare the observed (ferry) and predicted (VAST) abundance indices and to determine whether the model predictions have been biased by differences in the effort.

An additional analysis with predicted abundances was also conducted to identify areas in which significant spatio-temporal changes occurred over the study period. For that, predicted abundances per grid cell were analyzed as a function of year by means of a linear regression. The slope and the *p*-value obtained in each cell, as indicators of change rate and its significance, were then plotted over the standard grid covering the study area.

Distribution Shift Metrics

Shifts in distribution were summarized by calculating the centroid of the distribution for a given year (termed center of gravity, CoG) after having predicted the density associated with every knot and year in the previous step. By means of the SDF estimator implemented in the VAST model, the CoG was calculated for the BoB population domain and standardized by the total abundance predicted for the study area, so that our analysis focused on changes in distribution after controlling for changes in total abundance (Thorson et al., 2016). Shifts in CoG were displayed in terms of "Eastings" and "Northings," meaning km from the most western point of the study area and km from the Equator, respectively. Significant trends were identified using a linear regression against year.

Identification of Main Drivers

To understand spatio-temporal patterns, three main groups of drivers were analyzed (**Table 1**), classified into local and regional covariates according to their spatio-temporal structure (a local covariate varies across space while a regional covariate is a univariate time series representing the covariate over the entire study area; Thorson, 2019):

(1) Local oceanographic conditions integrated at 100 m depth, specifically temperature and chlorophyll *a* concentration (Chl-a), based on their direct relationship with climate change and their importance for predicting top predators distribution (Hazen et al., 2012; García-Barón et al., 2020).
(2) Regional climate indices, specifically North Atlantic Oscillation (NAO), East Atlantic Pattern (EA), and Atlantic Multidecadal Oscillation (AMO) climate indices (details in **Table 1**), due to their ability to extract the leading pattern in weather and climate variability over the North Atlantic and their relationship to cetacean and prey populations (Simmonds and Isaac, 2007; Borja et al., 2008; Evans et al., 2010; Evans and Waggitt, 2020).
(3) Regional biomasses of potential prey species, based on the assumption that climate change will affect cetaceans distribution through changes in their prey (Robinson et al., 2005; Simmonds and Isaac, 2007; Evans and Waggitt, 2020).

Temperature and Chl-a values were sourced from the Iberian Biscay Irish Ocean Reanalysis Model available at the Marine Environmental Monitoring Systems,[3] providing values at a 0.08° spatial resolution, a 1-month temporal resolution and at 22 discrete depth intervals ranging from surface to 100 m depth. To test their effect on the annual estimates predicted by the baseline spatio-temporal model, the annual mean of both temperature and Chl-a was estimated integrating the data available in the first 100 m of the water column and then resampled with the *raster* package (Hijmans et al., 2017) at 0.1° (∼10 km) resolution (Waggitt et al., 2020). The three climate indices were downloaded from the National Oceanic and Atmospheric Administration (NOAA) at a monthly scale and averaged to obtain annual values,[4] while the biomass of prey species was acquired from the International Council for The Exploration of Seas (ICES) website at annual scale.[5] We selected prey species based on their relative importance in the common dolphin's diet in the BoB (Meynier et al., 2008; Santos et al., 2013) as well as data availability and suitability because not every potential prey species (e.g., sprat, myctophids) was available for the spatio-temporal scale defined in this study. European anchovy (*Engraulis encrasicolus*) was the only prey species whose biomass had been estimated exclusively for the BoB. Horse mackerel (*Trachurus trachurus*) estimates were for the Northeast Atlantic, Atlantic mackerel (*Scomber scombrus*) and blue whiting (*Micromesistius poutassou*) for the Northeast Atlantic and adjacent waters and sardine (*Sardina pilchardus*) estimates for the Cantabrian-Atlantic Iberian waters (for information on the extent of stocks see **Table 1**). Although there is an assessment for the sardine stock of the BoB, data were only available from 2000 onward (ICES, 2019c), so we decided to use the biomass estimations from the Cantabrian sea and Atlantic Iberian waters instead after having checked that both indices were highly correlated (r = 0.87) and followed similar trends (**Supplementary Figure 2**). Finally, the biomasses of all species were summed and used as a proxy for total prey biomass available in the BoB.

For modeling purposes, local temperature and Chl-a variables were included as quadratic forms in the model to allow for non-linear responses (Perretti and Thorson, 2019). Regional climate indices were included as "spatially varying coefficients" as in Thorson (2019), which means that instead of estimating a single slope parameter presenting the effect of an oceanographic index on density, the model estimates a separate slope parameter for every modeled location (every knot). The biomass of each

[3]www.ncdc.noaa.gov
[4]https://marine.copernicus.eu/
[5]https://standardgraphs.ices.dk/

TABLE 1 | Summary of the local oceanographic, regional climatic and regional prey variables used in this study accompanied by a little description and the source from which they were obtained.

	Variable	Measure	Description	Source
Local oceanographic conditions	Temperature	°C	Mean annual temperature between 0 and 100 m depth	The Iberian Biscay Irish Ocean Reanalysis Model
	Chlorophyll *a*	Mg/m^3	Mean annual chlorophyll between 0 and 100 m depth	
Regional climatic indices	NAO	–	Both NAO and EA are estimated from the difference of atmospheric pressure at sea level between the Icelandic Low and Azores High, but the anomaly centers of the EA pattern are displaced southeastward to the approximate nodal lines of the NAO pattern	NOAA (National Oceanic and Atmospheric Administration)
	EA	–		
	AMO	–	Average anomalies of sea surface temperatures	
Regional prey biomasses	Anchovy	Tons	Mean spawning stock biomass in subarea 8 (Bay of Biscay)	ICES (International Council for The Exploration of Seas): stock assessment models
	Sardine	Tons	Mean spawning stock biomass in division 8.c and 9.a (Cantabrian Sea and Atlantic Iberian waters)	
	Mackerel	Tons	Mean spawning stock in subareas 1–8 and 14, and in Division 9.a (the Northeast Atlantic and adjacent waters)	
	Horse mackerel	Tons	Mean spawning stock biomass in Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c., and 7.e–k (the Northeast Atlantic)	
	Blue whiting	Tons	Mean spawning stock biomass in subareas 1–9, 12, and 14 (Northeast Atlantic and adjacent waters)	

prey species, as well as the total biomass index, were first log transformed and then included as spatially varying coefficients since they were also available as a single regional time-series.

As a preliminary analysis, potential drivers were correlated with the abundance and CoG of common dolphin obtained in the previous baseline spatio-temporal model. Then, covariates-based modeling was performed in two different ways to identify the most parsimonious drivers and to uncover the relative contribution of covariates:

(1) Univariate spatio-temporal models were fitted for each variable using the same configuration as in the baseline spatio-temporal model. Univariate models were then compared with the baseline model by means of the AIC (Sakamoto et al., 1986). Only a decrease in the AIC > 2 in relation to the baseline spatio-temporal model was considered an improvement. When models differed by less than 2 units of AIC ($\Delta AIC \leq 2$), they were considered statistically equivalent (Arnold, 2010). The way in which covariates were related to the spatio-temporal patterns of common dolphin was also explored by plotting the functional relationships from the model parameters.

(2) Univariate models were fitted for each variable after setting the spatio-temporal variation (i.e., spatio-temporal random effects) to 0. This was done to remove the contribution of random effects and isolate the effect of the covariates since in VAST, random fields can also account for changes in distribution over time by capturing the residual spatial patterns that cannot be attributed to the fixed effect (Thorson et al., 2017). The abundances and CoG obtained from these models were then compared with those from the baseline spatio-temporal model to determine the amount of variation attributable to covariates.

RESULTS

Spatio-Temporal Patterns

Sampling Effort

A total of 1728 sightings of common dolphin collected in 21 different surveys were analyzed (**Figure 1** and **Supplementary Table 1**). Those surveys mainly covered spring-summer months and showed a peak of maximum effort between the 2007 and 2012 period (**Supplementary Figure 3**). The mean latitude of sampling also varied and shifted significantly south over time ($p = 0.001$), while no significant change was observed in the mean longitude of sampling (**Figure 2**).

Common Dolphin

The common dolphin abundance estimated by the baseline spatio-temporal model showed a significant increase ($p < 0.001$) throughout the study period, accompanied by high variability (**Figure 3** and **Supplementary Table 3**). This increase was most pronounced over the more recent years (2011–2017) and mainly occurred in the southeast corner of the BoB (**Figure 4**). These results agreed with the ferry data, which also showed an increasing trend and a significant correlation ($r = 0.7$, $p = 0.003$) with the predicted abundances (**Supplementary Figures 4**, **5**).

The CoG also showed a high interannual variability, but no significant trend was found over time in either of the two axes (**Figures 5A,B**). In contrast, the correlation between eastings and northings showed as significant pattern ($p = 0.005$) in the direction of the shift, indicating that the distribution of common dolphins generally varied between the inner (southeast) and the outer part (northwest) of the BoB (**Figure 5C**).

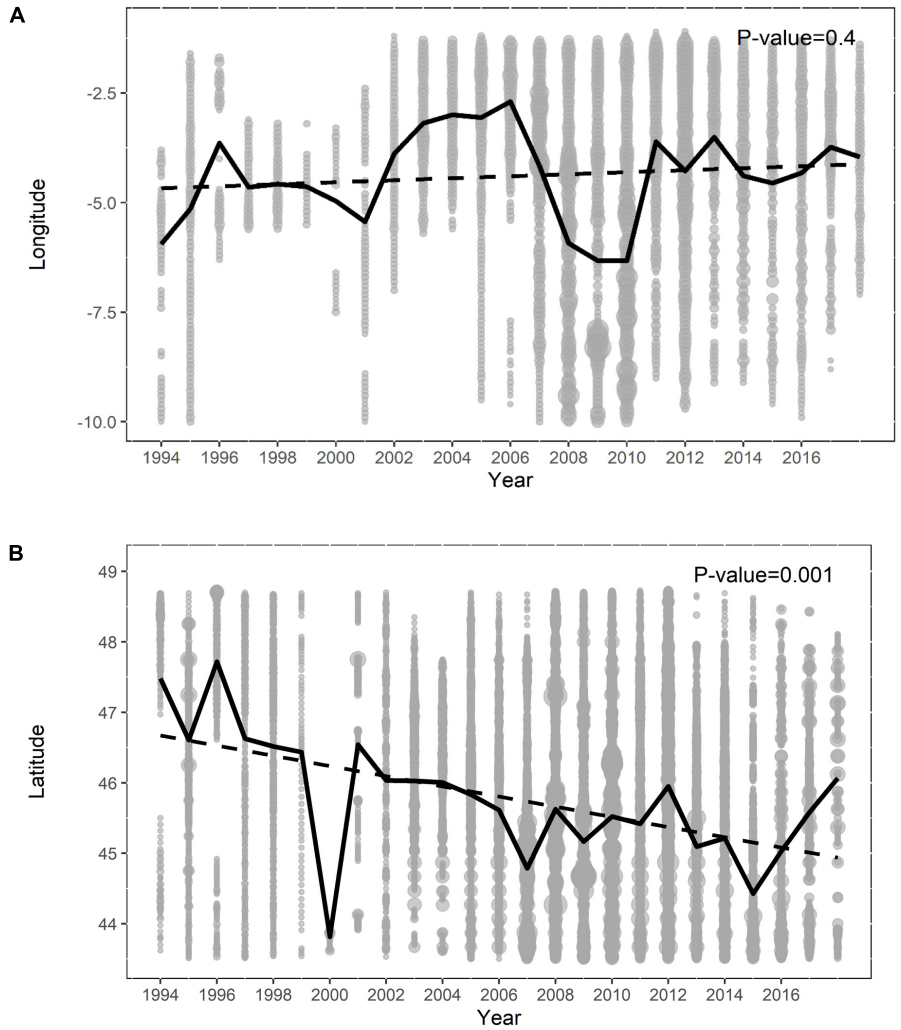

FIGURE 2 | Sampling effort (number of segments of up to 10 km) as a function of year and longitude **(A)**, and year and latitude **(B)**. In both figures the size of the circle is proportional to the sampling effort; the black line indicates the mean value and the dotted line the linear temporal trend.

Drivers and Covariate Contributions

Neither the annual temperature nor the Chl-a concentration integrated at 100 m depth revealed a significant ($p > 0.05$) temporal trend across the full BoB (**Supplementary Figure 6**). The climate index AMO has remained in a positive phase since 1997, whereas NAO and EA indices have shown a higher variability with alternation between positive and negative phases (**Supplementary Figure 7**). Both anchovy and mackerel biomasses showed a significant ($p \leq 0.05$) recovery after a period of low abundance, while sardine and horse mackerel underwent a severe decline ($p \leq 0.001$). In contrast, blue whiting did not show any significant temporal trend ($p = 0.2$). The prey biomass index, on the other hand, exhibited a significant increase ($p = 0.003$), despite the large variability (**Supplementary Figure 8**).

The correlation between the potential drivers and the CoG (easting and northings) of common dolphin only showed weak relationships. In contrast, predicted abundance revealed several strong relationships ($r > 0.5$) with prey species, specifically mackerel and anchovy (positive correlation), and sardine and horse mackerel (negative correlation) (**Figure 6**). After prey species, only EA and NAO climate indices showed a moderate correlation with abundance ($r \sim 0.40$). Blue whiting was not significant ($p > 0.05$), while temperature, Chl-a, AMO and the prey biomass index showed weak relationships ($r \sim 0.20$) (**Figure 6**).

For covariates-based models, the AIC score showed that the most substantial decrease was for the NAO index and regional prey species biomasses (especially anchovy and sardine). Local Chl-a concentration, as well as horse mackerel and mackerel, only contributed slightly, while remaining drivers (temperature, AMO, EA, blue whiting and prey species biomass index) were not relevant in terms of AIC (**Table 2**). Functional relationships of those important drivers revealed positive responses for NAO, anchovy, mackerel and negative for Chl-a, horse mackerel and sardine (**Supplementary Figure 9**).

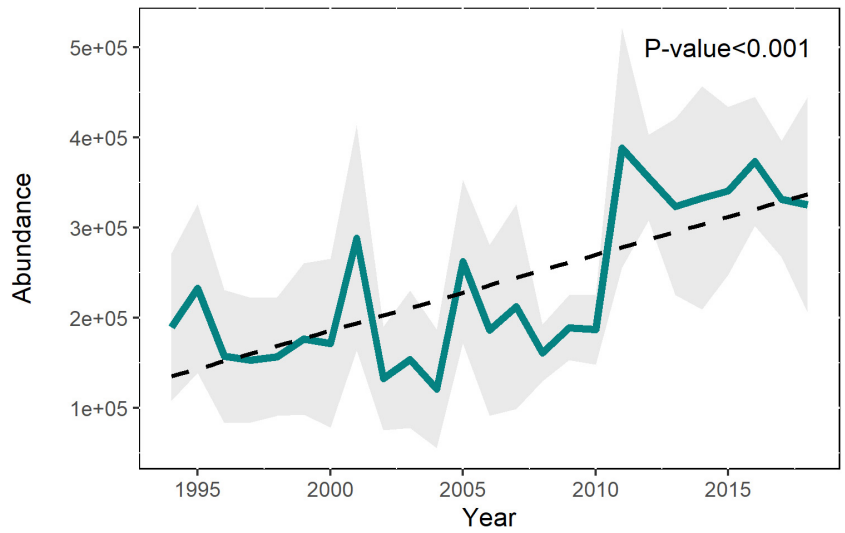

FIGURE 3 | Abundance of common dolphin in the BoB predicted by the baseline spatio-temporal model with standard deviation (shaded area), the linear trend, and its significance.

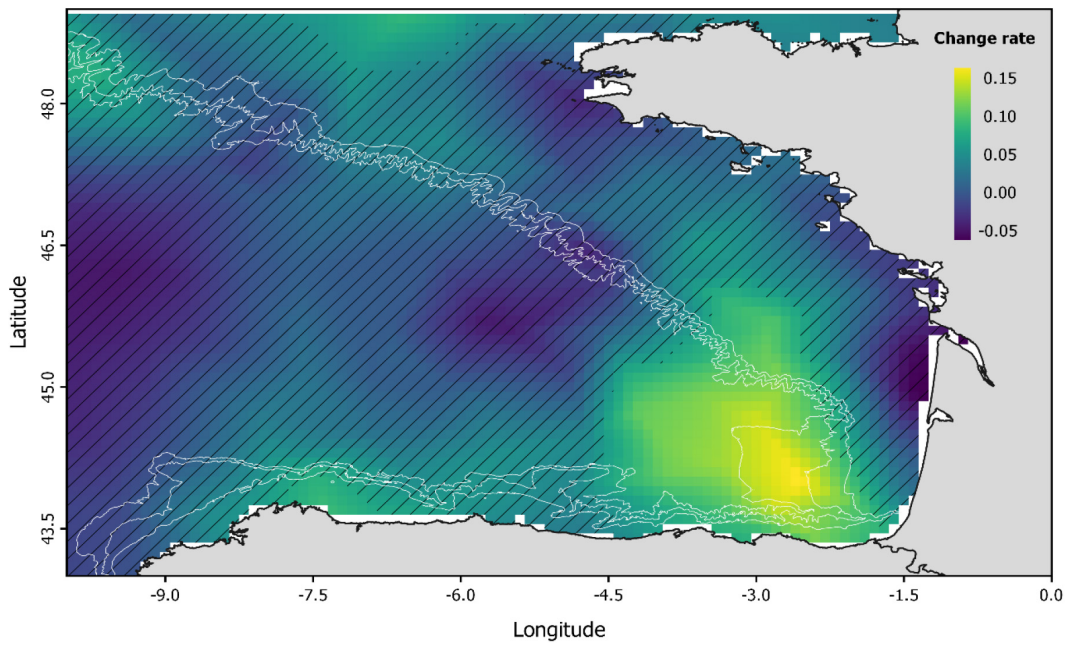

FIGURE 4 | Spatio-temporal changes in the abundance of common dolphin (predicted by the baseline model) illustrated by means of the change rate (the slope of the linear regression). Hatched areas indicate those areas where change rate is not significant ($p > 0.05$).

Similarly, covariate-only models (with no random effects) showed that the NAO index and prey species biomasses were able to explain the increase in region-wide abundance of common dolphin (**Figure 7**). Chl-a concentration, despite having shown a decrease in AIC score (**Table 2**), did not contribute to explain the variability in the relative abundance (**Figure 7**), and neither did temperature, AMO index, or blue whiting (**Supplementary Figure 10**). EA and biomass indices did show a higher contribution in terms of variability, but they were not identified as important drivers according to AIC score (**Supplementary Figure 10**).

In the case of CoG, only Chl-a and temperature contributed to explain the observed variability but, even then, only in a very small proportion (**Figure 8**). In fact, the variation in the CoG explained by these variables only accounted for about 10–20 km, while the spatio-temporal model suggested variation of 100–300 km.

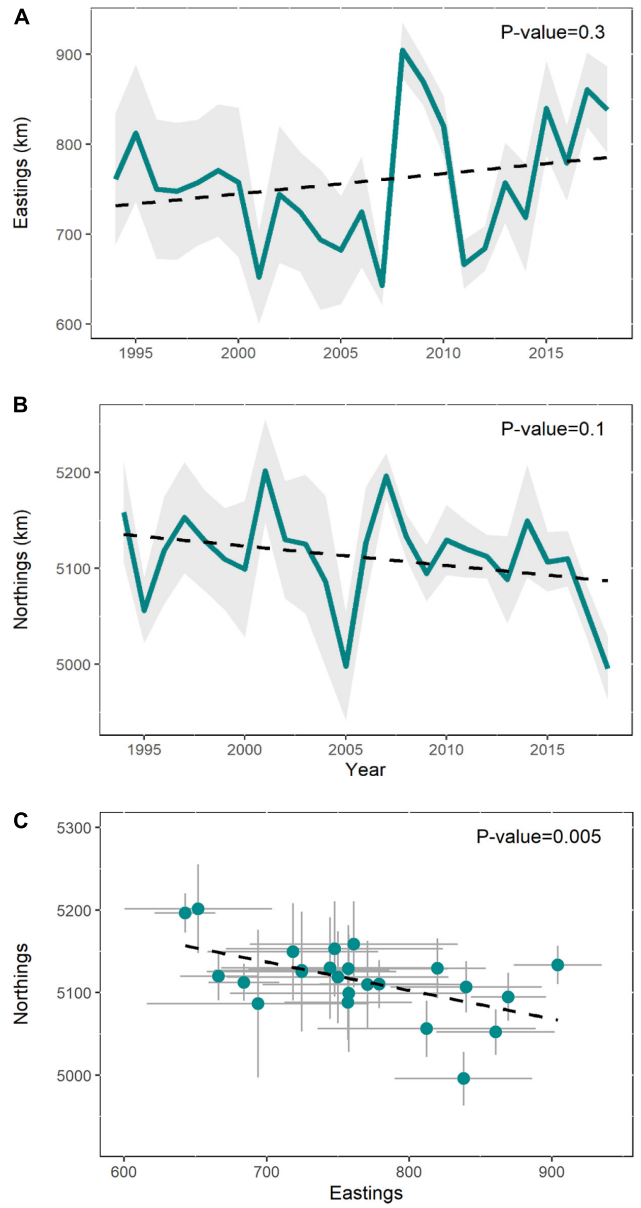

FIGURE 5 | The variation in the center of gravity (CoG) of common dolphin expressed in eastings **(A)** and northings axes **(B)**, and as a function of both **(C)**. Shaded area means the standard error, while the dashed line indicates the linear trend.

DISCUSSION

The evaluation of the spatio-temporal patterns of common dolphin in the BoB agrees with the MSFD aiming to assess the abundance and distribution of species in European waters. Surveys providing information on species distribution and abundance in this region, however, have shown significant shifts in the spatial distribution of observations, which make necessary the application of methods such as VAST to account for uneven sampling effort.

Spatio-Temporal Trends in Common Dolphin Abundance

The modeling of common dolphin sightings revealed a significant increase in abundance, which is in agreement with previous studies conducted in the BoB (Hemery et al., 2007; Authier et al., 2018; Saavedra et al., 2018) and in the wider Northeast Atlantic (Hammond et al., 2017; Evans and Waggitt, 2020) that also reported an increasing trend. In addition, data from ferry surveys, known to perform the same route every year, showed the same pattern and confirmed that the results were not biased by the detected latitudinal shift in effort.

In addition, the predicted abundance estimates were found to be quite coherent with those obtained in previous surveys conducted in summer 2012 in the BoB (Laran et al., 2017) and in summer 2016 in the Northeast Atlantic (ICES, 2020), in which 490,000 (95% CI: 340,000–720,000) small delphinids (common and striped dolphins) and 634,000 (95% CI: 353,000–1,140,000) common dolphins were estimated, respectively. Although it is not possible to make a direct comparison with our predictions, the ratios for common/striped dolphins and Northeast Atlantic/BoB estimated from Hammond et al. (2017) would lead to an approximate abundance of 360,000 (95% CI: 250,000–526,000) and 425,000 (95% CI: 237,000–764,000) individuals of common dolphin in the BoB for 2012–2016, respectively. These numbers were similar to our predictions in those years (359,000 ± 49,000 and 376,000 ± 71,500 individuals, respectively; **Supplementary Table 2**), and would indicate that, overall, abundance estimates from VAST were consistent with previous studies. This good agreement is remarkable, given the heterogeneity of the data used in this study that comprised 21 datasets, and emphasizes the importance of applying methods that are robust to shifts in sampling effort. In addition, the concordance between our results and those estimates made on summer also suggest that the spatio-temporal patterns obtained in this study should be interpreted as spring-summer trends, as this was the period of the year when most data were collected (**Supplementary Figure 2B**).

The increasing trend in abundance found in this study for the BoB, however, does not necessarily imply an overall population increase at the Northeast Atlantic level (i.e., species whole distribution range), and instead, could be due to the arrival of individuals from unsampled areas. That is why the results found in this study should be treated with caution and never be used to downplay the effects of incidental capture on common dolphin, especially when recent estimates suggest that the bycatch in the BoB is unsustainable for the population as a whole (ICES, 2020).

Regional vs. Locally Estimated Environmental Variables

Local environmental variables, such as temperature and Chl-a used in this study, are often unable to capture complex associations between environment and ecological process due to time lags in species responses coupled with the non-linear intrinsic nature of population dynamics (Hallett et al., 2004).

This can be particularly true for Chl-a and cetaceans species that feed on zooplanktivorous fishes, since the abundance of the latter has been related to a period of zooplankton grazing and

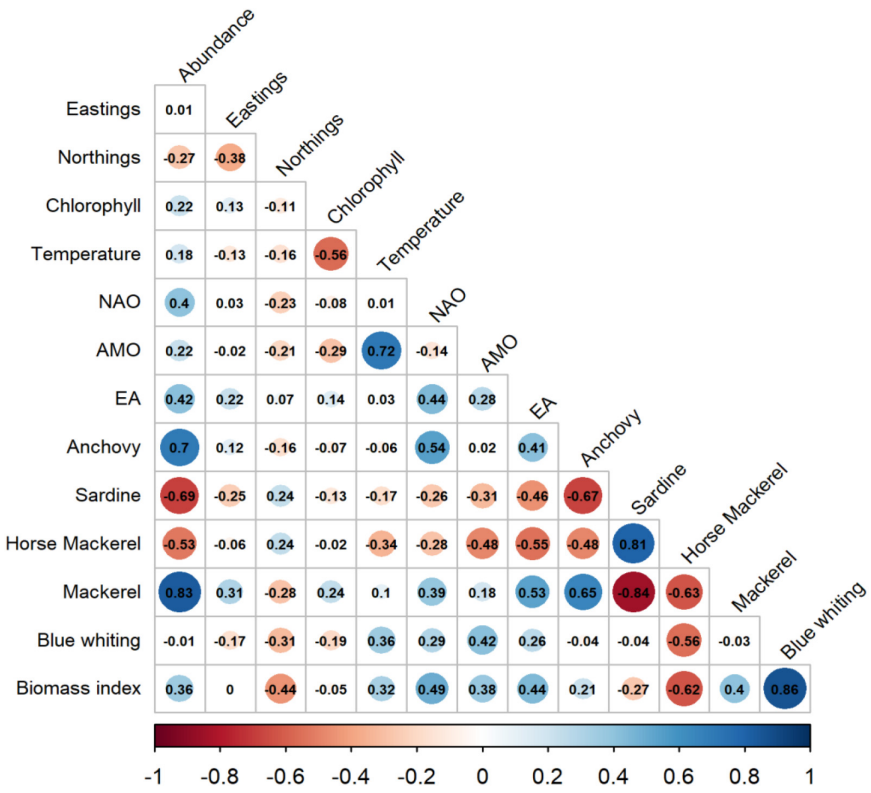

FIGURE 6 | Pearson correlation among the common dolphin's predicted abundance, CoG and potential drivers. Circle sizes are proportional to the correlation coefficient, which is indicated inside the circles. Non-significant correlations ($p > 0.05$) are shown without a circle.

a phytoplankton decay (Díaz López et al., 2019). Under such circumstances, many researchers working with cetaceans often apply time-lagged Chl-a concentration for one and/or 2 months prior to the sighting month (Tobeña et al., 2016; Prieto et al., 2017; Pérez-Jorge et al., 2020; Barlow et al., 2021).

In this study, however, predictors were introduced at an annual scale to match the available temporal scales of both prey and climatic indices, which prevented its incorporation in a lagged phase and likely led to the low contribution of Chl-a in explaining the spatio-temporal patterns of common dolphin. Similarly, the lack of importance shown by temperature could be also a consequence of this annual resolution or could instead suggest that, within the core of the species range, temperature is not such an important variable to explain its abundance and distribution.

On the contrary, regional indices of climate, spanning several months and considering wider areas of influence, are less disturbed by local variability and very often outperform locally estimated environmental variables (Hallett et al., 2004). In addition, they usually hold information about several environmental factors (e.g., temperature, storms and precipitation, mixed layer depths or circulation patterns), which make them act as an integrated measure of meteo-oceanographic conditions that tend to explain more of the variability of the system than just, for example, ocean temperature (Hurrell and Deser, 2009; Thorson, 2019).

The results found in this study are a good example of this, as the NAO climate index was found to be the best predictor explaining the abundance of common dolphin according to AIC scores. Specifically, results showed a positive relationship between both, meaning that common dolphin abundance is enhanced during positive phases of NAO, which are characterized by colder and drier conditions over Mediterranean regions, central and southern Europe (e.g., BoB), and warmer and wetter conditions in northern Europe (Visbeck et al., 2001; Aravena et al., 2009; Hurrell and Deser, 2009).

Although the NAO index and similar climate indices have been previously related to the abundance of wide ranging predators in the BoB (Hemery et al., 2007; Louzao et al., 2015), responses are likely mediated through the influence of the climate indices on food resources rather than directly on higher trophic predators such as cetaceans (Drinkwater et al., 2003; Lusseau et al., 2004). Indeed, the NAO climatic index has been related to some biologically important phenomena, such as upwelling (Pérez et al., 2010), river run-off (Dupuis et al., 2006) and Ekman transport (Guisande et al., 2004), which are known to influence the recruitment of some of the main prey species (i.e., anchovy, sardine) of common dolphin (Guisande et al., 2004; Borja et al., 2008; Planque and Buffaz, 2008). We could therefore hypothesize a potential bottom-up process, in which NAO affects common dolphins through its influence on prey. In fact, bottom-up control has been suggested for the continental shelf food web of the BoB,

TABLE 2 | Model terms.

	Model	AIC	Δ AIC
Baseline spatio-temporal	No covariates	27814.85	0
Local oceanographic conditions	Temperature	27820.78	5.93
	Chlorophyll	27811.99	**−2.86**
Regional climate indices	NAO	27806.3	**−8.55**
	EA	27816.38	1.53
	AMO	27817.57	2.72
Regional prey biomasses	Anchovy	27807.76	**−7.09**
	Sardine	27809.77	**−5.08**
	Mackerel	27812.81	**−2.04**
	Horse mackerel	27812.63	**−2.22**
	Blue whiting	27816.69	1.84
	Biomass index	27814.12	−0.73

Second column refers to the AIC score of each model, while the third column refers to the difference in the AIC (ΔAIC) resulting from the comparison of each univariate model with the spatio-temporal model (reference model). Positive values mean that higher AIC were obtained relative to the baseline spatio-temporal model while negative values mean that lower AIC scores were achieved. Numbers in bold mean improvement in model fitting (ΔAIC < −2) and hence, substantial contribution of the given variable.

where a highly diverse and abundant community of forage fishes regulates higher trophic levels (Lassalle et al., 2011).

The Role of Prey

Common dolphins are assumed to be opportunistic predators that feed on a wide variety of species, although a preference for energy-rich species, such as the anchovy, sardine, mackerel and horse mackerel investigated in this study, has been suggested (Meynier et al., 2008). Atlantic mackerel, however, is only present in large quantities during the first half of the year in the BoB, coinciding with its spawning period (Uriarte and Lucio, 2001), while Atlantic horse mackerel and the Iberian sardine are currently in serious decline (ICES, 2018, 2019b). European anchovy, in contrast, has been at a sustainable level since 2010, with an overall increasing trend that reached its maximum in 2019 (ICES, 2019a). The importance of prey species in common dolphin diet has been found to be related to their availability in terms of abundance (Santos et al., 2004; Meynier et al., 2008), which could explain the negative responses shown by species with low abundances (e.g., Iberian sardine and Atlantic horse mackerel) and the positive and larger contribution in terms of AIC made by those species with higher abundance (i.e., European anchovy). Blue whiting, on the other hand, did not seem to be relevant in explaining the variability of common dolphin over the study period, despite being more abundant than, for example, anchovy or mackerel. Evidence of blue whiting in the diet of the common dolphin was found in the BoB in the 1980s (Desportes, 1985), which could mean that it was important in the past but less so now, or that it is only important, given its poorer energetic condition (4.4 kJ g^{-1}), in the absence of other remarkable prey species (Santos et al., 2013).

However, not all potential prey species were included and differences in the distribution of stocks may have also affected the results. In fact, only anchovy's biomass had been estimated exclusively for the BoB. Remaining species biomasses were either estimated using adjacent areas (i.e., Iberian sardine) or distribution areas that extended considerably the observations range of common dolphin (i.e., blue whiting, mackerel and in a lesser extent horse mackerel), which could have contributed, for example, to the higher prominence of anchovy detected in this study.

Distributional Shifts

The common dolphin is considered a warm-temperate species, and accordingly, its range is expected to expand in response to increasing water temperature (MacLeod, 2009). This northward expansion seems to be already happening, at least at the northern limit of the species range, as evidenced by a higher frequency of strandings and sightings in northern Britain and southern Scandinavia (MacLeod et al., 2005; Evans and Waggitt, 2020). The BoB, however, does not constitute a range edge within common dolphin's distribution, which can explain why we did not find a northward shift in its CoG, but instead, switches between the inner (i.e., southeast) and the outer (i.e., northwest) part of the BoB. This pattern has also been detected when forecasting the future distribution of anchovy's egg density in the BoB for spring (Erauskin-Extramiana et al., 2019a) and was associated to the contraction (southeast) and expansion (northwest) of anchovy population (Motos et al., 1996). A prey driven distribution was already suggested for albacore tuna in the area (Lezama-Ochoa et al., 2010), so we could hypothesize that the distributional shifts of common dolphins in the BoB are also driven by the distribution of their main prey. Similarly, the increase in common dolphin abundance detected in the southeast corner of the BoB could be also related to a higher prey availability. Indeed, other important prey species of the diet of common dolphin (e.g., horse mackerel, sprat) also shifted to the southeast of the BoB in the past 30 years (Baudron et al., 2020).

The prey variables considered in this study, however, could not explain much of the observed spatio-temporal variability of the CoG as a result of being introduced as a biomass index that changed across time but not across space, and hence, could not confirm or reject the hypothesized prey-driven distribution.

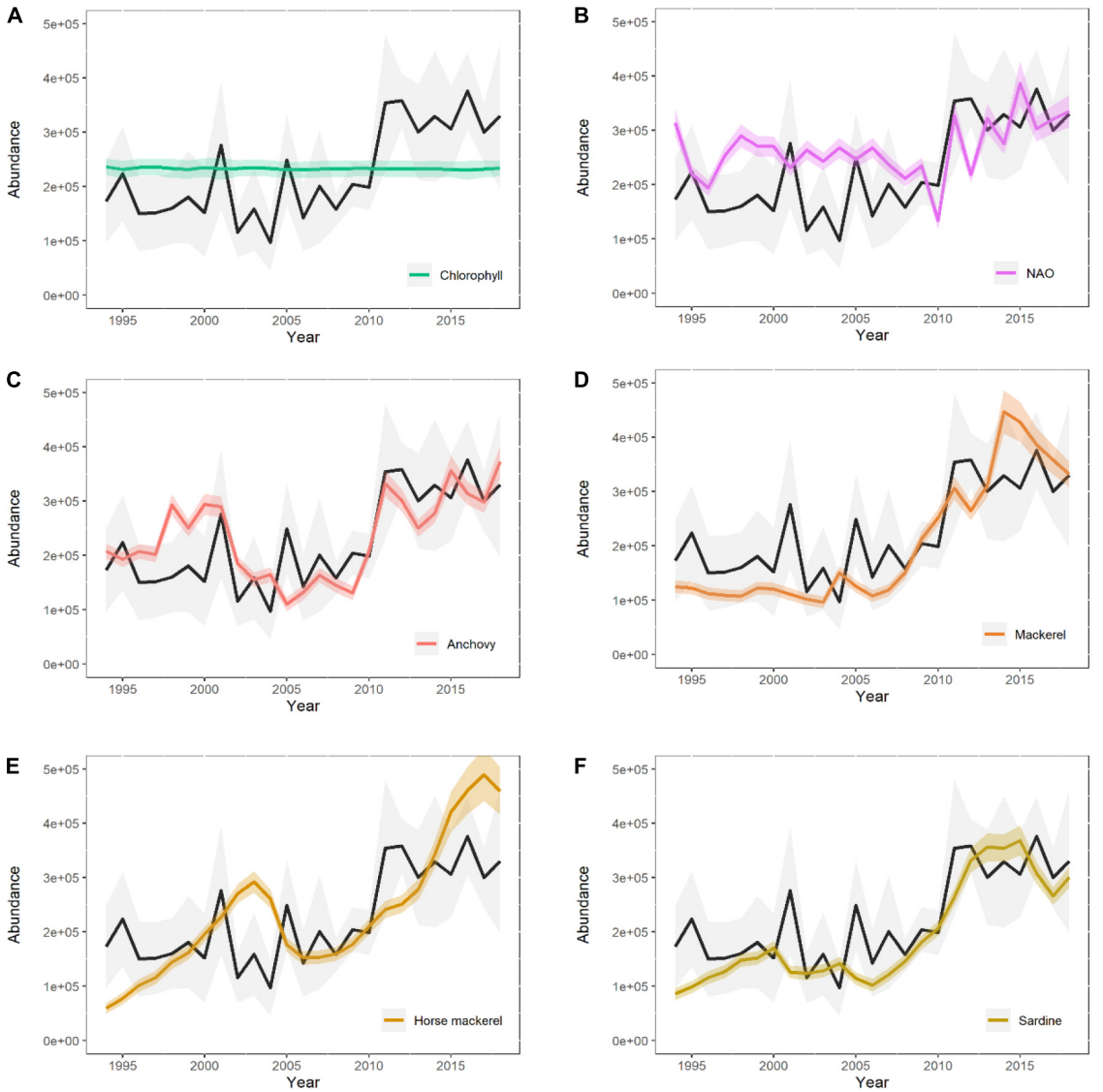

FIGURE 7 | Abundance estimates predicted by the baseline spatio-temporal model (black line) and by the covariates-based model (with no random effects, colored line) so that the contribution made by each variable **(A–F)** can be visualized. Only drivers identified as relevant by AIC score are shown.

Whether top predator abundance and distribution is driven by the environment or prey is a much debated question in ecology (Grinnell, 1917; Elton, 1927; Torres et al., 2008). However, acquiring co-occurring top predator and prey data in space and time to test these hypotheses is challenging. In this study, we have taken advantage of a large spatio-temporal compilation of top predator sightings, but in contrast, we have only been able to incorporate annual, non-spatial biomass indices of prey. Future work, therefore, should focus on improving prey data inputs to better understand their role in driving top predator distributional shifts in the BoB, a question that remains open. Climate indices, as for prey biomasses, were regional time-series rather than spatio-temporal datasets (i.e., changed across time but not across space), so their effect on the CoG is also difficult to understand. Local oceanographic variables did account for spatio-temporal changes, but even so, only explained a very small proportion of spatial shifts, which means that most of the distributional shifts occurred due to unidentified sources. This inability to attribute a source to distributional shifts was also found in previous studies with fishes (Thorson et al., 2017; Perretti and Thorson, 2019), and suggests that more effort must be made to understand when distributional shifts can be attributed to covariates in spatial random effects models (Hodges and Reich, 2010).

CONCLUSION

Climate change is believed to affect marine mammals through changes in their physical environment but also in their prey. However, many studies aimed at understanding climate impacts

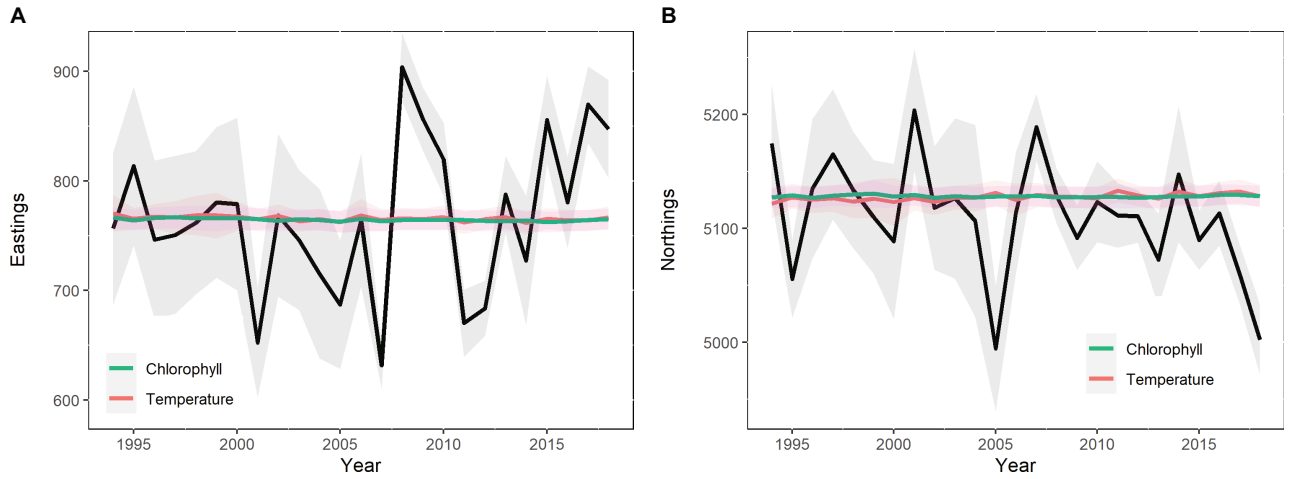

FIGURE 8 | Center of gravity estimates predicted by the baseline spatio-temporal model (black line) and by the temperature and chlorophyll-based models (with no random effects, colored line), expressed in easting **(A)** and northings **(B)** axes.

often employ environmental characteristics as proxies for prey distribution. In this study, we incorporated both environmental and prey variables estimated at local and regional scale and explored the relative importance of each of them in explaining the spatio-temporal variability in common dolphin data. Although we could not attribute much of the detected distributional shifts to the variables considered in this study, we could conclude that, in the BoB, climate indices and prey species biomasses can play an important role in driving the abundance patterns of the common dolphin.

Further research on climate change effects on common dolphin, however, should focus on comprising the whole distribution range of the species, given the increasingly feasible possibility for combining surveys across areas and regions provided by methods such as those used here. This way, we could address important knowledge gaps that have not been solved here, for example, if the increasing trend found in abundance is due to the arrival of new individuals or it is the result of an overall population growth. Answering to this question will undoubtedly help understand population dynamics and bycatch implications, but meanwhile, we reiterate our call for caution when interpreting the abundance patterns predicted in this study.

DATA AVAILABILITY STATEMENT

Any requests for survey data should be addressed to their owners. In future, some survey data may become open access. Please contact PE (peter.evans@bangor.ac.uk) for further details.

AUTHOR CONTRIBUTIONS

AA, MLo, and GC conceived and designed the study. AA applied the models and wrote the main text of the manuscript. JW collated and standardized survey data while MP and JT helped with the modeling approach. The remaining co-authors contributed survey data and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Basque and the Spanish Government through CHALLENGES (CTM2013-47032-R) and EPELECO projects and LIFE-IP URBAN KLIMA 2050 project (Grant agreement no. LIFE18 IPC/ES/000001) which has received funding from European Union's LIFE program. AA has benefited from a Basque Government scholarship (PRE_2016_1_0134) and MLo from Ramón y Cajal funding (Ministerio de Ciencia e Innovación, RYC-2012-09897).

ACKNOWLEDGMENTS

We thank the multiple organizations, institutions and surveys that provided data (ATLANCET, BIOMAN, CEMMA, CODA, ESAS, EVHOE, IBTS, IFAW, IWDG, JUVENA, KOSMOS, MARINELIFE, ORCA, PELACUS, PELGAS, SAMM, SCANS, SPEA, SWF, WDC; details in **Supplementary Table 1**), to the Bangor University, James Waggitt and Peter Evans for allowing a short visit, and to Rutgers University and the Pinsky Lab for hosting an academic exchange. We also thank Megan Ferguson and Leire Citores for their feedback and help as well as the editor and the two reviewers for their valuable suggestions and comments. This is contribution 1076 from AZTI Marine Research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.665474/full#supplementary-material

REFERENCES

Aravena, G., Villate, F., Iriarte, A., Uriarte, I., and Ibáñez, B. (2009). Influence of the North Atlantic Oscillation (NAO) on climatic factors and estuarine water temperature on the Basque coast (Bay of Biscay): comparative analysis of three seasonal NAO indices. *Continent. Shelf Res.* 29, 750–758. doi: 10.1016/j.csr.2008.12.001

Arnold, T. W. (2010). Uninformative parameters and model selection using Akaike's information criterion. *J. Wildl. Manage.* 74, 1175–1178. doi: 10.1111/j.1937-2817.2010.tb01236.x

Astarloa, A., Louzao, M., Boyra, G., Martinez, U., Rubio, A., Irigoien, X., et al. (2019). Identifying main interactions in marine predator–prey networks of the Bay of Biscay. *ICES J. Mar. Sci.* 76, 2247–2259. doi: 10.1093/icesjms/fsz140

Authier, M., Dorémus, G., Van Canneyt, O., Boubert, J.-J., Gautier, G., Doray, M., et al. (2018). Exploring change in the relative abundance of marine megafauna in the Bay of Biscay, 2004–2016. *Prog. Oceanogr.* 166, 159–167. doi: 10.1016/j.pocean.2017.09.014

Barlow, D. R., Klinck, H., Ponirakis, D., Garvey, C., and Torres, L. G. (2021). Temporal and spatial lags between wind, coastal upwelling, and blue whale occurrence. *Sci. Rep.* 11:6915. doi: 10.1038/s41598-021-86403-y

Baudron, A. R., Brunel, T., Blanchet, M. A., Hidalgo, M., Chust, G., Brown, E. J., et al. (2020). Changing fish distributions challenge the effective management of European fisheries. *Ecography* 43, 494–505. doi: 10.1111/ecog.04864

Blanchard, F., and Vandermeirsch, F. (2005). Warming and exponential abundance increase of the subtropical fish Capros aper in the Bay of Biscay (1973–2002). *C. R. Biol.* 328, 505–509. doi: 10.1016/j.crvi.2004.12.006

Borja, A., Fontán, A., Sáenz, J. O. N., and Valencia, V. (2008). Climate, oceanography, and recruitment: the case of the Bay of Biscay anchovy (*Engraulis encrasicolus*). *Fish. Oceanogr.* 17, 477–493. doi: 10.1111/j.1365-2419.2008.00494.x

Buckland, S. T., Anderson, D., Burnham, K., Laake, J., Thomas, L., and Borchers, D. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford university press.

Burrows, M. T., Schoeman, D. S., Buckley, L. B., Moore, P., Poloczanska, E. S., Brander, K. M., et al. (2011). The pace of shifting climate in marine and terrestrial ecosystems. *Science* 334, 652–655. doi: 10.1126/science.1210288

Cañadas, A., Desportes, G., and Borchers, D. (2004). The estimation of the detection function and g (0) for short-beaked common dolphins (*Delphinus delphis*), using double-platform data collected during the NASS-95 Faroese survey. *J. Cetacean Res. Manag.* 6, 191–198.

Carroll, G., Holsman, K. K., Brodie, S., Thorson, J. T., Hazen, E. L., Bograd, S. J., et al. (2019). A review of methods for quantifying spatial predator–prey overlap. *Glob. Ecol. Biogeogr.* 28, 1561–1577. doi: 10.1111/geb.12984

Castège, I., Soulier, L., Hémery, G., Mouchès, C., Lalanne, Y., Dewez, A., et al. (2013). Exploring cetacean stranding pattern in light of variation in at-sea encounter rate and fishing activity: lessons from time surveys in the south Bay of Biscay (East-Atlantic; France). *J. Mar. Syst.* 109-110, S284–S292. doi: 10.1016/j.jmarsys.2012.04.007

Chust, G., Castellani, C., Licandro, P., Ibaibarriaga, L., Sagarminaga, Y., and Irigoien, X. (2014b). Are Calanus spp. shifting poleward in the North Atlantic? A habitat modelling approach. *ICES J. Mar. Sci.* 71, 241–253. doi: 10.1093/icesjms/fst147

Chust, G., Allen, J. I., Bopp, L., Schrum, C., Holt, J., Tsiaras, K., et al. (2014a). Biomass changes and trophic amplification of plankton in a warmer ocean. *Glob. Chang Biol.* 20, 2124–2139. doi: 10.1111/gcb.12562

Chust, G., Goikoetxea, N., Ibaibarriaga, L., Sagarminaga, Y., Arregui, I., Fontán, A., et al. (2019). Earlier migration and distribution changes of albacore in the Northeast Atlantic. *Fish. Oceanogr.* 28, 505–516. doi: 10.1111/fog.12427

Costoya, X., Decastro, M., Gómez-Gesteira, M., and Santos, F. (2015). Changes in sea surface temperature seasonality in the Bay of Biscay over the last decades (1982–2014). *J. Mar. Syst.* 150, 91–101. doi: 10.1016/j.jmarsys.2015.06.002

Desportes, G. (1985). *La nutrition des Odontocètes en Atlantique Nord-Est(côtes Françaises-îles Feroë)*. Berlin: Springer.

Díaz López, B., and Methion, S. (2019). Habitat drivers of endangered rorqual whales in a highly impacted upwelling region. *Ecol. Indic.* 103, 610–616. doi: 10.1016/j.ecolind.2019.04.038

Díaz López, B., Methion, S., and Giralt Paradell, O. (2019). Living on the edge: overlap between a marine predator's habitat use and fisheries in the Northeast Atlantic waters (NW Spain). *Prog. Oceanogr.* 175, 115–123. doi: 10.1016/j.pocean.2019.04.004

Drinkwater, K. F., Belgrano, A., Borja, A., Conversi, A., Edwards, M., Greene, C. H., et al. (2003). The response of marine ecosystems to climate variability associated with the North Atlantic Oscillation. *Geophys. Monogr. Ser.* 134, 211–234. doi: 10.1029/134GM10

Dupuis, H., Michel, D., and Sottolichio, A. (2006). Wave climate evolution in the Bay of Biscay over two decades. *J. Mar. Syst.* 63, 105–114. doi: 10.1016/j.jmarsys.2006.05.009

Elton, C. (1927). *Animal Ecology*. London: Sedgwick & Jackson Ltd.

Erauskin-Extramiana, M., Arrizabalaga, H., Cabré, A., Coelho, R., Rosa, D., Ibaibarriaga, L., et al. (2019b). Are shifts in species distribution triggered by climate change? A swordfish case study. *Deep Sea Res. II Top. Stud. Oceanogr.* 175:104666. doi: 10.1016/j.dsr2.2019.104666

Erauskin-Extramiana, M., Alvarez, P., Arrizabalaga, H., Ibaibarriaga, L., Uriarte, A., Cotano, U., et al. (2019a). Historical trends and future distribution of anchovy spawning in the Bay of Biscay. *Deep Sea Res. II Top. Stud. Oceanogr.* 159, 169–182. doi: 10.1016/j.dsr2.2018.07.007

Evans, P., and Waggitt, J. (2020). Impacts of climate change on marine mammals, relevant to the coastal and marine environment around the UK. *MCCIP Sci. Rev.* 2020, 421–455.

Evans, P. G. H., Pierce, G. J., and Panigada, S. (2010). Climate change and marine mammals. *J. Mar. Biol. Assoc. U.K.* 90, 1483–1487. doi: 10.1017/S0025315410001815

Free, C. M., Thorson, J. T., Pinsky, M. L., Oken, K. L., Wiedenmann, J., and Jensen, O. P. (2019). Impacts of historical warming on marine fisheries production. *Science* 363, 979–983. doi: 10.1126/science.aau1758

García-Barón, I., Authier, M., Caballero, A., Vázquez, J. A., Santos, M. B., Murcia, J. L., et al. (2019). Modelling the spatial abundance of a migratory predator: a call for transboundary marine protected areas. *Divers. Distrib.* 25, 346–360. doi: 10.1111/ddi.12877

García-Barón, I., Santos, M. B., Saavedra, C., Astarloa, A., Valeiras, J., García Barcelona, S., et al. (2020). Essential ocean variables and high value biodiversity areas: targets for the conservation of marine megafauna. *Ecol. Indic.* 117:106504. doi: 10.1016/j.ecolind.2020.106504

Giralt Paradell, O., Díaz López, B., and Methion, S. (2019). Modelling common dolphin (*Delphinus delphis*) coastal distribution and habitat use: insights for conservation. *Ocean Coast. Manag.* 179:104836. doi: 10.1016/j.ocecoaman.2019.104836

Godefroid, M., Boldt, J. L., Thorson, J. T., Forrest, R., Gauthier, S., Flostrand, L., et al. (2019). Spatio-temporal models provide new insights on the biotic and abiotic drivers shaping Pacific Herring (*Clupea pallasi*) distribution. *Prog. Oceanogr.* 178:102198. doi: 10.1016/j.pocean.2019.102198

Grinnell, J. (1917). The niche-relationships of the California thrasher. *Auk* 34, 427–433. doi: 10.2307/4072271

Guisande, C., Vergara, A., Riveiro, I., and Cabanas, J. (2004). Climate change and abundance of the Atlantic-Iberian sardine (*Sardina pilchardus*). *Fish. Oceanogr.* 13, 91–101. doi: 10.1046/j.1365-2419.2003.00276.x

Hallett, T., Coulson, T., Pilkington, J., Clutton-Brock, T., Pemberton, J., and Grenfell, B. (2004). Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature* 430, 71–75. doi: 10.1038/nature02708

Hammond, P., Lacey, C., Gilles, A., Viquerat, S., Boerjesson, P., Herr, H., et al. (2017). *Estimates of Cetacean Abundance in European Atlantic Waters in Summer 2016 from the SCANS-III Aerial and Shipboard Surveys*. Yerseke: Wageningen Marine Research.

Hazen, E. L., Abrahms, B., Brodie, S., Carroll, G., Jacox, M. G., Savoca, M. S., et al. (2019). Marine top predators as climate and ecosystem sentinels. *Front. Ecol. Environ.* 17, 565–574. doi: 10.1002/fee.2125

Hazen, E. L., Jorgensen, S., Rykaczewski, R. R., Bograd, S. J., Foley, D. G., Jonsen, I. D., et al. (2012). Predicted habitat shifts of Pacific top predators in a changing climate. *Nat. Clim. Change* 3, 234–238. doi: 10.1038/nclimate1686

Hemery, G., D'amico, F., Castege, I., Dupont, B., D'elbee, J., Lalanne, Y., et al. (2007). Detecting the impact of oceano-climatic changes on marine ecosystems using a multivariate index: the case of the Bay of Biscay (North

Atlantic-European Ocean). *Glob. Change Biol.* 14, 1–12. doi: 10.1111/j.1365-2486.2007.01471.x

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., et al. (2017). *Package 'raster'. R package*.

Hodges, J. S., and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *Am. Statist.* 64, 325–334. doi: 10.1198/tast.2010.10052

Hurrell, J. W., and Deser, C. (2009). North Atlantic climate variability: the role of the North Atlantic Oscillation. *J. Mar. Syst.* 78, 28–41. doi: 10.1016/j.jmarsys.2008.11.026

ICES (2018). *Sardine (Sardina pilchardus) in Divisions 8.c and 9.a (Cantabrian Sea and Atlantic Iberian Waters) Advice Sheet*. Report of the ICES Advisory Commitee, 2018. Washington, DC: ICES.

ICES (2019c). *Sardine (Sardina pilchardus) in Divisions 8.a-b and 8.d (Bay of Biscay) Advice Sheet*. Report of the ICES Advisory Commitee, 2019. Washington, DC: ICES.

ICES (2019b). *Horse Mackerel (Trachurus trachurus) in Subarea 8 and Divisions 2.a,4.a,5.b,6.a,7.a-c,e-k (the Northeast Atlantic) Advice Sheet*. Report of the ICES Advisory Commitee, 2019. Washington, DC: ICES.

ICES (2019a). *Anchovy (Engraulis encrasicolus) in Subarea 8 (Bay of Biscay) Advice Sheet*. Report of the ICES Advisory Commitee, 2019. Washington, DC: ICES.

ICES (2020). *Workshop on Fisheries Emergency Measures to minimize BYCatch of Short-Beaked Common Dolphins in the Bay of Biscay and Harbour Porpoise in the Baltic Sea (WKEMBYC)*. Washington, DC: ICES.

IPCC (2019). "Technical summary," in *PCC Special Report on the Ocean and Cryosphere in a Changing Climate*, eds H.-O. PöRtner, V. Masson-Delmotte, P. Zhai, E. Poloczanska, K. Mintenbeck, M. Tignor, et al. (Geneva: IPCC).

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. (2016). TMB: automatic differentiation and Laplace approximation. *arXiv* [Preprint]. arXiv:1509.00660

Lambert, E., MacLeod, C. D., Hall, K., Brereton, T., Dunn, T. E., Wall, D., et al. (2011). Quantifying likely cetacean range shifts in response to global climatic change: implications for conservation strategies in a changing world. *Endang. Spec. Res.* 15, 205–222. doi: 10.3354/esr00376

Lambert, E., Pierce, G. J., Hall, K., Brereton, T., Dunn, T. E., Wall, D., et al. (2014). Cetacean range and climate in the eastern North Atlantic: future predictions and implications for conservation. *Glob. Chang. Biol.* 20, 1782–1793. doi: 10.1111/gcb.12560

Laran, S., Authier, M., Blanck, A., Doremus, G., Falchetto, H., Monestiez, P., et al. (2017). Seasonal distribution and abundance of cetaceans within French waters-Part II: The Bay of Biscay and the English Channel. *Deep Sea Res. II Top. Stud. Oceanogr.* 141, 31–40. doi: 10.1016/j.dsr2.2016.12.012

Lassalle, G., Lobry, J., Le Loc'h, F., Bustamante, P., Certain, G., Delmas, D., et al. (2011). Lower trophic levels and detrital biomass control the Bay of Biscay continental shelf food web: implications for ecosystem management. *Prog. Oceanogr.* 91, 561–575. doi: 10.1016/j.pocean.2011.09.002

Lezama-Ochoa, A., Boyra, G., Goñi, N., Arrizabalaga, H., and Bertrand, A. (2010). Investigating relationships between albacore tuna (Thunnus alalunga) CPUE and prey distribution in the Bay of Biscay. *Prog. Oceanogr.* 86, 105–114. doi: 10.1016/j.pocean.2010.04.006

Lindgren, F. (2012). Continuous domain spatial models in R-INLA. *ISBA Bull.* 19, 14–20.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73, 423–498. doi: 10.1111/j.1467-9868.2011.00777.x

Louzao, M., Afán, I., Santos, M., and Brereton, T. (2015). The role of climate and food availability on driving decadal abundance patterns of highly migratory pelagic predators in the Bay of Biscay. *Front. Ecol. Evol.* 3:90. doi: 10.3389/fevo.2015.00090

Louzao, M., García-Barón, I., Rubio, A., Martínez, U., Vázquez, J. A., Murcia, J. L., et al. (2019). Understanding the 3D environment of pelagic predators from multidisciplinary oceanographic surveys to advance ecosystem-based monitoring. *Mar. Ecol. Prog. Ser.* 617-618, 199–219. doi: 10.3354/meps12838

Lusseau, D., Williams, R., Wilson, B., Grellier, K., Barton, T. R., Hammond, P. S., et al. (2004). Parallel influence of climate on the behaviour of Pacific killer whales and Atlantic bottlenose dolphins. *Ecol. Lett.* 7, 1068–1076. doi: 10.1111/j.1461-0248.2004.00669.x

MacLeod, C. D. (2009). Global climate change, range changes and potential implications for the conservation of marine cetaceans: a review and synthesis. *Endang. Spec. Res.* 7, 125–136. doi: 10.3354/esr00197

MacLeod, C. D., Bannon, S. M., Pierce, G. J., Schweder, C., Learmonth, J. A., Herman, J. S., et al. (2005). Climate change and the cetacean community of north-west Scotland. *Biol. Conserv.* 124, 477–483. doi: 10.1016/j.biocon.2005.02.004

Marques, F., and Buckland, S. (2004). Covariate models for the detection function. *Adv. Distance Sampl.* 8, 31–47.

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., et al. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol. Lett.* 8, 1235–1246. doi: 10.1111/j.1461-0248.2005.00826.x

Maureaud, A. A., Frelat, R., Pecuchet, L., Shackell, N., Merigot, B., Pinsky, M. L., et al. (2021). Are we ready to track climate-driven shifts in marine species across international boundaries? - A global survey of scientific bottom trawl data. *Glob. Chang Biol.* 27, 220–236. doi: 10.1111/gcb.15404

Meynier, L., Pusineri, C., Spitz, J., Santos, M. B., Pierce, G. J., and Ridoux, V. (2008). Intraspecific dietary variation in the short-beaked common dolphin *Delphinus delphis* in the Bay of Biscay: importance of fat fish. *Mar. Ecol. Prog. Ser.* 354, 277–287. doi: 10.3354/meps07246

Motos, L., Uriarte, A., and Valencia, V. (1996). The spawning environment of the Bay of Biscay anchovy (*Engraulis encrasicolus* L.). *Sci. Mar.* 60, 117–140.

Parmesan, C., and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421, 37–42. doi: 10.1038/nature01286

Pérez, F. F., Padín, X. A., Pazos, Y., Gilcoto, M., Cabanas, M., Pardo, P. C., et al. (2010). Plankton response to weakening of the Iberian coastal upwelling. *Glob. Change Biol.* 16, 1258–1267. doi: 10.1111/j.1365-2486.2009.02125.x

Pérez-Jorge, S., Tobeña, M., Prieto, R., Vandeperre, F., Calmettes, B., Lehodey, P., et al. (2020). Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. *Divers. Distrib.* 26, 683–698. doi: 10.1111/ddi.13038

Perretti, C. T., and Thorson, J. T. (2019). Spatio-temporal dynamics of summer flounder (*Paralichthys dentatus*) on the Northeast US shelf. *Fish. Res.* 215, 62–68. doi: 10.1016/j.fishres.2019.03.006

Pinsky, M. L., Eikeset, A. M., Mccauley, D. J., Payne, J. L., and Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* 569, 108–111. doi: 10.1038/s41586-019-1132-4

Planque, B., and Buffaz, L. (2008). Quantile regression models for fish recruitment–environment relationships: four case studies. *Mar. Ecol. Prog. Ser.* 357, 213–223. doi: 10.3354/meps07274

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., et al. (2013). Global imprint of climate change on marine life. *Nat. Clim. Change* 3:919. doi: 10.1038/nclimate1958

Prieto, R., Tobeña, M., and Silva, M. A. (2017). Habitat preferences of baleen whales in a mid-latitude habitat. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 141, 155–167. doi: 10.1016/j.dsr2.2016.07.015

Robbins, J. R., Babey, L., and Embling, C. B. (2020). Citizen science in the marine environment: estimating common dolphin densities in the north-east Atlantic. *PeerJ* 8:e8335. doi: 10.7717/peerj.8335

Robinson, R. A., Learmonth, J. A., Hutson, A. M., MacLeod, C. D., Sparks, T. H., Leech, D. I., et al. (2005). *Climate Change and Migratory Species*. Thetford: British Trust for Ornithology The Nunnery.

Saavedra, C., Gerrodette, T., Louzao, M., Valeiras, J., García, S., Cerviño, S., et al. (2018). Assessing the environmental status of the short-beaked common dolphin (*Delphinus delphis*) in North-western Spanish waters using abundance trends and safe removal limits. *Prog. Oceanogr.* 166, 66–75. doi: 10.1016/j.pocean.2017.08.006

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. Dordrecht: D. Reidel.

Santos, M., Pierce, G., López, A., Martínez, J., Fernández, M., Ieno, E., et al. (2004). Variability in the diet of common dolphins (*Delphinus delphis*) in Galician waters 1991–2003 and relationship with prey abundance. *ICES CM* 9, 91–103.

Santos, M. B., German, I., Correia, D., Read, F. L., Cedeira, J. M., Caldas, M., et al. (2013). Long-term variation in common dolphin diet in relation to prey abundance. *Mar. Ecol. Prog. Ser.* 481, 249–268. doi: 10.3354/meps10233

Simmonds, M. P., and Isaac, S. J. (2007). The impacts of climate change on marine mammals: early signs of significant problems. *Oryx* 41, 19–26. doi: 10.1017/S0030605307001524

Spitz, J., Ridoux, V., Trites, A. W., Laran, S., and Authier, M. (2018). Prey consumption by cetaceans reveals the importance of energy-rich food webs in the Bay of Biscay. *Prog. Oceanogr.* 166, 148–158. doi: 10.1016/j.pocean.2017.09.013

Stefánsson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.* 53, 577–588. doi: 10.1006/jmsc.1996.0079

Thorson, J. T. (2019). Measuring the impact of oceanographic indices on species distribution shifts: the spatially varying effect of cold-pool extent in the eastern Bering Sea. *Limnol. Oceanogr.* 64, 2632–2645. doi: 10.1002/lno.11238

Thorson, J. T., and Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES J. Mar. Sci.* 74, 1311–1321. doi: 10.1093/icesjms/fsw193

Thorson, J. T., Ianelli, J. N., and Kotwicki, S. (2017). The relative influence of temperature and size-structure on fish distribution shifts: a case-study on Walleye pollock in the Bering Sea. *Fish Fish.* 18, 1073–1084. doi: 10.1111/faf.12225

Thorson, J. T., Pinsky, M. L., Ward, E. J., and Gimenez, O. (2016). Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. *Methods Ecol. Evol.* 7, 990–1002. doi: 10.1111/2041-210X.12567

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Am. Stat. Assoc.* 84, 710–716. doi: 10.2307/2289652

Tobeña, M., Prieto, R., Machete, M., and Silva, M. A. (2016). Modeling the potential distribution and richness of cetaceans in the azores from fisheries observer program data. *Front. Mar. Sci.* 3:202. doi: 10.3389/fmars.2016.00202

Torres, L. G., Read, A. J., and Halpin, P. (2008). Fine-scale habitat modeling of a top marine predator: do prey data improve predictive capacity. *Ecol. Appl.* 18, 1702–1717. doi: 10.1890/07-1455.1

Trisos, C. H., Merow, C., and Pigot, A. L. (2020). The projected timing of abrupt ecological disruption from climate change. *Nature* 580, 496–501. doi: 10.1038/s41586-020-2189-9

Uriarte, A., and Lucio, P. (2001). Migration of adult mackerel along the Atlantic European shelf edge from a tagging experiment in the south of the Bay of Biscay in 1994. *Fish. Res.* 50, 129–139. doi: 10.1016/S0165-7836(00)00246-0

Visbeck, M. H., Hurrell, J. W., Polvani, L., and Cullen, H. M. (2001). The North Atlantic Oscillation: past, present, and future. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12876–12877. doi: 10.1073/pnas.231391598

Waggitt, J. J., Evans, P. G. H., Andrade, J., Banks, A. N., Boisseau, O., Bolton, M., et al. (2020). Distribution maps of cetacean and seabird populations in the North-East Atlantic. *J. Appl. Ecol.* 57, 253–269. doi: 10.1111/1365-2664.13525

Xu, H., Lennert-Cody, C. E., Maunder, M. N., and Minte-Vera, C. V. (2019). Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. *Fish. Res.* 213, 121–131. doi: 10.1016/j.fishres.2019.01.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 14 December 2021
doi: 10.3389/fmars.2021.766915

Testing the Influence of Seascape Connectivity on Marine-Based Species Distribution Models

Giorgia Cecino[1]†, Roozbeh Valavi[1] and Eric A. Treml[1,2]*

[1] School of BioSciences, University of Melbourne, Melbourne, VIC, Australia, [2] School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Geelong, VIC, Australia

OPEN ACCESS

Edited by:
Mark J. Henderson,
United States Geological Survey,
United States

Reviewed by:
Xiaolong Yang,
National Marine Environmental Monitoring Center, China
Vera Rullens,
University of Waikato, New Zealand

***Correspondence:**
Giorgia Cecino
giorgia.cecino@fathompacific.com

†Present address:
Giorgia Cecino,
Fathom Pacific Pty Ltd., Melbourne,
VIC, Australia

Specialty section:
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

Received: 30 August 2021
Accepted: 24 November 2021
Published: 14 December 2021

Citation:
Cecino G, Valavi R and Treml EA (2021) Testing the Influence of Seascape Connectivity on Marine-Based Species Distribution Models. Front. Mar. Sci. 8:766915. doi: 10.3389/fmars.2021.766915

Species distribution models (SDMs) are commonly used in ecology to predict species occurrence probability and how species are geographically distributed. Here, we propose innovative predictive factors to efficiently integrate information on connectivity into SDMs, a key element of population dynamics strongly influencing how species are distributed across seascapes. We also quantify the influence of species-specific connectivity estimates (i.e., larval dispersal vs. adult movement) on the marine-based SDMs outcomes. For illustration, seascape connectivity was modeled for two common, yet contrasting, marine species occurring in southeast Australian waters, the purple sea urchin, *Heliocidaris erythrogramma,* and the Australasian snapper, *Chrysophrys auratus*. Our models illustrate how different species-specific larval dispersal and adult movement can be efficiently accommodated. We used network-based centrality metrics to compute patch-level importance values and include these metrics in the group of predictors of correlative SDMs. We employed boosted regression trees (BRT) to fit our models, calculating the predictive performance, comparing spatial predictions and evaluating the relative influence of connectivity-based metrics among other predictors. Network-based metrics provide a flexible tool to quantify seascape connectivity that can be efficiently incorporated into SDMs. Connectivity across larval and adult stages was found to contribute to SDMs predictions and model performance was not negatively influenced from including these connectivity measures. Degree centrality, quantifying incoming and outgoing connections with habitat patches, was the most influential centrality metric. Pairwise interactions between predictors revealed that the species were predominantly found around hubs of connectivity and in warm, high-oxygenated, shallow waters. Additional research is needed to quantify the complex role that habitat network structure and temporal dynamics may have on SDM spatial predictions and explanatory power.

Keywords: centrality measures, fragmented habitat, graph theory, machine learning, predictive model, seascape connectivity

INTRODUCTION

Conservation of biodiversity is a priority in management plans for conservation scientists and managers. Understanding species' spatial distribution patterns is critical to identify important habitats and improve management strategies (Monk et al., 2010; Foltête et al., 2012). Classic strategies used in conservation to manage species include the establishment of protected areas and reserves around key habitats. Today, connectivity is considered essential, and plays a fundamental role in characterizing the importance of protected areas within a broader network of habitat patches (Agardy, 1994). The movement of individuals among habitat patches (either as larvae or as adults), or connectivity, ensures species persistence and is critical to determine population dynamics, particularly when species are distributed across fragmented habitat patches (Hanski, 1998).

Species distribution models (SDMs) represent a key tool for the prediction of species distributions, driven by environmental parameters. SDMs have been applied to marine, freshwater and terrestrial species and demonstrated to perform well in predicting the geographic distribution of species in various contexts (Elith and Leathwick, 2009). Distribution modeling techniques have developed using presence/absence or abundance data, but recent research has focused on proposing methods which perform well when presence-only data are available (Elith et al., 2006). Though it can be hard to detect model errors and uncertainties in these cases, best practices are necessary to ensure that SDMs have strong predictive capability (Robinson et al., 2017). Correlative SDMs provide a valuable approach to predict distribution across a land/seascape, broadly applicable across diverse fields such as ecology, evolutionary biology and conservation biology (Pearson, 2007). Species distribution modeling approaches have been used to address different marine-related research goals (Robinson et al., 2017), for instance describing essential fish habitat (Monk et al., 2010), assessing the impact of climate change (Jones and Cheung, 2015), understanding habitat distribution shifts (Gormley et al., 2015), studying the spread of invasive species (Báez et al., 2010) or better designing conservation strategies (Adams et al., 2016).

Appropriate environmental parameters are crucial for the robust development and realistic predictions of SDMs, but global marine environmental datasets are often of coarse spatial resolution and coastal data are often missing or inaccurate. However, extensive work has been done to make data more reliable and available to researchers for marine species distribution modeling, such as Bio-ORACLE global environmental dataset (Tyberghein et al., 2012). Environmental parameters used in SDMs most often represent static *in situ* characteristics (e.g., annual mean temperature). But, the spatial distribution of populations is often equally as dependent on the dynamics or variability in these parameters (e.g., changes in weekly maximum temperature). In marine systems, larval dispersal is a critical component in population dynamics (i.e., population connectivity), fundamental for persistence of metapopulations inhabiting fragmented landscapes (Hanski, 1998) and in source-sink dynamics (Pulliam, 1988). Marine connectivity results from larval dispersal or adult movement and is governed by dynamic oceanic environmental variables as well as life history and biological attributes (Cowen and Sponaugle, 2009). This connectivity can largely determine the geographic range, as well as the presence/absence within habitat patches. As a result, when modeling the spatial distribution of populations, it is important to consider this dispersal as well as adult movement among habitat patches (Foltête et al., 2012). Movements of reef fishes are associated to diel movements within their home range and longer migrations toward spawning sites (Meyer et al., 2010). For fish, movements are also a density dependent process, where fish move to suboptimal habitats in response to variations in population density (Rose et al., 2001). Even though habitats might be suitable for their intrinsic environmental characteristics and potential value to the metapopulation, they might be difficult to reach and therefore not effectively contribute to the population. Habitat fragmentation can also impact connectivity, as well as species distributions. Smaller or more distant patches will be less functionally connected with surrounding habitats, increasing the isolation and vulnerability to extinction (O'Hara, 2002). In these isolated habitat patches, marine populations are often demographically closed, and species' persistence depends on replacement through local retention of larvae, whereby larvae are released and settled back to the natal habitat patch (Burgess et al., 2014). SDMs rarely directly consider dispersal of species (Robinson et al., 2011), effectively ignoring this potentially important process. Clearly, including dispersal dynamics and population connectivity into the study of species distributions is critical.

Seascape connectivity, representing the functional connectedness of marine habitat patches, combines environmental attributes and the geographic configuration of the seascape with information on the ability of the species to move (Weeks, 2017). Several studies utilize cost-surfaces incorporating the influence of ocean currents on marine species movements to determine least-cost paths connecting marine habitats of the same type, taking advantage of terrestrial examples (Caldwell and Gergel, 2013; Fischer et al., 2015). An increasingly popular approach to quantify seascape connectivity is based on biophysical models used to determine connectivity in marine systems, coupled with graph theory to study structure and properties of connectivity networks. Spatial predictions of population connectivity across the seascape are created based on habitat characteristics, ocean currents' velocity and species-specific biological parameters (Treml et al., 2008).

A well-known and appropriate framework to represent and analyze connectivity takes advantage of graph theory. Habitat connectivity, and all of its complexities, can be summarized as a network, where habitat patches are nodes and the presence and strength of connections between patches are represented by links or edges in the network (Urban and Keitt, 2001). In landscape and seascape ecology, network algorithms have been used in understanding and managing habitat fragmentation, reserve design and conservation planning (Urban and Keitt, 2001; Bodin and Norberg, 2007; Minor and Urban, 2007; Estrada and Bodin, 2008; Grober-Dunsmore et al., 2009). Few studies in landscape ecology effectively integrated graph-based metrics

into SDMs to summarize landscape connectivity, although these studies have been limited to terrestrial systems and simplified connectivity, including connectivity estimates improved the predictive performance of the SDMs (Foltête et al., 2012). These approaches have been used for terrestrial species impacted by urban development (Tarabon et al., 2019) and by linear infrastructures such as roads and railways (Clauzel et al., 2013; Girardet et al., 2013). This method focused on connectivity metrics such as recruitment, flux and betweenness centrality and predicted accurate species distributions (Foltête et al., 2012; Clauzel et al., 2013; Girardet et al., 2013). Among these metrics, betweenness centrality demonstrated to be a relevant SDM predictor (Clauzel et al., 2013). Throughout much of this work, network-based centrality measures have received much attention for summarizing patch-level connectivity attributes and determining patch-level contributions and metapopulation importance. Among these centrality metrics, betweenness centrality (BC) (Freeman, 1978; Newman, 2005) has often been used in the context of habitat prioritization and species conservation to identify stepping-stones habitats (Urban and Keitt, 2001; Bodin and Norberg, 2007; Estrada and Bodin, 2008; Bode et al., 2008; Bodin and Saura, 2010; Carroll et al., 2012). BC is defined as the number of shortest paths within an entire habitat network that pass through a given node and may indicate common or important stepping-stones habitats critical for maintaining network-wide connectivity. Eigenvector centrality (Bonacich, 1987), a similar network-wide measure of the most "influential" nodes in a network has also been used to identify important habitat patches in connected landscape networks (Estrada and Bodin, 2008) and has been shown to strongly correlate with metapopulation persistence (Watson et al., 2011). A local centrality measure, degree centrality, quantifies the number of incoming and/or outgoing connections and determines which habitat patches may act as local highly connected hubs of connectivity (Minor and Urban, 2007). These centrality measures may be ideal proxies for a habitat's connectivity importance and offer a useful pathway for integrating the connectivity process into SDMs (Foltête et al., 2012).

The main aims of this study are (i) to illustrate how centrality metrics, suitable proxies for seascape connectivity, can be incorporated in traditional marine-based SDMs and (ii) to test whether including connectivity in these models influences SDM predictions. This is the first study that uses a connectivity-enhance SDM approach in the marine environment to evaluate where, and to what degree, connectivity influences model predictions. We aim to integrate graph-based network metrics into SDMs for two types of marine species, a larval dispersing benthic invertebrate and a highly mobile pelagic fish. Here, we focused on two widely distributed marine species living across the south-east coast of Australia. This region consists of a mosaic of habitats and home to a broad group of species. We focused on the Australasian snapper *Chrysophrys auratus*, a species of fish, characterized by the ability to move across the region through the whole lifespan, and on a marine invertebrate, purple sea urchin *Heliocidaris erythrogramma*, where dispersal is limited to the larval stage. We quantify patch-level metrics using graph theory algorithms, defining centrality metrics for each habitat patch, and we integrate these metrics into our marine-based SDMs. We perform SDMs, comparing models' results and evaluating the contribution of seascape connectivity to models' performance. We assess the relative influence of centrality measures among other predictive variables identifying which metrics mostly influence SDMs. We investigate differences in the predicted geographic ranges of distribution, understanding whether these differences corresponded to critical areas for connectivity.

MATERIALS AND METHODS

Study Species

For this study we selected two representative and widely distributed species of the south-eastern Australian coast, the Australasian snapper, *Chrysophrys auratus* formerly known as *Pagrus auratus*, and the purple sea urchin, *Heliocidaris erythrogramma,* both usually associated with rocky reefs habitats (Vanderklift and Kendrick, 2004; Pederson and Johnson, 2006; Ling et al., 2010; Harasti et al., 2015; Terres et al., 2015). Snapper represents an important resource for commercial and recreational fisheries (Hamer et al., 2011). Purple sea urchin is well-known because of its role in altering coastal habitat toward a dominated urchin barren seascape (Ling et al., 2015).

Study Area

The spatial domain extends across the south-eastern coast of Australia (**Figure 1**), from the south coast of New South Wales, including Tasmania and Victoria waters, and as far west as Kangaroo Island in South Australia. This region consists of a mosaic of hard and soft bottom habitats, populated by a range of diverse species. It spans from warm temperate waters in the north to cooler waters around Tasmania. This region is also important in terms of conservation values, including both protected species and protected areas (Bax and Williams, 2000; Commonwealth of Australia, 2015). Overall, the south-east Australian waters have low productivity, however, there are localized spots of high productivity at the edges of the continental shelf, where the effects of currents, eddies and upwelling creates a rich habitat, that is fished commercially and recreationally. In this work we focused on the coastal areas of this region, which consists of rocky reefs and soft sediments supporting a broad range of species (Commonwealth of Australia, 2015).

We identified habitat patches using data available through Seamap Australia National Benthic Habitat Classification Scheme (Butler et al., 2017). Data from this dataset were downloaded at state-resolution then merged. The extent of the study domain is 1,990 km × 1,850 km. We selected only habitats that are classified as rocky reefs contained in the domain area, and we aggregated habitat patches that showed a very limited size (of order of less of 1 km^2) into a single patch, where possible. We defined 236 rocky reefs patches across the whole region for an extent of 15,248 km^2 of available habitat for the sea urchin and 264,050 km^2 for the snapper, which includes reefs surrounding area that could be used by this species.

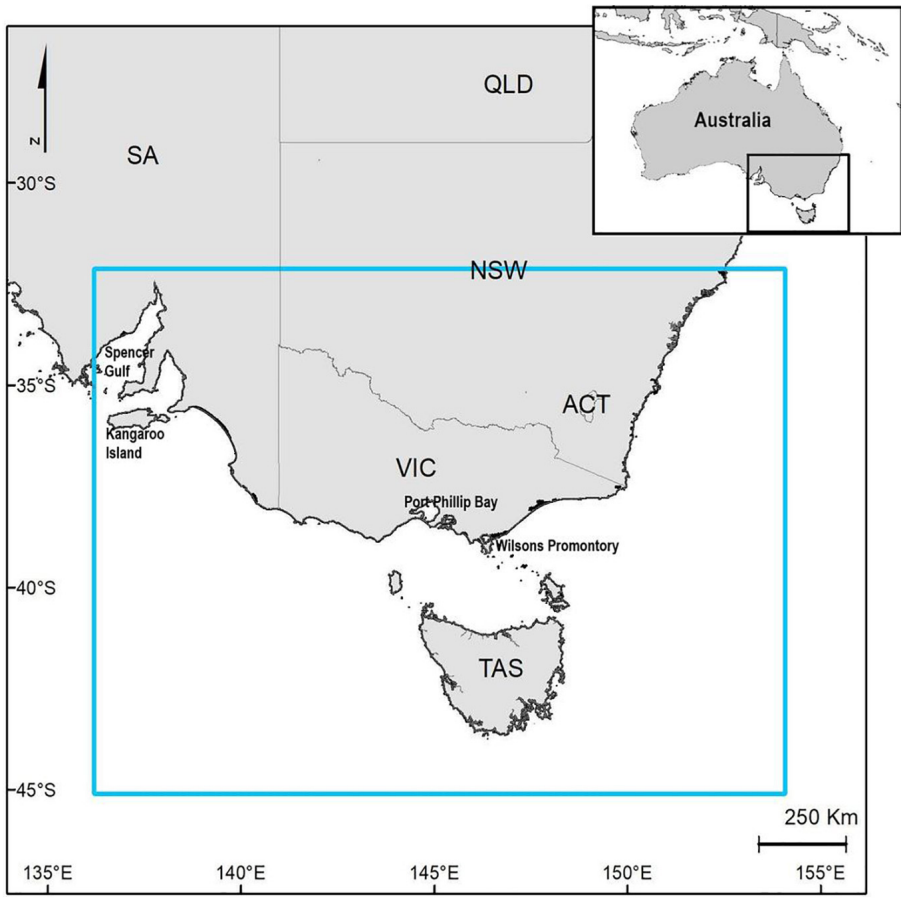

FIGURE 1 | Map of the study area.

Seascape Connectivity for Snapper (*C. auratus*)

To model adult snapper movements across the seascape and quantify habitat connectivity, we (1) built a cost surface layer based on magnitude and direction of oceanic currents, and (2) completed a least-cost path analysis to quantify seascape connectivity. The cost surface, required for the least cost path analysis (LCP), assumed fish movement was influenced by the magnitude and the direction of currents, with least cost following the direction of currents. Magnitude and direction of currents were derived from a global ocean circulation model (HYCOM)[1] using the Marine Geospatial Ecology Toolbox, MGET (Roberts et al., 2010) in ArcGIS® 10.5.1 (ESRI, 2017). Data were aggregated into single annual cumulative cost layers, representative of currents magnitude and currents direction for the entire region (see **Supplementary Figure 1**). Following examples from terrestrial habitat, first we created two cost surfaces, one for currents' magnitude and one currents' direction, quantifying the increasing relative cost of moving across the seascape. Generally, due to the dominant eastward flow of currents, the cost of moving in this direction was less than traveling westward (Caldwell and Gergel, 2013). We reclassified both layers and assigned a relative score representing the cost of traveling (Rayfield et al., 2010), ranging from 1 to 10, with a score of 1 representing the least cost, while 10 represented the greatest cost of travel, ten times more costly compared to cells with a value of 1. Finally, we combined the currents magnitude and currents direction cost surfaces, calculating the weighted mean and defining one cumulative movement cost surface among all study area, assuming parameters have equal weight. See **Supplementary Material** for further details.

We performed LCP analysis using Linkage Mapper 2.0.0 (McRae and Kavanagh, 2011) a toolbox freely available for ArcGIS® 10.5.1 (ESRI, 2017). To add realism to the model, we applied a maximum threshold of 100 km of traveled distance, based on maximum swimming linear distances recorded from acoustic tagging of snapper in South-east Australia (Fowler et al., 2017). We modeled only ecologically meaningful corridors among all habitat patches within the swimming range of snappers. Our LCP analysis resulted in maps representing seascape connectivity for adult snapper, with routes showing the least costly paths among all habitat patches (nodes). This LCP network was used to further quantify the structure of seascape connectivity (see **Supplementary Figure 3**).

[1]https://www.hycom.org

Seascape Connectivity for Purple Sea Urchin (*H. erythrogramma*)

For marine invertebrates such as *H. erythrogramma*, movements across the seascape are largely determined by the larval dispersal phase. We modeled larval connectivity using an existing spatially explicit biophysical marine connectivity model (Treml et al., 2012). In this model, we used (1) a map defining suitable rocky reef habitat patches, same data as above for snapper, where all the habitat patches are source and destination sites for larvae, (2) data describing the ocean currents (HYCOM), and (3) species-specific life history traits for *H. erythrogramma* (**Supplementary Table 1**), obtained from the literature (Okubo, 1971; Rumrill, 1987; Lamare and Barker, 1999; Huggett et al., 2008; Swanson et al., 2012; Williams and Hastings, 2013).

We simulated larval dispersal from 1992 to 2012 at a 3-hourly time-step, using all the available data for all spawning times. Clouds of larvae were released from source reef patches and the likelihood of larval settlement to all destination patches was estimated based on species-specific biological parameters and ocean characteristics. The model output was a dispersal matrix, recording the cumulative quantity of larvae released from each source patch that survived and settled to each destination patch, summarizing across all modeled dispersal events and years, and scaled by the size of the available habitat area. Migration matrices are commonly used to study larval connectivity, for this reason the dispersal matrices were converted to migration matrices, M, representing the proportion of settled larvae arriving at each destination (columns in the matrix) that came from each source patch (rows in the matrix). The migration matrix was used to build a network of seascape connectivity, where rocky reef patches correspond to graph nodes and presence of larval connectivity was represented as graph edges (**Supplementary Figure 4**).

Network Analysis and Spatial Generalization of Centrality Measures

The species-specific connectivity data were used to quantify patch-level metrics representing patch importance, a common spatial ecology approach (Estrada and Bodin, 2008; Bodin and Saura, 2010; Carroll et al., 2012). All metrics were calculated in R (R Core Team, 2019) with the "igraph" package (Csardi and Nepusz, 2006). For all patches, we calculated degree centrality, betweenness centrality and eigenvector centrality. Centrality metrics indicate how central a node is in a network, therefore a node with a high value of centrality is expected to have high habitat connectivity importance. Degree centrality is the number of outgoing and incoming links with each node in the network. Betweenness centrality is a measure based on shortest paths, and it is calculated as the number of shortest paths between all pairs of nodes in the graph that pass through that node (Freeman, 1978; Newman, 2005). Eigenvector centrality is a measure of importance of a node, essentially identifying highly connected nodes that are also connected to other highly connected nodes. Compared to other centrality metrics eigenvector centrality values are defined between 0 and 1, where a value of 1 is assigned to the most influential node in the network and 0 to the least influential. This metric assigns relative scores to all nodes in the network and is estimated as the principal eigenvector of the adjacency matrix defining the network (Borgatti, 2005).

Species distribution models require continuous explanatory variables, therefore we interpolated our centrality estimates across the seascape domain. The interpolation technique and distance used was dependent on each species' capacity to move throughout the seascape. In the case of the purple sea urchin, with its limited ability to move great distances following settlement, centrality values were interpolated locally only and assigned to all habitat cells in the focal rocky reef patch. For the snapper we assigned the corresponding centrality value to each patch cell, and due to the likelihood of movement at greater distances, we extrapolated the centrality measure into the neighboring seascape using a negative exponential function with respect to distance. Consistent with the 100 km threshold used in the LCP analysis (Fowler et al., 2017), a maximum dispersal distance of 100 km corresponds to a probability of presence of $p = 0.05$ (Urban and Keitt, 2001; Foltête et al., 2012). We multiplied this probability by the centrality value of the habitat patch. Where values from two or more patches intersect, the mean centrality value was used in these intervening areas. The results are continuous centrality surfaces which can then be appropriately integrated into SDMs.

Species Distribution Modeling and Comparison of Models' Performance

We developed SDMs for both species, including and excluding the species-specific centrality surfaces. Species occurrences data recorded inside our spatial domain were derived from the Atlas of Living Australia [ALA] (2019, 2020)[2], and contained reliable occurrence data for species around Australia. Environmental parameters were extracted using the "Bio-Oracle" package in R, which contains many marine data layers for ecological modeling (Tyberghein et al., 2012). Given that ALA data cover a temporal period of more than 100 years, we cleaned the data set to remove the oldest data and duplicates to better align to the temporal extent of the environmental data. For both species we selected only data from 1980. The final presence data for purple sea urchin consists of 875 observations, distributed across the study area, with the largest concentration within Port Phillip Bay, while snapper occurrence data consists of 780 observations, distributed across Victoria, South Australia and northern Tasmania, reflecting the known habitat of this species.

We selected a group of ecologically important parameters which were believed to contribute to the distribution of both species. Temperature, chlorophyll A concentration, primary production (measured as net primary productivity of carbon), current velocity, dissolved oxygen data were summarized by the long-term monthly mean, pH, bathymetry, and salinity were downloaded from models or summarized *in situ* measurements (for additional information see **Supplementary Material**). In addition to these environmental data, we included the seascape connectivity layers of betweenness centrality, degree centrality and eigenvector centrality (**Figure 2**). Collinearity among

[2]https://www.ala.org.au

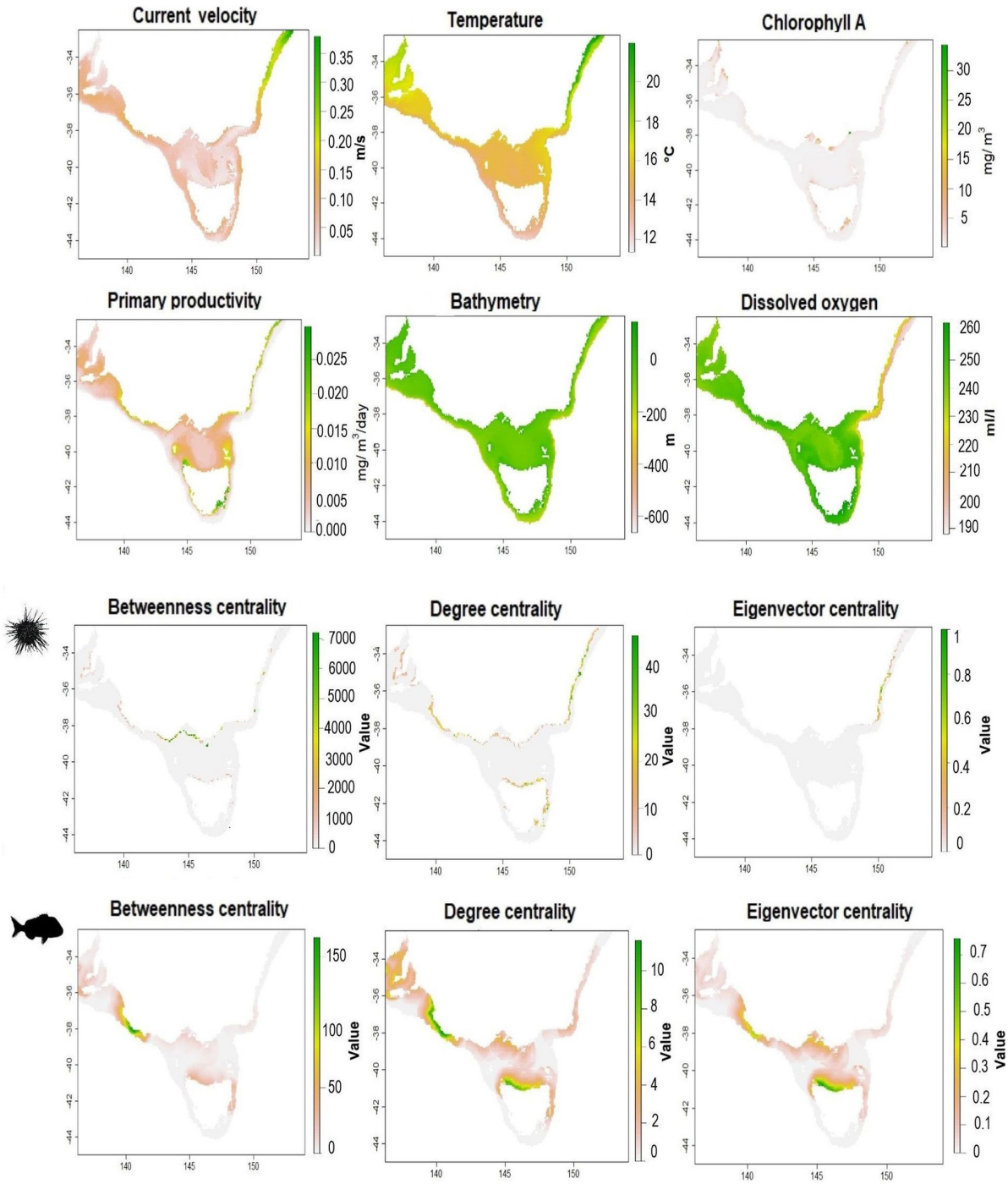

FIGURE 2 | Maps of environmental and connectivity predictors used in the SDMs for purple sea urchin, *H. erythrogramma* **(top)**, and snapper, *C. auratus* **(bottom)**. All Maps in WGS84.

predictors was quantitatively checked, and those with a Pearson correlation threshold of 0.7 or greater were identified and one was eliminated leaving the most ecologically meaningful parameter in the model. Collinearity is a known source of uncertainty, and when collinearity increases, the efficiency and statistical power of the model decrease (De Marco and Nóbrega, 2018).

Among the many SDMs algorithms available, we used a popular machine learning method, boosted regression trees (BRT) (Elith et al., 2008). BRT is a form of logistical regression using decision trees and a boosting algorithm, an optimization technique that reduces predictive deviance by combining numerous trees into a single model. BRT has a powerful

predictive performance and it has features such as handling different type of predictors, missing data, moderate collinearity, and complex non-linear relationships (Elith et al., 2008; Cimino et al., 2020). BRT also performs well for presence-only data (Elith et al., 2006). To model presence-only data with machine learning methods, a random sample of the background landscape is taken to represent unavailable "absence" data. In each model we defined 10,000 pseudo-absences, distributed equally across the coastal areas within the study area, representative of the potential species habitat. We followed the protocol for fitting BRT established within the "dismo" package (Hijmans et al., 2017) in R, to understand which was the best predictive model and compare the significance of including seascape connectivity in the models. We built a training dataset and a test dataset by resampling presence and background data, allocating them to cross validation (cv) folds. Evaluation was completed at two levels, first we used 10-folds to evaluate the models, then for each training fold a 5-folds internal cross validation procedure was completed for tuning the parameters of the BRT model using the "dismo" R package (Hijmans et al., 2020). Models' settings (**Table 1**) were selected according to the recommendations in the literature (see Elith et al., 2008). The selected settings directly affect the number of optimal trees. As a result, by keeping the learning rate and tree complexity constant, we can optimize the number of trees to fit a good model. The settings were selected to aim for a model with a high number of trees, so the model can reliably estimate our response (Elith et al., 2008). We used cross-validation to evaluate the predictive power of the models and assessed performance using AUC-ROC, or area under the curve – receiver operating characteristics curve approaches. Then, we quantified the relative influence of seascape centrality metrics with respect to all other environmental variables to assess their contribution in predicting species distributions. We also quantified pairwise interactions between environmental and connectivity variables and environmental variables themselves, which is useful to define the most suitable environment for the species (Elith et al., 2008). BRT automatically models predictor interactions, allowing their magnitude, and therefore ranking, to be calculated (Hastie et al., 2009). Interaction results can be visualized as three-dimensional partial dependence plots.

For each species we present results for two connectivity-enhanced models compared to a model without connectivity. First, we investigated the effect of connectivity adding to the model all centrality metrics (degree, betweenness, and eigenvector centrality) to understand which metric has the largest influence and we compared it to the model without centrality metrics. Then, to minimize overfitting and maximize predictive performance (Duan et al., 2014), we selected the single centrality metric with the largest relative influence in the model to remain in the model during fitting and we explored the SDMs results when we included or excluded connectivity. These additional models help to understand the role of connectivity and whether the SDM predictions were influenced by the number of connectivity parameters included in the model.

Finally, we mapped the spatial distribution of species across the study area to visualize and quantify differences in spatial predictions. To evaluate if there is a statistical relationship between centrality measures and models' predictions, we tested them for correlation. Spatial indicators were used to quantify the differences in predicted habitat suitability from integrating connectivity or excluding connectivity. An overlay analysis was performed to identify areas within the SDM predictions that corresponded to critical connectivity areas revealed in the network analysis.

TABLE 1 | Boosted regression trees settings applied to all models.

Model settings	
Tree complexity	5
Learning rate	0.005
Bag fraction	0.75
Maximum trees	10,000

RESULTS

Seascape Connectivity

We estimated seascape connectivity for both species (**Supplementary Figures 2, 3**) and no consistent spatial trend existed between species, revealing different connectivity structures across the seascape, according to species-specific dispersal characteristics. All centrality measures showed some spatial consistency within species, identifying similar areas of high and low values, revealing that hubs of connectivity (high degree centrality), populations stepping-stones (high betweenness centrality) and critical nodes (high eigenvector centrality) largely matched and were clustered in similar locations. Purple sea urchin showed well-connected areas, high eigenvector centrality and degree centrality, across north and east Tasmania, eastern Victoria, and New South Wales coast, while South Australia nodes had weak connections with the rest of the domain (**Figure 2**). Purple sea urchin stepping-stone habitats are clustered in central and eastern Victoria. Snapper connectivity revealed high values of centrality for patches along north of Tasmania and central Victoria coasts, while areas on the eastern and western boundaries of the domain, along South Australia and New South Wales coasts, showed less connectivity (**Figure 2**).

Species Distribution Modeling and Comparison of Models' Performance

The final SDMs included mean sea water temperature, chlorophyll A concentration, primary production, bathymetry, dissolved oxygen concentration, current velocity and centrality measures (**Supplementary Table 2**). Salinity and pH were removed for both species, due to strong correlation with other environmental variables, specifically salinity was highly correlated with temperature while pH was highly correlated with temperature and current velocity. Note that centrality measures displayed low correlation with the environmental variables included in the models, although they displayed greater correlation between centrality metrics, especially for snapper (**Supplementary Figures 4, 5**).

The optimal models' results are summarized in **Table 2**. For both species, SDMs used a tree complexity of 5, a learning rate of 0.005, bag fraction of 0.75, 5 folds for tuning and a maximum of 10,000 trees. The optimal model for sea urchin used 4,300 trees for the model integrating all centrality metrics, 4,700 trees for the model including degree centrality only and 5,700 for the model without centrality metrics. Models showed good predictive performance with same mean AUC score (0.95 ± 0.01) for all models (with all centrality metrics, degree centrality only and without connectivity). The optimal model for snapper used 4,000 trees for the model integrating centrality metrics, 3,300 trees when only degree centrality is included in the model and 3,200 trees when seascape connectivity was excluded. The mean AUC score was 0.91 ± 0.03 for the models including all centrality metrics and degree centrality only while it was slightly lower (0.90 ± 0.03) in the model without connectivity.

Environmental variables emerged as the most influential predictors for both species. For the sea urchin bathymetry showed the largest influence in all models, respectively, contributing between 25 and 30% to SDMs predictions, followed by temperature and dissolved oxygen which had a relative influence between 13 and 17% across the three sea urchin models (**Figures 3A–C**). Primary production, chlorophyll A and current velocity had a lower contribution, with relative influence varying between 7 and 15% (**Figures 3A–C**). For snapper, temperature was the most influential variable contributing between 36.5 and 40% to SDMs predictions in all the models (**Figures 3D–F**). Other environmental variables that showed an important relative influence for snapper were current velocity (13.2, 15.2, and 17.2%), followed by chlorophyll A (10.1, 11.6, and 13.3%). Dissolved oxygen, primary production and bathymetry were less influential with relative influence values varying from 7 to 11.5% (**Figures 3D–F**).

Centrality measures had some influence across both species with degree centrality emerging as the most important centrality measure. For the purple sea urchin SDM, connectivity contributed to a total of 18.6% to the final model, with degree centrality having the largest relative influence (8.2%), followed by betweenness centrality (7.2%) and eigenvector centrality (3.2%) (**Figure 3A**). Degree centrality was more influential than current velocity (7.3%) and similar to chlorophyll A concentration (9.2%). Centrality measures showed pairwise interactions with several of the environmental variables (see three-dimensional dependence plots **Supplementary Figures 7, 8**). Eigenvector centrality had the strongest interactions with current velocity and primary production, degree centrality had interactions with primary production and bathymetry, while betweenness centrality interacted with temperature and bathymetry. For snapper, all centrality measures had a lower relative influence than the environmental parameters, and contributed at most 17% to SDM predictions. Degree centrality was the most influential among the centrality metrics, with a relative influence of 6.9%, followed by eigenvector centrality (6.4%) and betweenness centrality (3.6%) (**Figure 3D**). Centrality measures interacted with environmental variables, and the strongest interactions were with temperature for degree centrality and eigenvector centrality, and bathymetry for betweenness centrality (**Supplementary Figures 7, 8**).

We selected only the network-based metric with the largest influence to reduce the number of variables and increase the predictive power. Degree centrality was selected for both species, and we therefore compared the SDM results between models with and without degree centrality included (**Table 2**). Note that a lower number of predictors is expected to result in an overall increase in relative influence across all variables. In the purple sea urchin model, the relative influence of degree centrality was maintained among models, more influential than current velocity and similar to chlorophyll A concentration, primary production and dissolved oxygen (**Figure 3B**). For the sea urchin, the order of variables based on relative influence remained the same for the model including all centrality measures (e.g., **Figures 3A** vs. **B**), and degree centrality maintained a similar order of influence, comparable to chlorophyll A and well above the influence of current velocity. Degree centrality had some interactions with all the environmental parameters, but the strongest interactions were with bathymetry and high dissolved oxygen (**Supplementary Figures 7, 8**). In the snapper model the order of influence changed when only degree centrality was used, moving ahead of primary productivity and bathymetry in influence. The relative influence of degree centrality, when used as the sole connectivity metric moved in front of both primary productivity and bathymetry, and was comparable in influence to dissolved oxygen concentration (**Figure 3E**). Degree centrality interacted with all environmental variables, particularly with warm temperature and high bathymetry (**Supplementary Figures 7, 8**). Across all models and both species, connectivity metrics appeared to maintain a relative influence between 9.5 and 18.5% on species distributions.

We predicted species distribution and compared maps of habitat suitability, highlighting differences in species range

TABLE 2 | Optimal SDMs models results for each species.

	SDM including all connectivity variables	SDM including degree centrality only	SDM excluding all connectivity variables
Purple sea urchin			
Number of trees	4,300	4,700	5,700
Mean AUC score	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
Snapper			
Number of trees	4,000	3,300	3,200
Mean AUC score	0.91 ± 0.03	0.91 ± 0.03	0.90 ± 0.03

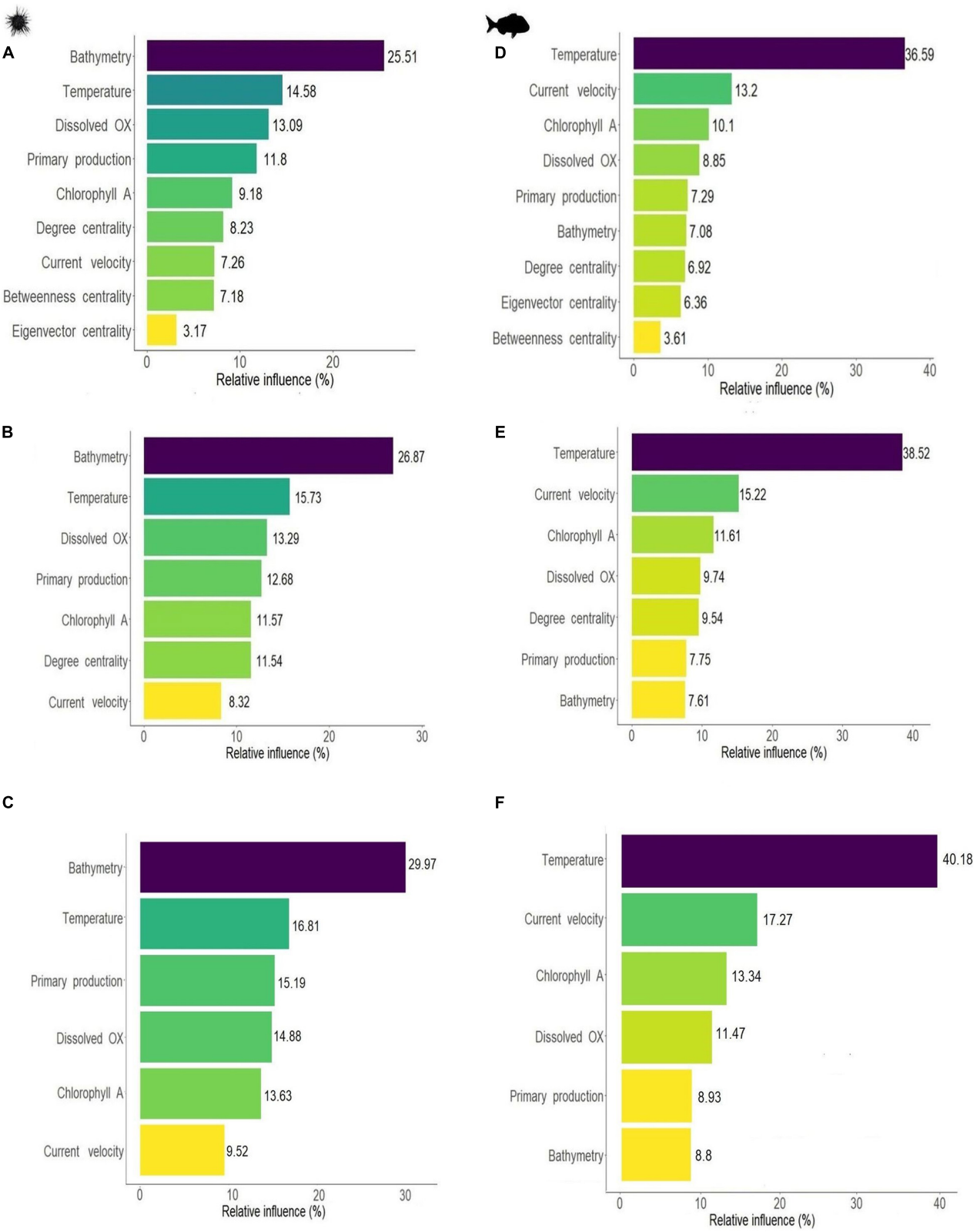

FIGURE 3 | Relative influence of environmental parameters and centrality metrics on SDM results for sea urchin *H. erythrogramma* (left) and snapper, *C. auratus* (right). Fitting BRT including all centrality measures **(A,D)** or selecting only the most influential variable degree centrality **(B,E)**, or excluding centrality **(C,F)**. Relative influence expressed in percentage (i.e., total influence sums up to 100%).

(see **Supplementary Figure 9** for habitat suitability predictions for all models). Despite these models predicted somewhat different species distribution range, when tested for pairwise correlation, the differences in spatial distribution showed very low correlation with the seascape centrality metrics for both species (**Supplementary Figure 10**).

The impact of including (or not) connectivity in the SDM predictions revealed geographic structure in terms of the magnitude of increase or decrease in modeled habitat suitability (**Figure 4**). For the purple sea urchin, most areas showed a decrease in habitat suitability when connectivity was included (i.e., these areas became less suitable in the model), particularly for Port Phillip Bay in Victoria and Spencer Gulf in South Australia and areas far from the coastline (**Figure 4**). Areas of increased habitat suitability were smaller and focused around "central" rocky reefs (**Figure 4B**). Located primarily around high degree centrality sites in central and western Victoria, north and east Tasmania and New South Wales (**Figure 4A**). Rocky reef patches with high betweenness centrality and eigenvector centrality did not correspond to key zones revealed from SDMs results (**Supplementary Figure 11**). Snapper habitat suitability predictions decreased for models including connectivity, especially for areas far from the coast. Areas associated to high degree centrality largely corresponded to higher suitability, particularly along north Tasmania, central Victoria and on the border between Victoria and New South Wales and South Australia habitats (**Figure 4D**). Areas of high eigenvector centrality in central Victoria and north Tasmania also correspond to high degree centrality, while there was no consistent spatial trend for betweenness centrality (**Supplementary Figure 11**).

DISCUSSION

Seascape connectivity is essential for ensuring long term species persistence and determining the distribution of species (Engelhard et al., 2017; Weeks, 2017), and as a result is expected to have a significant influence on predicting species distribution with SDMs. Graph-based centrality metrics may influence SDMs predictions and degree centrality appeared to be the most important metric among the centrality measures.

Degree centrality was the most significant among the centrality measures included in the model. Degree centrality identifies hubs of high connectivity, and it is critically important for benthic species dispersing only during the larval stage, representing the quantity of larval connectivity, identifying important sources and destinations of larvae (Treml et al., 2015; Zamborain-Mason et al., 2017). Hotspots of connectivity ensure persistence in marine metapopulations (Zamborain-Mason et al., 2017; Cecino and Treml, 2021), and in this work was also significant in defining the species spatial distribution, showing that highly central nodes identified areas of greater habitat suitability. Connectivity variables had interactions with the environmental parameters revealing that the most suitable habitat also corresponded to critical habitats for connectivity. Quantifying interactions among variables helps to define more clearly which is the most suitable habitat for the species (Elith et al., 2008), showing how the effect of one environmental predictor on a species changes according to the levels of other predictors. Recognizing these environmental interactions is critical to assess changing environmental conditions, and integrating environmental and ecological interactions produces more robust SDMs and improves understanding of causes of species' distributions (Guisan et al., 2006).

The results for degree centrality indicate that the sea urchin is predicted to occur in shallow waters, around high oxygen concentrations and in hubs of connectivity (degree centrality values between 10 and 20 ecological linkages). Both depth and connectivity are critical to define benthic species distribution, while dispersal largely influences the spatial distribution and range extension (Ling et al., 2009). Depth was found to influence reproduction in sea urchin, where higher gonad index was associated to individual occupying the intertidal zone compared to sea urchins living in the subtidal zone (Basch and Tegner, 2007). The role of temperature as an influential predictor of sea urchin distribution is particularly relevant to the management of sea urchin species due to their range expansion along with the tropicalization of south-eastern Australian waters and the consequential loss of kelp. Using mechanistic species distribution models, range shifts of sea urchins were predicted, revealing how these shifts are driven by climate, therefore leading to the contraction of habitat-forming species such as kelp (Castro et al., 2020).

Snapper, in contrast, is predicted to be found around hubs of connectivity (degree centrality of value 4 and 8), and in warm shallow waters. Elevated temperature is associated with increased larval size and survival influencing the snapper adult population dynamics (McMahon et al., 2020). Adult snapper movement appeared to concentrate around these warmer habitats, where conditions are optimal for larval rearing (Fielder et al., 2005). Current velocity emerged as another influential environmental parameter influencing the snapper SDMs (**Figure 3**). Current velocities proved to be critical to distinguish between juvenile and adult habitat for New Zealand snapper populations (Compton et al., 2012). Water column features such as currents largely influence species distribution predictions of south-east Australian nearshore temperate reef fishes, while other environmental variables like bathymetry appeared to be less important predictors (Young and Carr, 2015).

Machine learning methods such as BRT offer the advantage of exploring not only model performance but also the extent of each variable's relative influence. If predictors have no contribution, the model algorithm calculates the relative variable influence as zero or near zero. In our species distribution models, connectivity contributes to the model, yet the influence on predictions was not as strong as key environmental variables, such as temperature, currents and chlorophyll. As a result, if centrality metrics were omitted, the resultant models would have resulted in different habitat suitability predictions, especially affecting their spatial range. When we included the most influential metric, degree centrality, the influence of connectivity on SDMs predictions increased together with the other predictors remaining among the least influential variables,

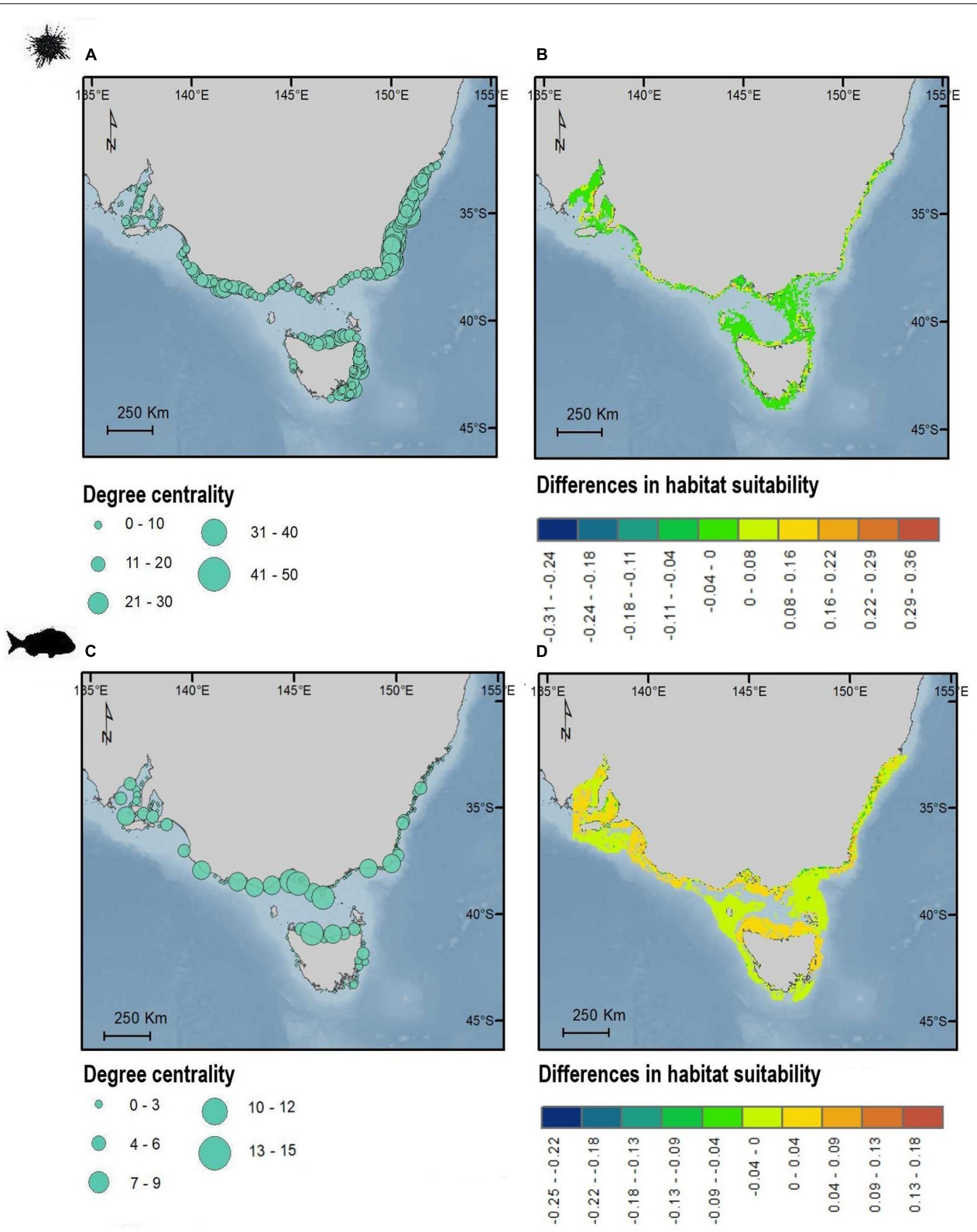

FIGURE 4 | Maps showing the geographic distribution of degree centrality **(A,C)** and differences in spatial predictions of habitat suitability **(B,D)** for sea urchin *H. erythrogramma* (top) and snapper, *C. auratus* (bottom). Degree centrality **(A,C)** is shown as dots corresponding to the habitat patches centroids. Values of habitat suitability are positive when predicted habitat suitability is larger for SDM incorporating connectivity compared to the SDM without connectivity. Values of habitat suitability are negative when predicted habitat suitability is lower for SDM incorporating connectivity compared to the SDM without connectivity. Maps in WGS84.

however, its influence showed a larger increase compared to other environmental predictors. In both species the area under the curve (AUC-ROC) of the models was close to one, indicating that the model performance and predictions were very good (Jiménez-Valverde, 2012). AUC scores were similar for models with and without connectivity, suggesting no differences in the models' predictive performance, however, the spatial range of habitat suitability predictions differed among the models, indicating that differences between the models exist. This apparent contradiction may be explained by the high accuracy typical of machine-learning algorithms (Bucklin et al., 2015). Independent to connectivity, the most influential environmental drivers were bathymetry and temperature for purple sea urchin, and temperature and current velocity for snapper (**Figure 3**), commonly found to have large influence across many marine-focused SDMs (Reiss et al., 2011; Tyberghein et al., 2012).

Despite the limited differences in habitat suitability magnitude, when incorporating connectivity in SDMs' spatial predictions revealed reduced suitability primarily for deep waters, defining a more restricted geographic range for snapper and sea urchins, limiting the distributions to shallow coastal waters. In addition, the inclusion of connectivity in the SDMs increase to a small extent the suitability around several clusters of connectivity hubs. Habitats with high degree centrality, were identified in central Victoria, in proximity of Wilsons Promontory particularly for snapper population, key habitats for metapopulation persistence across species and corresponding to marine protected areas and reserves (Cecino and Treml, 2021). This region significance is well known and includes several ecological features, which define the structure of the coastal communities. Eastern Victoria was identified as potential biogeographic break for many taxa often associated with limits in species' ranges and changes in community assemblages (Colton and Swearer, 2012). Habitat patches in northern Tasmania may also be essential for ensuring connectivity between Tasmania and Victoria coasts. In Eastern Tasmania, the oceanographic mixing zone where subantarctic water masses, driven by westerly winds, interact with eddies from the East Australian current, lead to enhanced productivity, and phytoplankton blooms and mass aggregations of coastal temperate taxa occur (Hosack and Dambacher, 2012; Dambacher et al., 2012; Commonwealth of Australia, 2015). Snapper's hubs of connectivity in South Australia are also consistent with key habitat sites for the snapper fishery and for spawning grounds (Fowler and Jennings, 2003).

Both study species used for this work have somewhat limited dispersal ability, especially in relationship to the extent of the model domain. This choice of model domain was made to highlight an ecologically and economically important Australian seascape, and the influence of ocean dynamics and life histories. However, further research effort may be needed where species have extended home ranges, long-distance swimming capacity, or where dispersal periods extend for many weeks or months.

Applying SDMs to marine species can be particularly challenging. Challenges in understanding how species are distributed across space arise when comprehensive sampling is not possible, for example for species with high degree of niche specialization, and/or restricted range (Araujo and Guisan, 2006). Several issues are somewhat unique of the marine environment. For example, a strong spatial bias in data collection, since different effort is required to collect data in shallow waters compared to deep waters, and the widespread spatial-temporal bias in global satellite-derived ocean measurements, due to unpredicted or unusual atmospheric properties affecting the algorithm interpretation, and the lack of *in situ* data to use for tuning (Robinson et al., 2011, 2017). In our models, occurrence data collected from the Atlas of Living Australia include data from early 1900s, while environmental data were based on information collected from 2000 (see **Supplementary Material** for details) and the connectivity models used ocean current data for the period 1992–2012. This might result in an underestimation of the importance of connectivity and its influence on model predictions. Despite the large temporal extent of the ALA data sets the oldest data largely corresponded the distribution of occurrences recorded in recent years. However, we focused our analysis on a cleaned and reduced data set, reducing the temporal differences among species data, environmental predictors and centrality metrics. The lack of true absence data may be another limitation when developing SDMs, especially for marine species, where presence data sampling is biased toward coastal waters and areas near ports (Robinson et al., 2011). Though we addressed this limitation to some degree by choosing BRT methods, an appropriate procedure when working with species presence and pseudo-absence data (Cerasoli et al., 2017). We selected BRTs over presence-only methods such as Maxent because BRTs allow for better control and quantification of predictors interactions, allows appropriate model complexity and tunes model parameters with internal cross-validation. Moreover, the predictive performance of BRT are comparable to Maxent for predicting presence-only species data (Valavi et al., 2021). BRTs outperform other approaches like generalized linear and additive models, as well as combine many decision trees to improve model's accuracy, include stochasticity, reducing variance and improving predicting performance (Cimino et al., 2020). That said, BRTs are often criticized for their tendency to overfit models. Other limitations common to SDMs include changes in habitat conditions due to climate change and human impacts, and attempting to predict species around range shifts. For exploited taxa like snapper, the distribution of fishing effort likely influences species distribution and presence/absence data. Our models could potentially be improved by including data on fishing pressure and environmental changes to producing more realistic spatial predictions.

Across two very different marine taxa, centrality measures proved to be appropriate and flexible proxies to describe seascape connectivity and can effectively identify hotspots and stepping-stones of connectivity. Using these patch-level metrics to describe seascape connectivity is an efficient way to incorporate connectivity information into marine-based SDMs. Centrality metrics proved to have a limited contribution to SDMs, yet they contribute to define the spatial distribution patterns and the most suitable habitat patches. Importantly, centrality metrics interact with other environmental predictors, highlight the suitability of habitats combining environmental and connectivity

characteristics. Connectivity is fundamentally important for marine species and should be considered in models of species distribution or abundance. Our new methods chart a pathway forward for efficiently incorporating connectivity into marine-based SDMs and open the door for exploring the broader influence of dispersal and movement on species distributions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://doi.org/10.26197/5e337e7061fa7, https://doi.org/10.26197/5d673e22c2429, https://www.hycom.org/, and https://www.bio-oracle.org/code.php.

AUTHOR CONTRIBUTIONS

GC: conceptualization, methodology, formal analysis, data curation, investigation, visualization, and writing – original draft preparation. RV: methodology, formal analysis, and writing – review and editing. ET: supervision, conceptualization, software, resources, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

Funding was provided through a Melbourne Research Scholarship to GC.

ACKNOWLEDGMENTS

Simulations were completed on the University of Melbourne High Performance Computing cluster, Spartan. We thank the reviewers for their thoughtful comments and efforts toward improving our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.766915/full#supplementary-material

REFERENCES

Adams, M. P., Saunders, M. I., Maxwell, P. S., Tuazon, D., Roelfsema, C. M., Callaghan, D. P., et al. (2016). Prioritizing localized management actions for seagrass conservation and restoration using a species distribution model. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 26, 639–659. doi: 10.1002/aqc.2573

Agardy, M. T. (1994). Advances in marine conservation: the role of marine protected areas. *Trends Ecol. Evol.* 9, 267–270. doi: 10.1016/0169-5347(94)90297-6

Araujo, M. B., and Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688. doi: 10.1111/j.1365-2699.2006.01584.x

Atlas of Living Australia [ALA] (2019). *Chrysophrys auratus Occurrence Data*. ALA. doi: 10.26197/5d673e22c2429

Atlas of Living Australia [ALA] (2020). *Heliocidaris erythrogramma Occurrence Data*. ALA. doi: 10.26197/5e337e7061fa7

Báez, J. C., Olivero, J., Peteiro, C., Ferri-Yáñez, F., Garcia-Soto, C., and Real, R. (2010). Macro-environmental modelling of the current distribution of *Undaria pinnatifida* (Laminariales, Ochrophyta) in northern Iberia. *Biol. Invas.* 12, 2131–2139. doi: 10.1007/s10530-009-9614-1

Basch, L. V., and Tegner, M. J. (2007). Reproductive responses of purple sea urchin (*Strongylocentrotus purpuratus*) populations to environmental conditions across a coastal depth gradient. *Bull. Mar. Sci.* 81, 255–282.

Bax, N., and Williams, A. (2000). *Habitat and Fisheries Productivity in the South East Fishery Ecosystem*. Deakin West, ACT: Fisheries Research and Development Corporation.

Bode, M., Burrage, K., and Possingham, H. P. (2008). Using complex network metrics to predict the persistence of metapopulations with asymmetric connectivity patterns. *Ecol. Model.* 214, 201–209. doi: 10.1016/j.ecolmodel.2008.02.040

Bodin, O., and Norberg, J. (2007). A network approach for analyzing spatially structured populations in fragmented landscape. *Landsc. Ecol.* 22, 31–44. doi: 10.1007/s10980-006-9015-0

Bodin, Ö, and Saura, S. (2010). Ranking individual habitat patches as connectivity providers: integrating network analysis and patch removal experiments. *Ecol. Model.* 221, 2393–2405. doi: 10.1016/j.ecolmodel.2010.06.017

Bonacich, P. (1987). Power and centrality – a family of measures. *Am. J. Soc.* 92, 1170–1182.

Borgatti, S. P. (2005). Centrality and network flow. *Soc. Netw.* 27, 55–71. doi: 10.1016/j.socnet.2004.11.008

Bucklin, D. N., Basille, M., Benscoter, A. M., Brandt, L. A., Mazzotti, F. J., Romañach, S. S., et al. (2015). Comparing species distribution models constructed with different subsets of environmental predictors. *Divers. Distrib.* 21, 23–35. doi: 10.1111/ddi.12247

Burgess, S. C., Nickols, K. J., Griesemer, C. D., Barnett, L. A., Dedrick, A. G., Satterthwaite, E. V., et al. (2014). Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. *Ecol. Appl.* 24, 257–270. doi: 10.1890/13-0710.1

Butler, C., Proctor, R. D., Flukes, E. D., Walsh, P., Johnson, C. P., and Lucieer, V. D. (2017). *Seamap Australia - A National Seafloor Habitat Classification Scheme*. Hobart TAS: University of Tasmania.

Caldwell, I. R., and Gergel, S. E. (2013). Thresholds in seascape connectivity: influence of mobility, habitat distribution, and current strength on fish movement. *Landsc. Ecol.* 28, 1937–1948. doi: 10.1007/s10980-013-9930-9

Carroll, C., McRAE, B. H., and Brookes, A. (2012). Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. *Conserv. Biol.* 26, 78–87. doi: 10.1111/j.1523-1739.2011.01753.x

Castro, L. C., Cetina-Heredia, P., Roughan, M., Dworjanyn, S., Thibaut, L., Chamberlain, M. A., et al. (2020). Combined mechanistic modelling predicts changes in species distribution and increased co-occurrence of a tropical urchin herbivore and a habitat-forming temperate kelp. *Divers. Distrib.* 26, 1211–1226. doi: 10.1111/ddi.13073

Cecino, G., and Treml, E. A. (2021). Local connections and the larval competency strongly influence marine metapopulation persistence. *Ecol. Appl.* 31:e02302. doi: 10.1002/eap.2302

Cerasoli, F., Iannella, M., D'Alessandro, P., and Biondi, M. (2017). Comparing pseudo-absences generation techniques in Boosted Regression Trees models for conservation purposes: a case study on amphibians in a protected area. *PLoS One* 12:e0187589. doi: 10.1371/journal.pone.0187589

Cimino, M. A., Santora, J. A., Schroeder, I., Sydeman, W., Jacox, M. G., Hazen, E. L., et al. (2020). Essential krill species habitat resolved by seasonal upwelling and ocean circulation models within the large marine ecosystem of the California current system. *Ecography* 43, 1536–1549. doi: 10.1111/ecog.05204

Clauzel, C., Girardet, X., and Foltête, J.-C. (2013). Impact assessment of a high-speed railway line on species distribution: application to the European tree frog (*Hyla arborea*) in Franche-Comté. *J. Environ. Manage.* 127, 125–134. doi: 10.1016/j.jenvman.2013.04.018

Colton, M. A., and Swearer, S. E. (2012). Locating faunal breaks in the nearshore fish assemblage of Victoria, Australia. *Mar. Freshw. Res.* 63, 218–231. doi: 10.1071/MF10322

Commonwealth of Australia (2015). *South-East Marine Region Profle: A Description of the Ecosystems, Conservation Values and Uses of the South-East Marine Region*. Canberra ACT: department of agriculture water and the environment.

Compton, T. J., Morrison, M. A., Leathwick, J. R., and Carbines, G. D. (2012). Ontogenetic habitat associations of a demersal fish species, *Pagrus auratus*, identified using boosted regression trees. *Mar. Ecol .Prog. Ser.* 462, 219–230. doi: 10.3354/meps09790

Cowen, R. K., and Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.* 1, 443–466.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Inter J. Complex Syst.* 1695, 1–9.

Dambacher, J. M., Hayes, K. R., Hosack, G. R., Lyne, V., Clifford, D., Dutra, L. X., et al. (2012). *National Marine Ecological Indicators*. Hobart, Tas: CSIRO Mathematics, Informatics and Statistics.

De Marco, P. J., and Nóbrega, C. C. (2018). Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS One* 13:e0202403. doi: 10.1371/journal.pone.0202403

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Fan, W.-Y., and Wang, Z.-G. (2014). The predictive performance and stability of six species distribution models. *PLoS One* 9:e112764. doi: 10.1371/journal.pone.0112764

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

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. doi: 10.1146/annurev.ecolsys.110308.120159

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Engelhard, S. L., Huijbers, C. M., Stewart-Koster, B., Olds, A. D., Schlacher, T. A., and Connolly, R. M. (2017). Prioritising seascape connectivity in conservation using network analysis. *J. Appl. Ecol.* 54, 1130–1141.

ESRI (2017). *ArcGIS*. Redlands, CA: Environmental Systems Research Institute Inc.

Estrada, E., and Bodin, O. (2008). Using network centrality measures to manage landscape connectivity. *Ecol. Appl.* 18, 1810–1825. doi: 10.1890/07-1419.1

Fielder, D. S., Bardsley, W. J., Allan, G. L., and Pankhurst, P. M. (2005). The effects of salinity and temperature on growth and survival of Australian snapper, *Pagrus auratus* larvae. *Aquaculture* 250, 201–214. doi: 10.1016/j.aquaculture.2005.04.045

Fischer, J., Kleidon, A., and Dittrich, P. (2015). Thermodynamics of random reaction networks. *PLoS One* 10:e0117312. doi: 10.1371/journal.pone.0117312

Foltête, J. C., Clauzel, C., Vuidel, G., and Tournant, P. (2012). Integrating graph-based connectivity metrics into species distribution models. *Landsc. Ecol.* 27, 557–569. doi: 10.1007/s10980-012-9709-4

Fowler, A., Huveneers, C., and Lloyd, M. (2017). Insights into movement behaviour of snapper (*Chrysophrys auratus*, Sparidae) from a large acoustic array. *Mar. Freshw. Res.* 68, 1438–1453. doi: 10.1071/MF16121

Fowler, A. J., and Jennings, P. R. (2003). Dynamics in 0+ recruitment and early life history for snapper (*Pagrus auratus*, Sparidae) in South Australia. *Mar. Freshw. Res.* 54, 941–956. doi: 10.1071/MF02172

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Soc. Netw.* 1, 215–239. doi: 10.1016/0378-8733(78)90021-7

Girardet, X., Foltete, J. C., and Clauzel, C. (2013). Designing a graph-based approach to landscape ecological assessment of linear infrastructures. *Environ. Impact Assess. Rev.* 42, 10–17. doi: 10.1016/J.EIAR.2013.03.004

Gormley, K. S., Hull, A. D., Porter, J. S., Bell, M. C., and Sanderson, W. G. (2015). Adaptive management, international co-operation and planning for marine conservation hotspots in a changing climate. *Mar. Policy* 53, 54–66. doi: 10.1016/j.marpol.2014.11.017

Grober-Dunsmore, R., Pittman, S. J., Caldow, C., Kendall, M. S., and Frazer, T. K. (2009). "A landscape ecology approach for the study of ecological connectivity across tropical marine seascapes," in *Ecological Connectivity Among Tropical Coastal Ecosystems*, ed. I. Nagelkerken (New York, NY: Springer), 493–530.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R., et al. (2006). Making better biogeographical predictions of species' distributions. *J. Appl. Ecol.* 43, 386–392. doi: 10.1111/j.1365-2664.2006.01164.x

Hamer, P. A., Acevedo, S., Jenkins, G. P., and Newman, A. (2011). Connectivity of a large embayment and coastal fishery: spawning aggregations in one bay source local and broad-scale fishery replenishment. *J. Fish Biol.* 78, 1090–1109. doi: 10.1111/j.1095-8649.2011.02921.x

Hanski, I. (1998). Metapopulation dynamics. *Nature* 396, 41–49. doi: 10.1038/23876

Harasti, D., Lee, K. A., Gallen, C., Hughes, J. M., and Stewart, J. (2015). Movements, home range and site fidelity of snapper (*Chrysophrys auratus*) within a temperate marine protected area. *PLoS One* 10:e0142454. doi: 10.1371/journal.pone.0142454

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edn. New York, NY: Springer series in statistics.

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. (2020). *dismo: Specie Distribution Modeling. R package version 1.3-3*. Available online at: https://CRAN.R-project.org/package=dismo (accessed October 11, 2021).

Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., and Hijmans, M. R. J. (2017). Package 'dismo'. *Circles* 9, 1–68.

Hosack, G., and Dambacher, J. (2012). *Ecological Indicators for the Exclusive Economic Zone of Australia's South East Marine Region*. CSIRO;2012. Canberra, ACT: CSIRO, doi: 10.4225/08/584c44e1a4389

Huggett, M. J., Crocetti, G. R., Kjelleberg, S., and Steinberg, P. D. (2008). Recruitment of the sea urchin *Heliocidaris erythrogramma* and the distribution and abundance of inducing bacteria in the field. *Aquat. Microb. Ecol.* 53, 161–171. doi: 10.3354/AME01239

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 498–507. doi: 10.1111/J.1466-8238.2011.00683.X

Jones, M. C., and Cheung, W. W. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES J. Mar. Sci.* 72, 741–752. doi: 10.1093/icesjms/fsu172

Lamare, M. D., and Barker, M. F. (1999). In situ estimates of larval development and mortality in the New Zealand sea urchin *Evechinus chloroticus* (Echinodermata : Echinoidea). *Mar. Ecol. Prog. Ser.* 180, 197–211. doi: 10.3354/meps180197

Ling, S., Ibbott, S., and Sanderson, J. (2010). Recovery of canopy-forming macroalgae following removal of the enigmatic grazing sea urchin *Heliocidaris erythrogramma*. *J. Exp. Mar. Biol. Ecol.* 395, 135–146. doi: 10.1016/j.jembe.2010.08.027

Ling, S., Scheibling, R., Rassweiler, A., Johnson, C., Shears, N., Connell, S., et al. (2015). Global regime shift dynamics of catastrophic sea urchin overgrazing. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20130269. doi: 10.1098/rstb.2013.0269

Ling, S. D., Johnson, C. R., Ridgway, K., Hobday, A. J., and Haddon, M. (2009). Climate-driven range extension of a sea urchin: inferring future trends by analysis of recent population dynamics. *Glob. Change Biol.* 15, 719–731. doi: 10.1111/j.1365-2486.2008.01734.x

McMahon, S. J., Parsons, D. M., Donelson, J. M., Pether, S. M., and Munday, P. L. (2020). Elevated temperature and CO2 have positive effects on the growth and survival of larval Australasian snapper. *Mar. Environ. Res.* 161:105054. doi: 10.1016/j.marenvres.2020.105054

McRae, B. H., and Kavanagh, D. M. (2011). *Linkage Mapper Connectivity Analysis Software*. Arlington, TX: The Nature Conservancy.

Meyer, C. G., Papastamatiou, Y. P., and Clark, T. B. (2010). Differential movement patterns and site fidelity among trophic groups of reef fishes in a Hawaiian marine protected area. *Mar. Biol.* 157, 1499–1511. doi: 10.1007/s00227-010-1424-6

Minor, E. S., and Urban, D. L. (2007). Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecol. Appl.* 17, 1771–1782. doi: 10.1890/06-1073.1

Monk, J., Ierodiaconou, D., Versace, V. L., Bellgrove, A., Harvey, E., Rattray, A., et al. (2010). Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. *Mar. Ecol. Prog. Ser.* 420, 157–174. doi: 10.3354/meps08858

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Soc. Netw.* 27, 39–54. doi: 10.1016/j.socnet.2004.11.009

O'Hara, T. (2002). Endemism, rarity and vulnerability of marine species along a temperate coastline. *Invertebr. Syst.* 16, 671–684. doi: 10.1071/IT01034

Okubo, A. (1971). Oceanic Diffusion Diagrams. *Deep-Sea Research* 18, 789–802. doi: 10.1016/0011-7471(71)90046-5

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners synthesis. *Am. Museum Nat. Hist.* 50, 54–89.

Pederson, H. G., and Johnson, C. R. (2006). Predation of the sea urchin Heliocidaris erythrogramma by rock lobsters (*Jasus edwardsii*) in no-take marine reserves. *J. Exp. Mar. Biol. Ecol.* 336, 120–134. doi: 10.1016/j.jembe.2006.04.010

Pulliam, H. R. (1988). Sources, sinks, and population regulation. *Am. Natural.* 132, 652–661.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rayfield, B., Fortin, M.-J., and Fall, A. (2010). The sensitivity of least-cost habitat graphs to relative cost surface values. *Landsc. Ecol.* 25, 519–532. doi: 10.1007/s10980-009-9436-7

Reiss, H., Cunze, S., König, K., Neumann, H., and Kröncke, I. (2011). Species distribution modelling of marine benthos: a North Sea case study. *Mar. Ecol. Prog. Ser.* 442, 71–86. doi: 10.3354/meps09391

Roberts, J. J., Best, B. D., Dunn, D. C., Treml, E. A., and Halpin, P. N. (2010). Marine geospatial ecology tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environ. Model. Softw.* 25, 1197–1207. doi: 10.1016/j.envsoft.2010.03.029

Robinson, L., Elith, J., Hobday, A., Pearson, R., Kendall, B., Possingham, H., et al. (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20, 789–802. doi: 10.1111/j.1466-8238.2010.00636.x

Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E., and Lundquist, C. J. (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Front. Mar. Sci.* 4:421. doi: 10.3389/fmars.2017.00421

Rose, K. A., Cowan, J. H. Jr., Winemiller, K. O., Myers, R. A., and Hilborn, R. (2001). Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish Fish.* 2, 293–327. doi: 10.1046/j.1467-2960.2001.00056.x

Rumrill, S. S. (1987). *Differential Predation Upon Embryos and Larvae of Temperate Pacific Echinoderms*. Edmonton, AB: University of Alberta.

Swanson, R. L., Byrne, M., Prowse, T. A. A., Mos, B., Dworjanyn, S. A., and Steinberg, P. D. (2012). Dissolved histamine: a potential habitat marker promoting settlement and metamorphosis in sea urchin larvae. *Mar. Biol.* 159, 915–925. doi: 10.1007/s00227-011-1869-2

Tarabon, S., Bergès, L., Dutoit, T., and Isselin-Nondedeu, F. (2019). Environmental impact assessment of development projects improved by merging species distribution and habitat connectivity modelling. *J. Environ. Manage.* 241, 439–449. doi: 10.1016/j.jenvman.2019.02.031

Terres, M. A., Lawrence, E., Hosack, G. R., Haywood, M. D., and Babcock, R. C. (2015). Assessing habitat use by Snapper (*Chrysophrys auratus*) from baited underwater video data in a coastal marine park. *PLoS One* 10:e0136799. doi: 10.1371/journal.pone.0136799

Treml, E. A., Ford, J. R., Black, K. P., and Swearer, S. E. (2015). Identifying the key biophysical drivers, connectivity outcomes, and metapopulation consequences of larval dispersal in the sea. *Mov. Ecol.* 3:16. doi: 10.1186/s40462-015-0045-6

Treml, E. A., Halpin, P. N., Urban, D. L., and Pratson, L. F. (2008). Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landsc. Ecol.* 23, 19–36. doi: 10.1007/s10980-007-9138-y

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., and Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integr. Comp. Biol.* 52, 525–537. doi: 10.1093/icb/ics101

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 272–281. doi: 10.1111/j.1466-8238.2011.00656.x

Urban, D., and Keitt, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology* 82, 1205–1218.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., and Elith, J. (2021). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* e1486. doi: 10.1002/ecm.1486

Vanderklift, M. A., and Kendrick, G. A. (2004). Variation in abundances of herbivorous invertebrates in temperate subtidal rocky reef habitats. *Mar. Freshw. Res.* 55, 93–103. doi: 10.1071/MF03057

Watson, J. R., Siegel, D. A., Kendall, B. E., Mitarai, S., Rassweiller, A., and Gaines, S. D. (2011). Identifying critical regions in small-world marine metapopulations. *Proc. Natl. Acad. Sci. U.S.A.* 108, E907–E913. doi: 10.1073/pnas.1111461108

Weeks, R. (2017). Incorporating seascape connectivity in conservation prioritisation. *PLoS One* 12:16. doi: 10.1371/journal.pone.0182396

Williams, P. D., and Hastings, A. (2013). Stochastic dispersal and population persistence in marine organisms. *Am. Natural.* 182, 271–282.

Young, M., and Carr, M. H. (2015). Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef fishes. *Divers. Distrib.* 21, 1428–1440. doi: 10.1111/ddi.12378

Zamborain-Mason, J., Russ, G. R., Abesamis, R. A., Bucol, A. A., and Connolly, S. R. (2017). Network theory and metapopulation persistence: incorporating node self-connections. *Ecol. Lett.* 20, 815–831. doi: 10.1111/ele.12784

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fmars.2022.711522

OPEN ACCESS

EDITED BY
Mary C. Fabrizio,
Virginia Institute of Marine Science,
United States

REVIEWED BY
Barbara Muhling,
University of California, Santa Cruz,
United States
Kelly Ortega-Cisneros,
University of Cape Town, South Africa

*CORRESPONDENCE
Rebecca G. Asch
aschr16@ecu.edu

SPECIALTY SECTION
This article was submitted to Marine Conservation and Sustainability, a section of the journal Frontiers in Marine Science

RECEIVED 18 May 2021
ACCEPTED 27 July 2022
PUBLISHED 25 August 2022

CITATION
Asch RG, Sobolewska J and Chan K (2022) Assessing the reliability of species distribution models in the face of climate and ecosystem regime shifts: small pelagic fishes in the California Current System. *Front. Mar. Sci.* 9:711522. doi: 10.3389/fmars.2022.711522

Assessing the reliability of species distribution models in the face of climate and ecosystem regime shifts: Small pelagic fishes in the California Current System

Rebecca G. Asch[1,2]*, Joanna Sobolewska[1] and Keo Chan[1]

[1]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, United States, [2]Department of Biology, East Carolina University, Greenville, NC, United States

Species distribution models (SDMs) are a commonly used tool, which when combined with earth system models (ESMs), can project changes in organismal occurrence, abundance, and phenology under climate change. An often untested assumption of SDMs is that relationships between organisms and the environment are stationary. To evaluate this assumption, we examined whether patterns of distribution among larvae of four small pelagic fishes (Pacific sardine *Sardinops sagax*, northern anchovy *Engraulis mordax*, jack mackerel *Trachurus symmetricus*, chub mackerel *Scomber japonicus*) in the California Current remained steady across time periods defined by climate regimes, changes in secondary productivity, and breakpoints in time series of spawning stock biomass (SSB). Generalized additive models (GAMs) were constructed separately for each period using temperature, salinity, dissolved oxygen (DO), and mesozooplankton volume as predictors of larval occurrence. We assessed non-stationarity based on changes in six metrics: 1) variables included in SDMs; 2) whether a variable exhibited a linear or non-linear form; 3) rank order of deviance explained by variables; 4) response curve shape; 5) degree of responsiveness of fishes to a variable; 6) range of environmental variables associated with maximum larval occurrence. Across all species and time periods, non-stationarity was ubiquitous, affecting at least one of the six indicators. Rank order of environmental variables, response curve shape, and oceanic conditions associated with peak larval occurrence were the indicators most subject to change. Non-stationarity was most common among regimes defined by changes in fish SSB. The relationships between larvae and DO were somewhat more likely to change across periods, whereas the relationships between fishes and temperature were more stable. Respectively, *S. sagax*, *T. symmetricus*, *S. japonicus*, and *E. mordax* exhibited non-stationarity across 89%, 67%, 50%, and 50% of indicators. For all species except *E. mordax*, inter-model variability had a larger impact on projected habitat suitability for larval

fishes than differences between two climate change scenarios (SSP1-2.6 and SSP5-8.5), implying that subtle differences in model formulation could have amplified future effects. These results suggest that the widespread non-stationarity in how fishes utilize their environment could hamper our ability to reliably project how species will respond to climatic change.

KEYWORDS

species distribution models, small pelagic fish, forage fish, climate change projections, non-stationarity, California Current

1 Introduction

Marine fishes in many ecosystems have shifted their distribution poleward and deeper as climate change has warmed the oceans (Murawski, 1993; Perry et al., 2005; Hsieh et al., 2008; Hsieh et al., 2009; Nye et al., 2009; Pinsky et al., 2013; Poloczanska et al., 2013; Walsh et al., 2015). Many of these changes are occurring at a rate faster than in terrestrial habitats (Sunday et al., 2012; Poloczanska et al., 2013; Blowes et al., 2019; Pinsky et al., 2019). Climate velocity, a measure of the rate of temperature change across spatial gradients, has proven to be an accurate predictor of the magnitude and direction of shifts in species distributions in many ecosystems (Chen et al., 2011; Pinsky et al., 2013), although other aspects of a species' ecological niche also influence distribution changes (McHenry et al., 2019). Throughout the 21st century, climate models project that changes in species distribution will continue unabated or further accelerate (Cheung et al., 2009; Cheung et al., 2016b; Morley et al., 2018). Shifts in fish distribution have implications for trophic interactions (Selden et al., 2018), global biodiversity patterns (Cheung et al., 2009), and food security (Golden et al., 2016; Free et al., 2019).

Many projections of changes in fish distribution, biomass, and phenology under climate change are based on statistical models referred to as species distribution models (SDMs), ecological niche models, or bioclimate envelope models. These models link spatial and temporal variations in organismal occurrence with environmental variables (Elith and Leathwick, 2009). Based on these empirical relationships, changes in environmental conditions derived from climate models are used to project future shifts in species occurrence or abundance. Due to the growing importance of climate change, there has been a rise in studies using SDMs and aligned models over the last 20 years (Figure 1).

A key assumption of SDMs is that the relationship between organisms and environmental conditions is stationary and not subject to changes due to variations in organismal abundance, climate, or ecosystem state. Since statistically derived relationship between a species and the environment form the basis for SDM projections, non-stationarity in this relationship could result in inaccurate projections of climate change impacts. Assumptions about stationarity in relationships between fishes and climatic variables have rarely been investigated (Litzow et al., 2019), but it is imperative to do so to assess the uncertainty associated with projections about how marine conservation initiatives will fare under climate change. Among planktonic organisms, such as dinoflagellates, diatoms, and copepods, SDMs developed using data from one decade failed to accurately project in species distribution during other decades (Brun et al., 2016). This reflects the patchy distribution of plankters, boom-bust cycles in abundance, and the potential for advection of plankton by currents outside their preferred habitat. Since projections made for copepods had greater model skill than those for primary producers, SDMs may have improved predictability for higher trophic level organisms, such as fishes. Nonetheless, recent work suggests that non-stationarity might be a common, albeit understudied, feature among SDMs that project changes in fish distribution (Litzow et al., 2018; Litzow et al., 2019; Puerta et al., 2019; Roberts et al., 2019; Litzow et al., 2020; Muhling et al., 2020).

At least seven ecological, climatic, and statistical mechanisms can lead to non-stationary fish-climate relationships. First, non-stationarity could arise if key variables influencing a species' ecological niche are excluded from an SDM. For example, many SDMs neglect to account for interspecific relationships, such as predator-prey dynamics (Fernandes et al., 2013). Second, over-parameterization of models can lead to the appearance of non-stationarity if this results in a relationship between an environmental variable and fish distribution that is solely due to a statistical artifact. Third, non-stationarity can result from density-dependent occurrence patterns where a fish is found in its optimal habitat at low

1 SPF refer to small-bodied fishes that live in the epipelagic zone (0-200 m), typically exhibit schooling behavior, and consume primarily a planktivorous diet. The largest fisheries for SPF target species in the order Clupeiformes, which includes sardines, anchovies, herrings, menhadens, and shads.

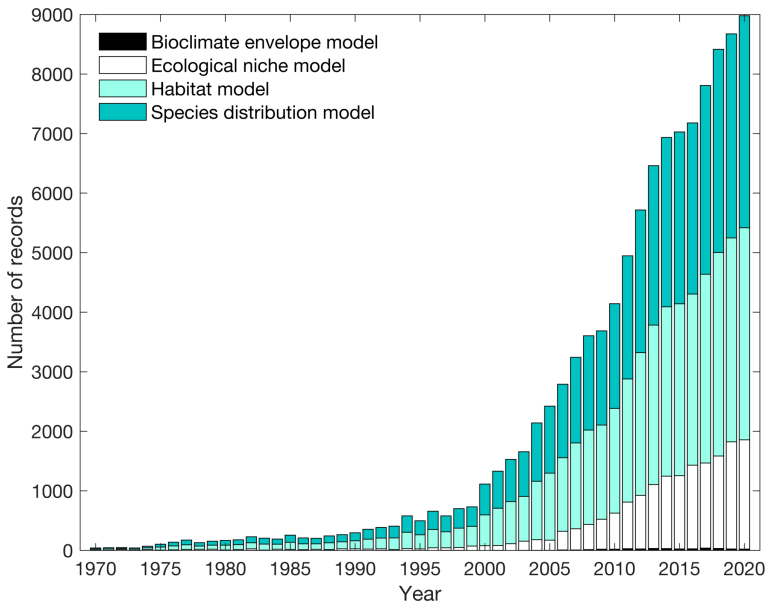

FIGURE 1
Web of Science search examining the cumulative number of records in the scientific literature on species distribution models, habitat models, ecological niche models, and bioclimate envelope models between 1970-2020. Five Web of Science searches were performed: (1) species AND distribution AND model*; (2) habitat AND model*; (3) ecolog* AND niche AND model*; (4) environment* AND niche AND model*, and; (5) bioclimate AND envelope AND model*. Results from the third and fourth search were combined in this figure.

density but, as its abundance increases, spreads to additional habitats to reduce interspecific competition (MacCall, 1990). Such dynamics are especially common among small pelagic fishes (SPF)[1] (Barange et al., 2009). Fourth, overfishing can truncate fish age structure, which can increase sensitivity to climatic variables since younger and smaller fishes often exhibit heightened sensitivity (Anderson et al., 2008). Fifth, at times, fish distribution has been related to basin-scale climate indices, such as the Pacific Decadal Oscillation (PDO), North Pacific Gyre Oscillation, and North Atlantic Oscillation. Many of these indices represent statistical compilations of several climatic variables. If the relationship between these indices and local climate variables changes over time (Joyce, 2002; Litzow et al., 2018; Litzow et al., 2020), this can lead to non-stationarity between species distribution and climate indices (Litzow et al., 2018; Litzow et al., 2019; Puerta et al., 2019; Litzow et al., 2020). Also, some species have been shown to react differently to environmental conditions, such as temperature, depending on the phase of climate oscillations likely due to the influence of these oscillations on larval advection or interspecific interactions (Roberts et al., 2019). Lastly, non-stationarity across climate oscillations could occur because some climate indices, such as the PDO, are detrended. Sixth, the distribution of some species may be constrained by non-climatic factors, such as depth, reliance on biogenic habitats, or lack of dispersal corridors (Reglero et al., 2012; Asch et al., 2019). When such constraints exist, organisms may be retained in their historical habitats, even though the climate of those habitats has shifted. This can result in a non-stationary relationship between species and climate. Lastly, phenotypic plasticity, acclimation to new conditions, or rapid adaptation could lead to changes in how species distribution is related to climate (Donelson et al., 2012; Anderson et al., 2013).

Despite numerous reasons why non-stationarity may occur, there have been relatively few assessments of non-stationarity in SDMs for marine fishes due to a paucity of spatially resolved, long-term datasets that can be used to test historical changes in how fish react to the environment. One such dataset that is well suited to examine non-stationary, fish-climate relationships is California Cooperative Ocean Fisheries Investigations (CalCOFI). This program has surveyed ichthyoplankton along six transects in its core region off southern California since 1951. This region has been subject to several climate regime shifts that affected living marine resources (McGowan et al., 2003; Di Lorenzo et al., 2008; Peabody et al., 2018; Litzow et al., 2020), making it a useful testbed for evaluating whether fishes react differently to environmental variables during each phase of a regime. Also, some of the fastest rates of species distribution change in U.S. waters are projected to occur in this area (Morley et al., 2018), making it an important region for studying non-stationarity.

Our analysis of non-stationarity focuses on SPF since these species account for approximately one-third of global fish catch (Smith et al., 2011). Also, pelagic fishes are often more sensitive

to climate-induced range shifts than demersal fishes (Murawski, 1993; Cheung et al., 2009; Walsh et al., 2015). SPF connect lower trophic organisms in upwelling systems with higher trophic level predators, such as piscivorous fishes, squid, seabirds, and marine mammals (Cury et al., 2011; Pikitch et al., 2014; Kaplan et al., 2017). Furthermore, their potential sensitivity to non-stationary dynamics is likely since SPF exhibit boom-bust cycles of abundance over multi-decadal periodicities (Schwartzlose et al., 1999; Chavez et al., 2003; McClatchie et al., 2017).

More specifically, we focus on four species managed under the Coastal Pelagic Species Fisheries Management Plan: northern anchovy (*Engraulis mordax*), Pacific sardine (*Sardinops sagax*), chub mackerel (*Scomber japonicus*), and jack mackerel (*Trachurus symmetricus*) [PFMC (Pacific Fishery Management Council), 2019]. Previous research has shown that these species are sensitive to fluctuations in oceanic conditions connected to climate variability and change (Lluch-Belda et al., 1991; Checkley Jr et al., 2000; Reiss et al., 2008; Rykaczewski and Checkley Jr, 2008; Weber and McClatchie, 2010; Zwolinski et al., 2011; Weber and McClatchie, 2012; Asch and Checkley Jr, 2013; Koslow et al., 2013; Howard et al., 2020).

Non-stationary relationships between SPF and environmental conditions were observed in the California Current System (CCS) in 2014-2017 when a marine heat wave (MHW) resulted in sea surface temperature (SST) anomalies exceeding three standard deviations above normal conditions (Di Lorenzo and Mantua, 2016). Historically the probability of adult *S. sardinops* occurrence declines when temperature exceeds 18°C, but during this event the probability of encountering *S. sardinops* peaked in some areas warmer than >19°C (Muhling et al., 2020). While this study did not detect similar incidents of non-stationarity when examining data from 1980 through present, it was unclear whether the rapid environmental change during the MHW was the main cause for non-stationarity or if similar non-stationary events might be observed if a longer time series were examined (Muhling et al., 2020). We addressed the latter question by determining if non-stationarity is prevalent in SDMs developed for larval *E. mordax*, *S. sardinops*, *S. japonicus*, and *T. symmetricus* between 1951-2015. This time series emphasizes the period prior to the MHW. We first determined if there were change points in time series of climate indices, oceanic variables, and fish spawning stock biomass (SSB). These change points are proxies for regime shifts. For each period associated with a different regime, we constructed a SDM for each species. Six metrics for identifying non-stationarity were inspected to determine if the relationships between fishes and oceanic conditions changed across regimes. Lastly, we examined whether SDMs developed under different regimes produce equivalent projections of future changes in fish habitat suitability under low and high greenhouse gas emissions.

2 Materials and methods

2.1 Data sources

2.1.1 Larval fish data

CalCOFI has sampled *E. mordax*, *S. sagax*, *S. japonicus*, and *T. symmetricus* larvae since 1951, with the highest concentration of samples from a core region of southern California that extends offshore from San Diego (33.0°N) to north of Point Conception (35.1°N). CalCOFI data are publicly available from the NOAA ERDDAP server.[2] Data on oblique ring and bongo net tows from January 1951 through April 2015 were downloaded for CalCOFI lines 76-93.3. Study sites farther offshore than CalCOFI Station 120 were filtered from this dataset because these stations were sampled less consistently. These criteria resulted in selection of 18,899 net tows. Sample collection occurred monthly during the 1950s, near monthly during the 1960s, 1970s, and early 1980s, albeit with substantial gaps during the 1970s, and quarterly since 1985. The methods for collecting and processing bongo and ring net samples were described in Kramer et al. (1972) and changes to sampling methodology were documented in Ohman and Smith (1995) and Thompson et al. (2017).

2.1.2 Oceanic data

Four environmental variables were selected for inclusion in SDMs because they were measured since 1951 concurrently at stations where CalCOFI ichthyoplankton samples were collected and because these variables were previously shown to influence target species (Checkley Jr et al., 2000; Lynn, 2003; Rykaczewski and Checkley Jr, 2008; Weber and McClatchie, 2010; Zwolinski et al., 2011; Weber and McClatchie, 2012; Asch and Checkley Jr, 2013; Weber et al., 2018; Howard et al., 2020). These variables included potential temperature, salinity, dissolved oxygen (DO), and mesozooplankton displacement volume (abbreviated as ZDV for zooplankton displacement volume). Both salinity and DO can be interpreted as indicators of water masses with distinct characteristics (e.g., Pacific subarctic water has low temperature and salinity, but high DO, whereas North Pacific Central water has high temperature and salinity, with low DO; McClatchie, 2013). Low DO can also act as a stressor affecting the physiology, distribution, and abundance of SPF (Howard et al., 2020). Upwelling of hypoxic and anoxic waters on the inner shelf has been observed in the northern CCS (Chan et al., 2008). In the southern CCS where upwelling is less vigorous, hypoxic waters do not frequently encroach into depths where SPF larvae reside (Dussin et al., 2019), so we interpret variations in DO primarily as an indicator of water mass properties. Temperature, salinity, and DO from Niskin bottles were averaged over the upper 50 m. This depth was selected because SPF eggs are most concentrated

2 https://coastwatch.pfeg.noaa.gov/erddap/index.html

across this range (Curtis et al., 2007). Environmental data were downloaded from ERDDAP between January 1951 and February 2015 and extending between 29.7-35.3° N and 117.2-125.8° W. This area corresponded to transects selected for fish larvae. Within these constraints, 18,925 environmental observations were identified for analysis.

ZDV was obtained from the same bongo and ring nets as larval fishes. We used displacement volumes where gelatinous organisms with biovolumes >5 cm^3 were removed (Kramer et al., 1972). Bias corrections from Ohman and Smith (1995) were applied to account for a change in tow depth (switch from 140 m to 210 m) and net type (switch from a 550-µm silk mesh net to a 505-µm nylon mesh net) in 1969 and a second change in net type (switch from a 1.0-m diameter ring net to a 0.71-m diameter bongo net) in 1977. ZDV measurements were ln(x +1) transformed prior to analysis. As a result, measurements of ZDV are presented with units of the log of the zooplankton volume measured in cm^3 divided by the standardized volume of seawater filtered during a plankton net tow (1,000 m^3). 18,746 observations of ZDV were available for analysis.

Oceanic and biological data were matched based on the year, month, transect, and station number. If multiple sets of environmental variables were matched to a single tow, data were averaged. After matching, a final sample size of 14,767 was obtained.

During initial SDM development, we considered including month and station number (a proxy for distance from shore) as independent variables. While these factors improved model fit, we decided to exclude them because they would constrain future shifts in species distribution and phenology. Since our research goal was to assess model performance over a multidecadal period as a proxy to better understand how such models would perform when detecting future shifts in species distribution and seasonal occurrence, including independent variables that constrain such shifts would be counter to achieving this objective. Also, since many environmental variables in this ecosystem exhibit onshore-offshore gradients (McClatchie, 2013), multicollinearity between station number and environmental variables could also influence our ability to detect non-stationary relationships. Similarly, latitude and longitude were not included in SDMs as independent variables since they would also constrain future shifts in species distribution. Previous studies have shown that stock size can influence the amount of suitable habitat occupied by our target species (Weber and McClatchie, 2010; Weber and McClatchie, 2012; Muhling et al., 2020). However, since earth system models (ESMs) cannot directly project future stock size, this is not a covariate that could be easily included in a model of future changes in species distribution or phenology. Since our goal is to provide a framework for assessing performance of such models, we did not include stock size as a covariate here.

2.2 Classification of change points in ocean ecosystems

The term regime shift describes low-frequency and high-amplitude changes in biological and physical conditions. However, there are disagreements about key characteristics of regime shifts. Different authors use this term to describe stochastic processes characterized by red noise; non-linear, alternative stable states; changes at multiple levels of ecological organization (e.g., species, assemblage, community, ecosystem); and processes related to both external perturbations and internal reorganization of ecological communities (Collie et al., 2004; Overland et al., 2008). Due to this multiplicity of definitions, we used three approaches to determine if relationships between fish and the environment were stable across different regimes. Since most of our regime shifts were defined based on changes in time series, we use the terms regime shift and change point synonymously.

2.2.1 Pacific Decadal Oscillation

The PDO is the first principal component of detrended winter SST in the North Pacific (Hare et al., 1999). During the latter half of the 20th century, this index exhibited decadal variability characterized by predominantly negative values during 1947-1976 and positive values during 1977-1998. Negative (positive) PDO values correspond to cool (warm) conditions in the southern CCS. The 1976/1977 shift in PDO sign coincided with large changes in the abundance of marine organisms across several trophic levels (Chavez et al., 2003; McGowan et al., 2003). In the CalCOFI region, this shift was associated with a 1.0°C increase in temperature over the upper 50 m of the water column and a ZDV decline of 68.4 cm^3/1,000 m^3 (Figure S1). Statistically significant, albeit smaller, changes in mean salinity and DO coincided with this regime shift (Figure S1). Since 1998, the PDO has displayed oscillations at an interannual rather than decadal scale (Peterson, 2009). Furthermore, the PDO has recently exhibited a decreased correlation with North Pacific climatic and ecological indicators (Puerta et al., 2019; Litzow et al., 2020). Consequently, we assessed whether non-stationary relationships between fish and environmental variables were evident across the 1976/1977 shift but did not consider years after 1998.

2.2.2 Change points in oceanic variables

Beyond the PDO, we took an empirical approach to identify change points associated with regime shifts in times series of environmental variables and SSB. First, we estimated change points separately for temperature, salinity, DO, and ZDV. To accomplish this, we performed a principal component analysis (PCA) on each variable to identify its dominant mode of

temporal variability. Since PCA cannot be performed on datasets with missing observations, we binned data into seven groups that represented an onshore-offshore gradient. Our seven bins were based on the following CalCOFI stations: ≤40 (closest to shore), 40-50, 50-60, 60-70, 80-90, 90-100, and ≥100 (farthest offshore). Stations in each bin were annually averaged. In cases when no observations were available in a bin for a year, linear interpolation across the onshore-offshore gradient was used to fill this gap. The years 1951, 1984, and 1982 were removed due to persistent gaps in coverage. Such gaps were more widespread for DO than other variables, which necessitated removal of additional years (1953-1955, 1957, 1960, 1967, 1975, 1980-1981). PCA was performed after these data processing steps.

Change point analysis was applied to the first principal component of each environmental variable using the Bayesian change point detection algorithm developed by Ruggieri (2013). Change point analysis was performed in MATLAB (version R2017a). The Ruggieri (2013) algorithm detected changes in time series mean, variance, or slope. We used uninformative priors. Algorithm parameters were set such that a maximum of three change points could be detected over a time series and change points needed to be separated by ≥10 years. Other parameters were set following guidance from Ruggieri (2013) ($k_0 = 0.01$, $\nu_0 = 2$, and σ_0^2=observed variance). 500 iterations of this algorithm were run for each time series to generate posterior probability distributions. Subsequent analyses examining non-stationarity across regimes were based on the number of change points with the highest posterior probability and years with the highest probability of a change point. In a sensitivity test, parameters related to maximum number of change points and minimum regime duration were varied between 2-4 and 8-12 years, respectively. This was found to affect the years of some change points by ±3 years or less.

2.2.3 Change points in SSB

Change point analysis was also applied to assess whether habitat use among SPF varied as a function of stock size. For this analysis, we used stock assessment data from Thayer et al. (2017) for 1951-2015 for *E. mordax* and Crone and Hill (2015) for 1983-2014 for *S. japonicus*. For *S. sagax*, we combined data from three stock assessments to obtain information for 1951-1963 (Jacobson and MacCall, 1995), 1981-2008 (Hill et al., 2008), and 2009-2015 (Hill et al., 2018). No stock assessment was available for *T. symmetricus*, so this species was excluded from this analysis. SSB was log transformed prior to analysis since histograms indicated SSB had a log-normal distribution. Change point detection parameters were the same as listed above, except the minimum duration for a regime was set to five years for *S. sagax* and *S. japonicus* since shorter SSB time series were available. For *S. sagax*, results were not sensitive to the choice of the minimum regime duration or to the use of only the more recent stock assessments by Hill et al. (2008; 2018).

2.3 Species distribution modeling

We used generalized additive models (GAMs) to assess non-stationarity across change points. While a variety of SDMs exists, GAMs were selected because this technique has been widely used in fisheries science (e.g., Bell et al., 2015; Morley et al., 2018; McHenry et al., 2019). GAMs were run separately for each species and period associated with a change point to determine if there were differences in model characteristics across regimes. Since our goal was to examine environmental influences on species distribution, presence/absence of larvae was used as the response variable. Independent variables included temperature, salinity, DO, and log-transformed ZDV. Any bongo and ring net tows that did not have a full suite of environmental variables associated with them were removed from analysis. GAMs were formulated using the binomial family and logit link. GAMs were parameterized to have a maximum of four knots to prevent overfitting (Weber and McClatchie, 2010; Lindegren and Eero, 2013; Tommasi et al., 2015). This step was important because an overparameterized model is more likely to be non-stationary when that model is applied to a different period. The decision to limit the number of knots was a conservative choice aimed at decreasing the likelihood of detecting non-stationarity. For each species and regime, 16 GAMs with different combinations of environmental variables were run. The Akaike Information Criteria (AIC) was minimized to select which of these models was the most parsimonious and determine the number of knots to include in that model. If the AIC for several models differed by ≤2, we used a multi-model approach including results from several models (Burnham and Anderson, 2002). Akaike weights (w_i) for the selected models were examined to assess the degree of confidence in the selection process.

GAMs can be fit using either the gam or mgcv package in R (version 4.1.1). The latter uses a Bayesian approach for variance estimation, which results in smaller confidence intervals than those from the gam package (Wood, 2006). Since smaller confidence intervals may increase the likelihood of detecting differences across regimes, we used the gam package since it would provide more conservative results regarding non-stationarity. Nonetheless, a comparison of the gam and mcgv packages for *E. mordax* produced similar models. Tests for multicollinearity between independent variables, spatial autocorrelation, and inspection of GAM residuals for outliers are described in the Supplementary Material 1.1, Table S1 and Figure S2.

2.4 Indicators for detecting non-stationarity

We used six metrics to assess non-stationarity across regimes. These metrics evaluated whether there were changes

in: 1) variables included in SDMs; 2) linearity of partial environmental variable responses in SDMs; 3) relative importance of environmental variables; 4) response curve shape; 5) degree of responsiveness of fishes to a variable, and; 6) the range of conditions associated with maximum larval occurrence. Changes in any metrics between regimes was interpreted as an indicator of non-stationarity. In cases where multiple models were selected for a regime, differences needed to be observed amongst the full suite of candidate models for periods to be classified as non-stationary.

Each non-stationarity metric has pros and cons but when viewed together they provide a complementary and comprehensive picture of the occurrence of non-stationary environmental relationships. For example, some metrics are quantitative and can be evaluated for statistical significance, whereas other metrics are qualitative (e.g., response curve shape). Some metrics principally detect large changes in model formulation, such as the lack of significance of a previously important variable, whereas others identify subtler changes, such as a shift in the relative ranking of variables affecting fishes. By considering multiple metrics, one can avoid the pitfalls associated with any one metric. For example, changes in maximal larval occurrence or degree of responsiveness are more likely to be affected by extrema. Shifts in rank importance of environmental variables could be due to a small change among two variables with similar effect sizes (Planque et al., 2007). When using a combination of metrics, biases affecting a single metric can be avoided, producing more reliable results. Details on how each metric was calculated are provided below.

2.4.1 Inclusion of variables in SDMs

Model selection was based on AIC minimization.

2.4.2 Linearity

Selected model(s) could include an environmental variable with either one, two, or three equivalent degrees of freedom (edf) in its partial response function. An edf of 1 was indicative of a linear model, whereas increasing edfs indicated greater non-linearity (Hastie, 1991). Changes in edf between regimes were used to assess changes in linearity.

2.4.3 Relative importance of variables

To assess the relative importance of environmental variables, we compared the change in deviance (ΔD) in GAM outputs between a full model and models when one variable was removed. ΔD was compared across variables to assess the rank importance of variables. Changes in ranking between regimes were interpreted as a qualitative indicator of non-stationarity. This is a qualitative indicator because at times changes in rank can reflect small differences in ΔD among nearly equally ranked variables.

2.4.4 Response curve shape

Response curve shape refers to the graphical relationship between an environmental variable and the probability of fish occurrence. The y-axis of response curves was presented on a logit scale. Response curve shape was assessed in a semi-quantitative manner in two stages. First, we qualitatively inspected shifts in shape. This step went beyond looking at changes in linearity, maximum value of the response curve, and response curve amplitude. Secondly, we inspected the 95% confidence intervals of response curves to evaluate overlap between different periods. If the confidence intervals had a substantial amount of overlap, periods were classified as similar to each other regardless of qualitative differences in response curve shape. In contrast, if confidence intervals did not overlap in entirety and response curve shape also differed, this was interpreted as an indication of non-stationarity.

2.4.5 Degree of responsiveness

The degree of responsiveness of a fish to an environmental variable was estimated based on the amplitude of the SDM response curve. A larger amplitude suggested that a fish was more responsive to a variable. To assess whether this metric differed between periods, we ran a bootstrap analysis in which observations were selected randomly with replacement 1,000 times for each species and regime (Efron and Tibshirani, 1998). The number of observations randomly selected during each bootstrap iteration was the same as the sample size for each SDM (Table S2). No spatio-temporal weighting was used when resampling data during bootstrap analysis. GAMs were recalculated for each dataset and response curves were plotted. We performed this analysis only for the most parsimonious model(s) selected with the AIC. Bootstrap permutations were used to develop 95% confidence intervals for response curve amplitude. In cases where multiple models were selected based on AIC scores, bootstraps were run separately for each model and confidence intervals were constructed jointly across models by weighting each model based on w_i. A lack of overlap between confidence intervals across regimes was an indication of non-stationarity.

2.4.6 Range of environmental variables associated with maximum larval occurrence

The sixth non-stationarity metric was the range of an environmental variable that maximized the probability of fish occurrence. A bootstrap was used to determine environmental conditions associated with maximum larval occurrence across 1,000 SDM realizations. For each bootstrap iteration, we identified the maximum value of the response curve and the corresponding value of the environmental variable at this maximum. These values were sorted from smallest to largest and we identified the lower 2.5th and upper 97.5th percentiles of this empirical distribution. These 95% confidence intervals were

used to assess whether the range of conditions associated with maximum larval habitat suitability differed between regimes. Weighted means of confidence intervals were used in cases where multiple models were selected for a regime.

2.5 Future projections

An ESM was used to make future projections of habitat suitability. ESM projections focused specifically on quantifying uncertainty associated with ecological and climatic change points and determining their importance compared to other sources of projection uncertainty. ESM output was obtained from the World Climate Research Programme's Coupled Model Intercomparison Project – Phase 6 (CMIP6). CMIP6 output is publicly available from Lawrence Livermore National Laboratory.[3] Our criteria for model selection from the CMIP6 ensemble were that ensemble members needed to contain output on all environmental variables used in SDMs for a historical simulation (1980-1999) and two future simulations (2080-2099). The historical period was selected to be 100 years earlier than the period used for future simulations. The two future climate change scenarios considered were Shared Socioeconomic Pathways (SSP) 5-8.5 and 1-2.6, which corresponded, respectively, to a high-end greenhouse gas emissions scenario and a climate change mitigation scenario consistent with the Paris Agreement (O'Neill et al., 2016). When data were downloaded from the CMIP6 archive (18 December 2019), only one ESM had full data available for all four variables, all three simulations, and both 20-year periods. This model, known as CNRM-CERFACS-ESM2.1 (abbreviated name: CNRM-ESM2), was developed by the French National Centre for Meteorological Research and couples the CNRM-CM6-1 atmosphere-ocean general circulation model with the PISCESv2-gas ocean biogeochemistry model (Séférian et al., 2019). The ESM has an approximately 100-km latitudinal/longitudinal resolution and 75 depths. PISCESv2-gas tracks 26 biogeochemical state variables and four plankton functional groups (diatoms, nanophytoplankton, microzooplankton, and mesozooplankton).

Monthly CNRM-ESM2 data on environmental variables were extracted from the core CalCOFI region (29.8-35.2°N and 117.3-125.9°W). This included 63 model grid cells, resulting in a similar number of grid cells to the number of CalCOFI stations. CNRM-ESM2 included 19 depth layers over the upper 50 m of the water column. Shape-preserving piecewise cubic interpolation was used to calculate the temperature, salinity, and DO exactly at 50 m by interpolating between the 18th and 19th model depth layers. We computed the mean of each variable over the upper 50 m, weighting this average by the width of each depth layer. Units of DO and mesozooplankton concentration differed between CNRM–ESM2 and CalCOFI. Unit conversions were applied to allow CNRM-ESM2 output to be used as independent variables in GAMs developed for SPF species (Supplementary Material 1.2).

Many ESMs overestimate coastal temperatures and underestimate primary production in Eastern Boundary Upwelling Systems (Stock et al., 2011; van Oostende et al., 2018). To compensate for this, we performed a bias correction on variables from CNRM-ESM2 using the delta method (Hare et al., 2012). Biases were estimated using the monthly mean climatology from CalCOFI observations for 1980-1999. Next separate GAMs were run for each species and regime using CNRM-ESM2 data as independent variables. Projections were made for 1980-1999 and 2080-2099 with the SSP5-8.5 and SSP1-2.6 scenarios. Mean habitat suitability for SPF species was computed for each grid cell and month, with 95% confidence intervals based on variations between years during each period. In this context, habitat suitability is equivalent to the modeled probability of larval occurrence and has a range between 0 (larval absence) and 1 (larval presence). Spatio-temporally integrated habitat suitability (IHS) for a given year was also calculated by summing suitability scores across CRNM-ESM2 grid cells during spring (i.e., the peak season for occurrence of most SPF species, Supplementary Material 1.2). IHS is unitless and its value is dependent on the number of grid cells and months in the integration. In cases where multiple models were selected, IHS was calculated based on the weighted means of models. A two-way crossed ANOVA assessed whether SSP scenario and GAM model period had a significant effect on IHS. The mean coefficient of variation (CV) was calculated for the historical and SSP5-8.5 scenarios to assess if variations in IHS were projected to increase under unmitigated climate change. Mean CVs were calculated as a function of species, regime shift type, environmental variables, and indicators of non-stationarity. For environmental variables and non-stationarity indicators, CV calculations only included GAMs where there was some indication of non-stationarity for a particular variable or metric. Instances of non-stationarity associated with the rank importance of variables were not included in CV calculations since it was not possible to attribute changes to a single environmental variable.

3 Results

3.1 Change point detection

3.1.1 Oceanic variables

Across all oceanic variables, the first principal component (PC1) of their time series accounted for 63.3-91.2% of variance, whereas the second principal component (PC2) accounted for a reduced percentage of variance (4.9-17.2%; Table 1). PC1

3 https://esgf-node.llnl.gov/projects/cmip6/

TABLE 1 Principal components analysis (PCA) performed on environmental variables binned by onshore-offshore strata.

		CalCOFI station numbers						
Principal Component (PC)	**Variance explained (%)**	**≤40**	**40-50**	**50-60**	**60-70**	**70-80**	**80-90**	**>90**
Temperature PC1	72.6	0.389	0.366	0.383	0.359	0.370	0.382	0.397
Temperature PC2	14.4	-0.230	-0.445	-0.320	-0.005	0.004	0.163	0.788
Salinity PC1	73.8	0.335	0.295	0.315	0.429	0.420	0.443	0.381
Salinity PC2	11.1	-0.506	-0.522	-0.237	-0.005	0.113	0.330	0.542
Oxygen PC1	63.3	0.415	0.261	0.452	0.362	0.386	0.373	0.369
Oxygen PC2	17.2	0.393	0.671	0.188	-0.306	-0.376	-0.333	-0.117
Zooplankton PC1	91.2	0.355	0.377	0.379	0.391	0.404	0.393	0.343
Zooplankton PC2	4.9	-0.576	-0.408	-0.182	0.102	0.181	0.296	0.579

Data on the percent variance explained by each principal component (PC) and loadings of the PC on each stratum are presented above.
Strata are indicated by station numbers from California Cooperative Ocean Fisheries Investigations (CalCOFI).

captured region-wide variations in temperature, salinity, DO, and ZDV at an interannual scale. PC2 was characterized by onshore-offshore differences where nearshore and offshore stations exhibited PCA loadings in different directions. This pattern was consistent across PC2 for all variables.

Each oceanic variable's principal component time series exhibited distinct temporal patterns (Figure 2). PC1 for temperature was primarily negative at the start of the time series, exhibited mainly positive values during the warm phase of the PDO between 1977-1998, displayed anomalies centered around zero during much of the 2000s and early 2010s, and rose sharply at the end of the time series in 2014-2015 coincident with MHW onset (Figure 2A; Di Lorenzo and Mantua, 2016). In contrast, PC1 of salinity was less closely correlated with the PDO, as has also been shown by Di Lorenzo et al. (2008). Instead, this PC exhibited greater variability at the interannual rather than decadal scale (Figure 2B). PC1 for DO was characterized by heightened variability at the start of the time series, with greater stability in more recent years (Figure 2C). Similar to the results for temperature, the PDO seemed to have a substantial influence on the zooplankton PC1 (Pearson correlation coefficient r=-0.49, p=0.0001, d.f. = 54). Zooplankton PC1 was characterized primarily by positive anomalies up until the mid-to-late 1970s and experienced a

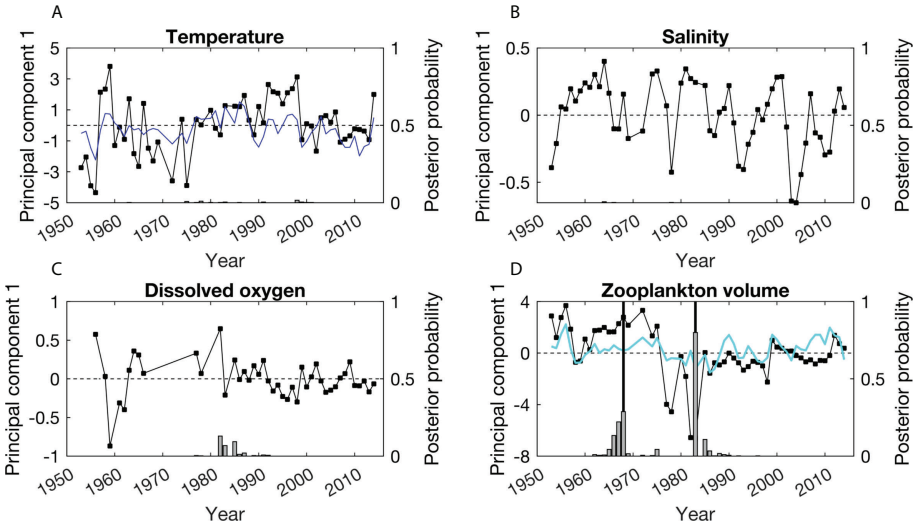

FIGURE 2
Time series of the first principal component of **(A)** temperature, **(B)** salinity, **(C)** dissolved oxygen (DO) concentration, and **(D)** mesozooplankton displacement volume from the southern California Current System. Note that there are some gaps in DO measurements during the early years of the CalCOFI time series. Horizontal, dashed lines indicate principal component scores of zero, while thick, vertical lines represent the timing of break points identified in time series. Gray bars show the posterior probability of a change point occurring each year in the time series of each oceanic variable. The winter Pacific Decadal Oscillation (PDO) is included as a blue line in **(A)** and its inverse is included as a turquoise line in **(D)** to illustrate correlations among principal components and this regional climate index.

period dominated by negative anomalies after the PDO entered its warm phase (Figure 2D).

The change point detection algorithm did not identify any regime shifts in the PC1 time series of temperatures, salinity, or DO. There was a 93.6% probability of zero change points detected in the temperature time series, 99.3% probability of zero salinity change points, and 63.7% probability of zero DO change points. In contrast, the posterior probability distribution indicated a 75.8% probability of two change points in the zooplankton PC1 time series, with 24.2% chance of one change point. The highest probabilities of change points were detected in 1968 and 1983. Prior to 1968, the PC1 time series for ZDV consistently exhibited positive anomalies (Figure 2D). During 1969-1983, ZDV was characterized by highly variable and declining abundance, while after 1983 this time series was fairly stable with anomalies close to zero.

3.1.2 SSB

With a posterior probability of 78.1%, one change point was detected in the time series of *E. mordax* SSB (Figure 3A). This change occurred in 1963, separating a period of low, but recovering SSB from a period when this species was fairly abundant. A decline in *E. mordax* biomass was observed at the end of this time, but there was only a 21.9% posterior probability that this decline was associated with a second change point.

For *S. sagax*, there was a 99.9% probability that its time series contained two change points, which were detected in 1963 and 1997 (Figure 3B). The 1963 change was associated with a decline in *S. sagax* biomass and its subsequent recovery. The precise date of this change is uncertain because of a discontinuity in the *S. sagax* time series due to a lack of stock assessments between 1964-1980. However, the fact that *E. mordax* also exhibited a change point during 1963 bolsters confidence in this result for *S. sagax* and suggests asynchronous dynamics between species. The second change point for *S. sagax* detected in 1997 was associated with stable, high fish biomass, with some declines near the time series end.

Log-transformed *S. japonicus* SSB was in decline throughout most of the period when biomass estimates were available (Figure 3C). With a posterior probability of 92.5%, no change points were detected for *S. japonicus*.

3.2 Non-stationarity detection using GAMs

Assessment of non-stationarity in models of all four species for each of the three types of regime shifts is described in the Supplementary Material 2.1-2.3 and Figures S3-11. Here we provide an in-depth, illustrative summary for one species as a case study and then compare general trends across all species and regime shift types.

3.2.1 Case study – changes points in *S. sagax* SSB

For each SSB regime, a single model was selected for *S. sagax* where the selected GAM had an Akaike weight >0.8 (Table S3). This indicated a >80% likelihood that the selected model was the most parsimonious choice of the candidate models.

Evidence of non-stationarity in how *S. sagax* relates to oceanic variables was found across all indicators. For the first indicator (inclusion of different variables in the selected GAM), non-stationarity was indicated by the fact that the model formulation changed across regimes. During the first two SSB regimes (1951-1963 and 1964-1997), temperature, salinity, and ZDV were included in the selected model, but DO was excluded (Table S2). In contrast, during the regime from 1998-2015, ZDV was excluded from the model.

The second indicator of non-stationarity was related to changes in whether fishes had linear or non-linear relationships with oceanic variables. In most models, the best-fit GAM included non-linear terms, with an edf of 3 (Table S2). Evidence of non-stationarity was observed since salinity initially had a linear relationship with larvae occurrence, which later became non-linear (Table S2; Figure 4).

Non-stationarity changes in the ranked importance of oceanic variables were also observed. Ranking of salinity declined over time, while DO ranking increased (Figure 5J). Temperature and ZDV exhibited variability in their ranking, but without long-term trends.

Changes in response curve shape was the fourth indicator of non-stationarity. Temperature response curves had a negative, parabolic shape during the 1951-1963 and 1964-1997 regimes. During 1998-2015, the temperature response curve had a flatter shape, and a higher probability of encountering *S. sagax* larvae at low temperatures was observed (Figure 4). The flattened response curve shape during the third regime may indicate a reduced influence of temperature on sardine distribution, which is also consistent with changes in the relative ranking of temperature during this regime (Figure 5J). *S. sagax* were most frequently encountered at higher salinities throughout all periods, but the salinity response curve shape changed across periods. During 1951-1963, this species had a positive, linear relationship with salinity; during 1964-1997, this relationship had a negative, parabolic form; from 1998-2015, *S. sagax* distribution was less responsive to variations in salinity as indicated by a flattened response curve (Figure 4). Less change in response curve shape was observed for ZDV since it exhibited a negative, parabolic response curve during both periods when included in GAMs (Figure 4). Changes in curve shape could not be assessed for DO, since this variable was only included in the selected model during the third SSB regime.

Changes in the amplitude (or range) of the response curve was the fifth indicator of non-stationarity. A decrease in response curve amplitude is suggestive of a reduced influence of a variable on larval fishes. For temperature, response curve range was significantly larger during 1964-1997 than 1998-2015 (Figure 6I). The period when *S. sagax* was most sensitive to

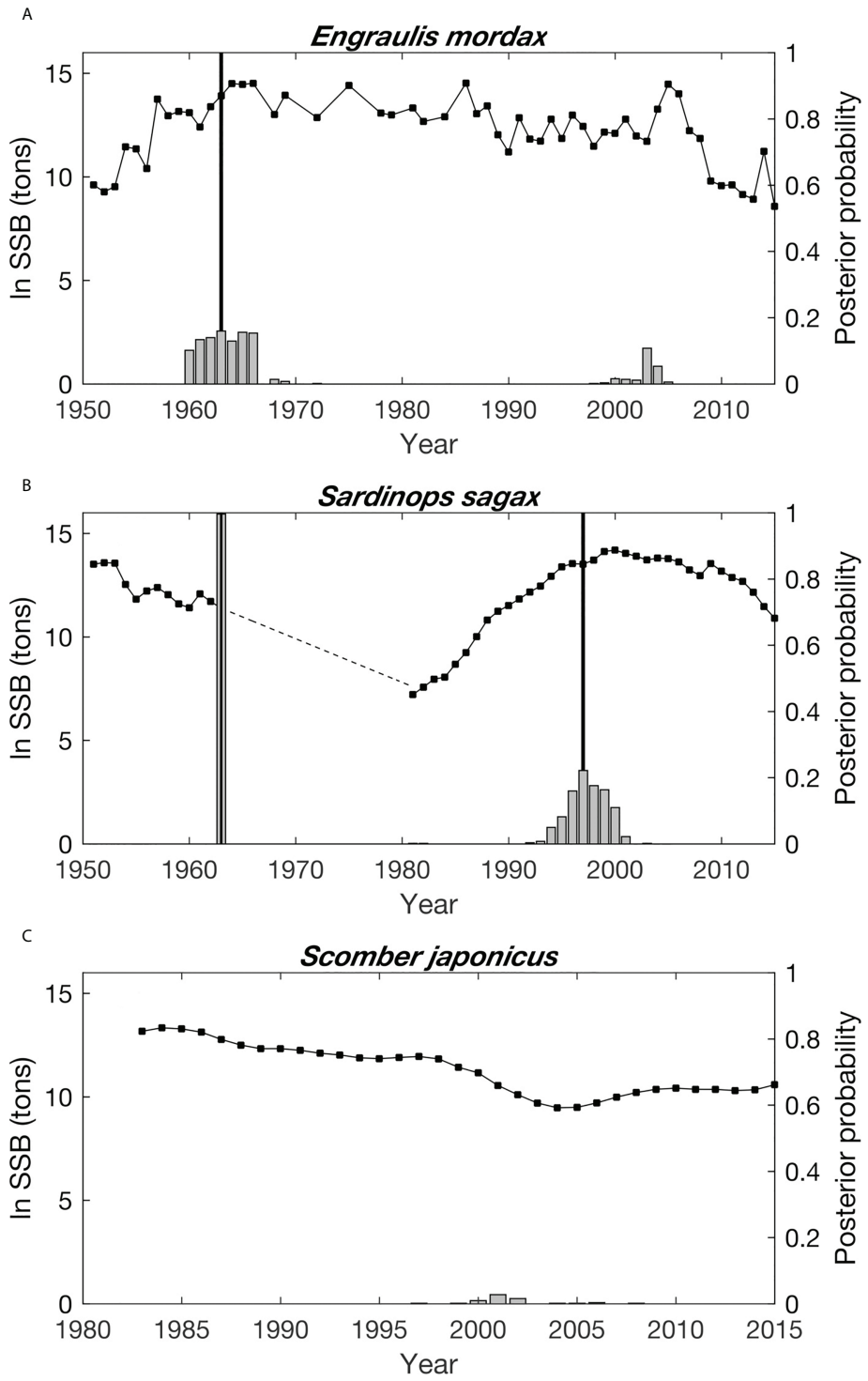

FIGURE 3
Time series of the natural log transformed spawning stock biomass (SSB) of **(A)** (*E*) *mordax*, **(B)** *S. sagax*, and **(C)** *S. japonicus*. No SSB data are available for *T. symmetricus*. Dashed line indicates the time period of low *S. sagax* biomass when no stock assessments were conducted to estimate this species' SSB. Gray bars show the posterior probability of a change point in the SSB time series occurring each year. Black, vertical lines indicate the timing of break points identified in each time series.

temperature based on this indicator coincided with low biomass of this species (Figure 3B). No significant changes were seen in response curve range for salinity and ZDV. Changes could not be assessed for DO since it was only included in the selected model during a single regime.

Significant changes in the sixth indicator of non-stationarity (shifts in the peak of the response curve) were observed for several oceanic variables. For temperature, *S. sagax* was most commonly found in areas with significantly cooler temperatures during 1998-2015 compared to prior periods (Figure 7I). The maximum likelihood of detecting larvae occurred at significantly lower salinities in 1964-1997 than 1951-1963 (Figure 7J). Sardine larvae were found in areas with significantly less zooplankton during 1964-1997 than 1951-1963 (Figure 7L). Since the former period was characterized by reduced ZDV (Figure S1), this might reflect a change in the availability of zooplankton rather than an active shift in habitat selection.

3.2.2 Comparisons across species and regime shift types

Every combination of species and regime type exhibited at least one indication of non-stationarity, implying that non-stationarity is ubiquitous across SPF in the CCS. A summary of patterns observed across non-stationarity indicators, oceanic variables, species, and regime types is included below.

A change in oceanic variables included in GAMs was observed across 60% of the combinations of species and regime shifts (Table 2). Nearly half of the selected of the selected models contained all four environmental variables, but in several cases the most parsimonious model(s) excluded DO or ZDV (Table S2). In a smaller number of cases, a simplified model containing 1-2 environmental variables was selected.

Changes in the linearity of the relationship between fishes and environmental variables also occurred across 60% of the combinations of species and regime shifts (Tables 2, S2). Salinity and DO were the most common variables to exhibit changes in linearity.

Changes in the ranked importance of oceanic variables were very common, with evidence of non-stationarity occurring across all species (Figure 5). Temperature and salinity were frequently ranked as having the greatest or second greatest influence on fish larvae, with lower rankings more common among DO and ZDV. Among *S. sagax* and *T. symmetricus*, the relative ranking of DO increased during recent periods.

Changes in response curve shape were observed across 80% of species and regime combinations (Table 2). The only cases where pronounced changes in response curve shape were not detected was among shifts between PDO phases for *E. mordax* and *S. japonicus* (Figures S3, S5). Of the four oceanic variables,

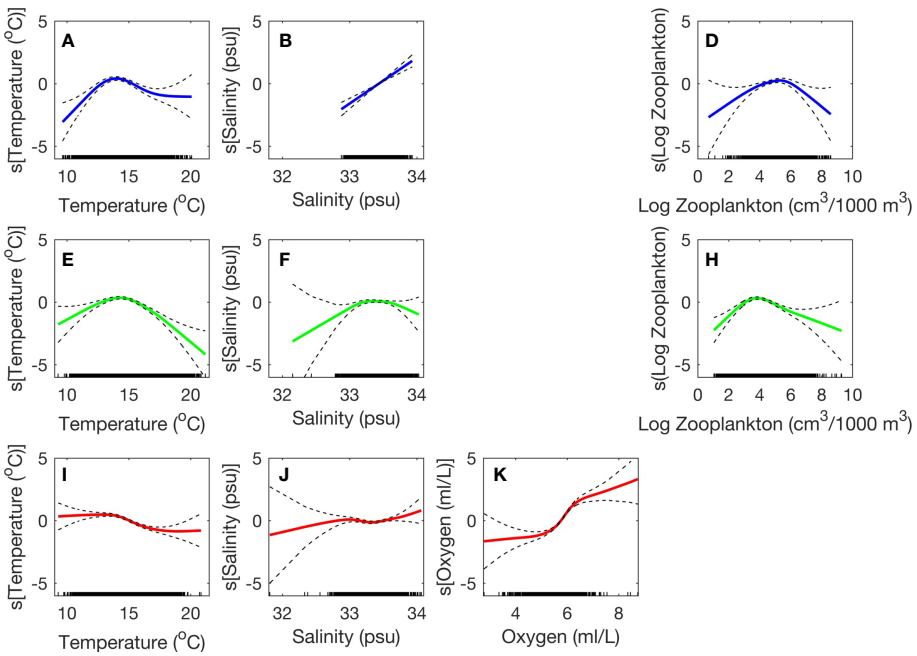

FIGURE 4
Generalized additive model (GAM) response curves for *S. sagax* during three different spawning stock biomass (SSB) change points: 1951-1963 (**A-D**; blue), 1964-1997 (**E-H**; green), and 1998-2015 (**I-K**; red). Dashed lines indicate that 95% confidence intervals for each response curve. Missing subplots (log zooplankton during 1998-2015 and oxygen in 1951-1963 and 1964-1997) are indicative that a particular oceanic variable was not included in the most parsimonious GAM. Rug plots are displayed at the bottom of each subplot.

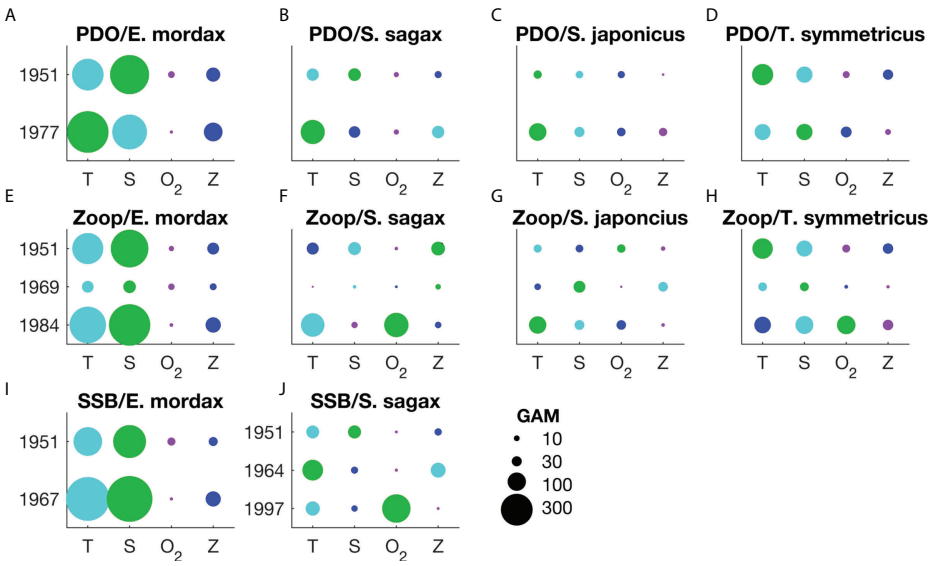

FIGURE 5
Rank order comparison between the influence of each oceanic variable on the presence/absence of larvae of *E. mordax*, *S. sagax*, *S. japonicus*, and *T. symmetricus*. Results are shown for change points designated based on changes in the sign of the Pacific Decadal Oscillation (PDO; **A-D**); break points in the mesozooplankton volume time series (**E-H**; the abbreviation "zoop" is used when labeling the title of these subplots), and; break points in the time series of *E. mordax* and *S. sagax* spawning stock biomass (SSB; **I, J**). Unless otherwise specified, time periods for each type of change point are the same across all species. Only the start year of a particular regime is listed here. Oceanic variables are abbreviated as follows: T – temperature, S – salinity, O_2 – dissolved oxygen concentration, Z – mesozooplankton volume. Comparisons between variables are based on the change in deviance (ΔD) when one variable is removed relative to the deviance of the full model. The scale for ΔD is shown in the lower, right corner of the figure. Note that ΔD is influenced by sample size so this metric is comparable across from a single regime, but not across multiple regime types due to variations in sample size. The rank order of different environmental variables for each period is shown based on circle size and color: green – 1st rank, turquoise – 2nd rank, blue – 3rd rank, purple – 4th rank.

temperature was the least likely to have a change in response curve shape, usually displaying a negative, parabolic shape (Figures 4, S3-S11). Like temperature, ZDV often exhibited a negative, parabolic response curve shape, especially at the start or mid-point of time series. In many cases (e.g., Figures S6-S8, S11), ZDV response curves displayed a flatter shape during later periods, indicating a reduced influence of this variable. The response curves for salinity and DO usually displayed wide confidence intervals at extrema, indicating reduced certainty in how fishes respond to these variables under conditions deviating from the mean. Lastly, compared to other species, *S. sagax* displayed a greater propensity for changes in response curve shape (Figures 4, S4, S8).

The amplitude of response curves, which is an indicator of sensitivity to oceanic variables, displayed non-stationarity across four of the ten combinations of species and regime shifts (Table 2). Only one significant change in this indicator was observed across PDO and SSB regimes, whereas deviations from stationarity were more common among zooplankton regimes (Figure 6). Deviations from stationarity for this indicator were most common among *S. sagax*.

Shifts in peak habitat use tied other indicators for the second most incidences of non-stationarity. This indicator refers to changes in the range of environmental variables associated with maximum larval occurrence. For 80% of species and regime shift combinations, at least one oceanic variable exhibited non-stationarity for this indicator (Table 2). Multiple species exhibited changes in the temperature and salinity at which their response curve peaked (Figure 7), but no overarching pattern of change between periods was identified amongst these variables. In contrast, whenever there was a significant change in peak DO use, fishes tended to occur in areas with higher DO in more recent years (Figures 7G, K). In four out of five cases where there was a significant change in peak use of ZDV, fishes occurred in areas with less ZDV during more recent years (Figures 7D, H, L). This may be related to long-term declines in ZDV in this ecosystem (Roemmich and McGowan, 1995; Lavaniegos and Ohman, 2007). Compared to other species, *S. sagax* was most likely to display significant changes in this indicator.

When integrating across all indicators, *S. sagax* was the species whose relationship with oceanic variables displayed the most signs of non-stationarity (Table 2). *S. japonicus* and *E. mordax* displayed the fewest indications of non-stationarity, even though some non-stationarity was detected for them across >50% of the indicators and regime shift types. Non-stationarity was most

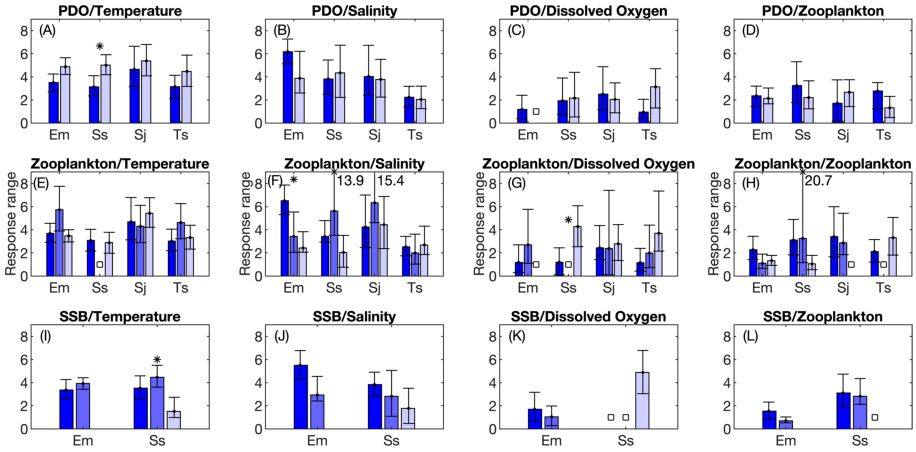

FIGURE 6
Changes between periods in the response curve range from generalized additive models (GAMs). Response curve range is defined as the difference between the maximum and minimum value in a GAM response curve and is indicative of how strongly an environmental variable influences larval fish occurrence. Median values and 95% confidence intervals from bootstrap analysis are shown. Results are shown for change points designated based on changes in the sign of the Pacific Decadal Oscillation (PDO; **A-D**); break points in the mesozooplankton volume time series **(E-H)**, and; break points in the time series of *E. mordax* and *S. sagax* spawning stock biomass (SSB; **I-L**). GAM results for different periods are displayed in groups, with the first period represented by the left most bar in a group (dark blue color) and the last period displayed to the right (light lavendar color). Intermediate periods are displayed in the middle of each group. Stars (*) indicate that periods are significantly different from each other for a given species and environmental variable based on non-overlapping 95% confidence intervals. White squares indicate that a particular variable was not included in the best fit GAM model(s). Numbers shown in some subplots indicate the maximum response curve range in a few cases where the maximum value exceeds the y-axis limit of a graph. Species names are abbreviated based on the first letter of the genus and the first letters of the species name: Em, *Engraulis mordax*; Ss, *Sardinops sagax*; Sj, *Scomber japonicus*; Ts, *Trachurus symmetricus*.

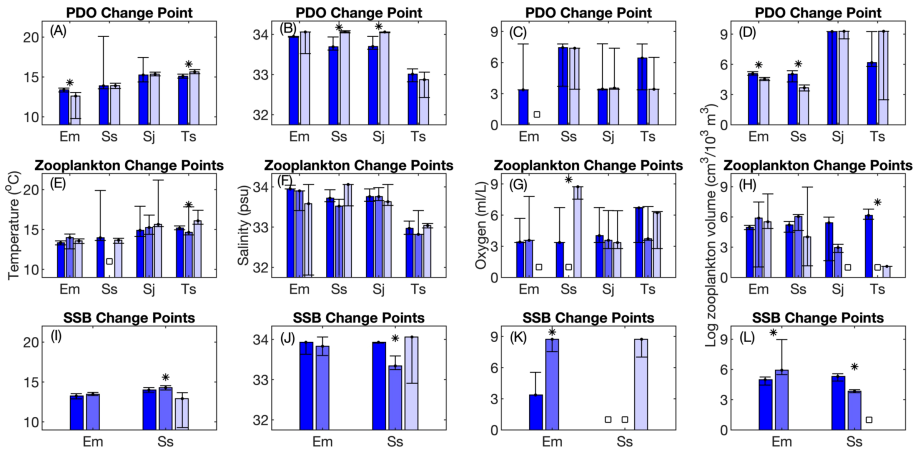

FIGURE 7
Changes between periods in the peak value of generalized additive model (GAM) response curves. The peak in response curves is indicative of the environmental conditions that maximize the likelihood of occurrence of *E. mordax*, *S. sagax*, *S. japonicus*, and *T. symmetricus*. Median values and 95% confidence intervals from bootstrap analysis are shown. Results are shown for change points designated based on changes in the sign of the Pacific Decadal Oscillation (PDO; **A-D**); break points in the mesozooplankton volume time series **(E-H)**, and; break points in the time series of *E. mordax* and *S. sagax* spawning stock biomass (SSB; **I-L**). Bar colors, symbols, and species name abbreviations are the same as in Figure 6.

TABLE 2 Percent incidence of non-stationarity by indicator metric, species, oceanic variable, and change point type for generalized additive models (GAMs).

(A) Percent incidence of non-stationarity by metric	
Variables included in model	60%
Degree of non-linearity	60%
Rank order of deviance explained	70%
Response curve shape	80%
Degree of responsiveness	40%
Peak environmental range	80%
(B) Percent incidence of non-stationarity by species	
Engraulis mordax	50%
Sardinops sagax	89%
Scomber japonicus	50%
Trachurus symmetricus	67%
(C) Percent incidence of non-stationarity by oceanic variable	
Temperature	25%
Salinity	33%
Dissolved oxygen	37%
Zooplankton volume	33%
(D) Percent incidence of non-stationarity by type of change point	
PDO	58%
Zooplankton volume	67%
SSB	75%

In (a), (b), and (d), non-stationarity is assessed at the model level, whereas in (c) it is assessed across each oceanic variable included in a model.

common among salinity and DO, whereas the relationship between fish presence/absence, temperature, and ZDV exhibited slightly more stability. Among different regimes, non-stationarity was observed most frequently for SSB regimes when integrated across indicators (Table 2).

3.3 Future projections

CNRM-ESM2 was used to produce end of the 21st century projections of suitable habitat for larval fishes and assess whether these projections differed significantly depending on which ecological or climatic regime was used to parameterize projection models. For *E. mordax*, *S. sagax*, and *S. japonicus*, habitat suitability declined during future projections, with a steeper loss in suitable habitat under SSP5-8.5 (Figure 8). For this scenario, decreases in mean IHS varied between 40.5-90.8% relative to the historical baseline. Under SSP1-2.6, declines in suitable habitat never exceeded 53.1% for any species or regime. In contrast to other species, *T. symmetricus* habitat suitability was projected to increase under SSP1-2.6 and SSP5-8.5 during spring (Figures 8D, H).

Two-way ANOVAs indicated that GAM model choice had a significant effect on habitat suitability in most cases (Table 3). The two exceptions to this occurred among *E. mordax* during regimes defined by PDO and SSB changes. For most species and regime shift types, *F* statistics from ANOVAs were larger for the GAM effect than the SSP effect, implying that the period used to parameterize the GAM had a larger impact on habitat suitability than SSP scenario. Furthermore, most species exhibited significant interactions between SSPs and GAMs from different regimes. One common pattern among interaction terms was that GAMs parameterized during periods with greater habitat suitability tended to undergo larger changes under future climate scenarios.

Changes in the mean CV between the historical and SSP5-8.5 scenarios were assessed to determine if variability in suitable habitat may increase under climate change. Increased variability was observed for all species, except *T. symmetricus*, under SSP5-8.5 (Table 4A). Variance in IHS was greater under regimes defined by changes in ZDV than other types of regimes (Table 4B). Regimes characterized by non-stationarity in salinity and ZDV exhibited greater variability than regimes with non-stationarity in temperature and DO (Table 4C). However, many regimes exhibited concurrent non-stationarity across multiple environmental variables, making it challenging to partition these effects among variables. The largest increases in variability under climate change were observed when there was non-stationarity associated with shifts in which variables were included in GAMs and changes in response curve amplitude (Table 4D).

4 Discussion

Non-stationary relationships between organismal distribution and climate can result in inaccurate projections of how species respond to climate change, but this subject has not been widely investigated across ecosystems (Litzow et al., 2019). We found that indications of non-stationarity were nearly ubiquitous among SPF species when models were constructed for three types of regime shifts. Non-stationarity most frequently resulted in changes in response curve shape, shifts in the peak range of conditions where larvae occurred, and changes in the relative importance of oceanic variables. Non-stationarity was most frequently associated with changes in ecological conditions, such as shifts in fish SSB or ZDV, rather than changes in the PDO. Relationships between fishes and temperature were more stable than other environmental variables. This might partially reflect greater uncertainty in relationships between fish distribution, salinity, and DO, which is indicated by the large confidence intervals associated with these variables' response curves. For several combinations of regimes and species, DO had a greater influence on

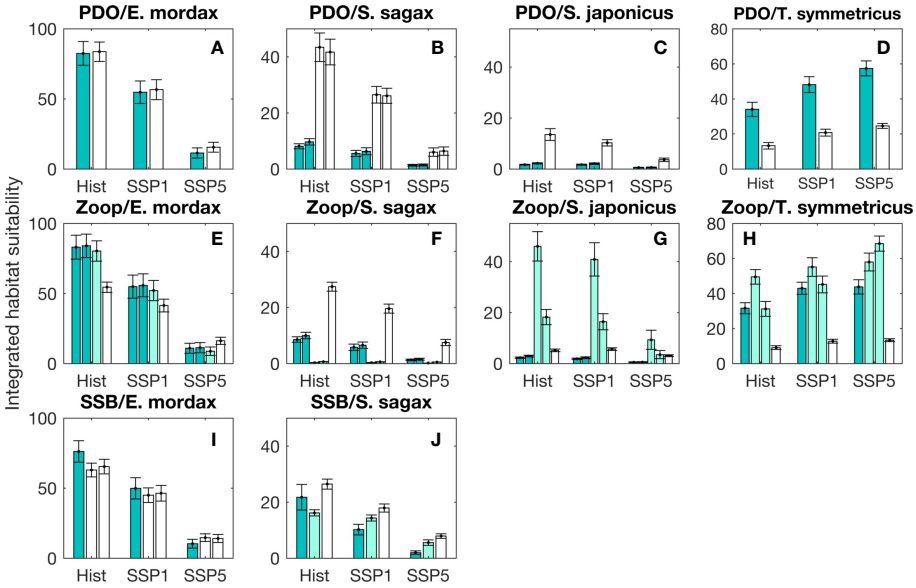

FIGURE 8
Integrated habitat suitability (IHS) for larval fishes during spring months based on projections from the CNRM Earth System Model (CNRM-ESM-2-1). Annual mean IHS scores (± 95% confidence intervals) are shown for a historical simulation for the years 1980-1999 (abbreviated as Hist) and two future simulations (SSP1-2.6 and SSP5-8.5) for the years 2080-2099. Using the climate forcing from each CNRM-ESM-2-1 simulation, habitat suitability was projected based on generalized additive models (GAMs) parameterized with data from different regimes and species. Results are shown for change points designated based on changes in the sign of the Pacific Decadal Oscillation (PDO; **A-D**); break points in the mesozooplankton volume time series (**E-H**; abbreviated as Zoop), and; break points in the time series of *E. mordax* and *S. sagax* spawning stock biomass (SSB; **I, J**). GAM results for different periods are displayed in groups, with the first period represented by the left most bar in a group (green color) and the last period displayed to the right (white color). Intermediate periods are displayed in the middle of each group (pale green color). In cases where multiple models were selected for a particular regime, separate bars are shown for each model using the color coding described above.

distribution in recent years (Figures 5, 7). Often the effects of non-stationarity on larval habitat suitability were larger than changes projected under high and low greenhouse gas emissions.

4.1 Non-stationary fish-environment relationships

Among fishes, non-stationarity can affect how environmental factors influence species distribution, recruitment, and fisheries productivity. Here we integrate our discussion across these types of non-stationarity. While non-stationarity has not been frequently considered in the scientific literature, when it has been investigated, results are similar to ours in that changes in organismal-environmental relationships are widespread. In studies comparing whether fish and invertebrate density, biomass, recruitment, and catch can be best modeled with stationary or non-stationary models, there is a pattern where the best fit model is usually non-stationary (Ciannelli et al., 2007; Lindegren and Eero, 2013; Beggs et al., 2014; Litzow et al., 2018; van der Sleen et al., 2018; Puerta et al., 2019). Similar results have been seen among non-marine taxa. For example, among British butterflies, changes in distribution in response to warming were not consistent across periods (Mair et al., 2012).

Among SPF, non-stationarity has been observed in multiple ecosystems and may be related to the boom-bust cycles of abundance common to this functional group. In the Northwest Atlantic, Atlantic menhaden (*Brevoortia tyrannus*) occurrence has a non-stationary relationship with temperature modulated by the North Atlantic Oscillation (Roberts et al., 2019). Changes in sardine (*S. sagax*) and anchovy (*E. encrasicolus*) spawning habitat preferences in the southern Benguela could be partially, but not fully, explained by warming, suggesting non-stationarity relationships occur among these stocks (Mhlongo et al., 2015). Among Japanese anchovy (*E. japonicus*), temperature where fish occurred as eggs and larvae differed between 1978-1991 and 1992-2004, which is suggestive of non-stationarity (Takasuka et al., 2008).

It is unclear whether SPF are more likely to exhibit non-stationary dynamics than other fishes. SPF are adapted to environments with a high degree of climate variability (Checkley et al., 2017), which could be indicative of resilience to fluctuating conditions. Conversely, SPF are more subject to population collapse than other fishes (Pinsky

TABLE 3 Two-way crossed analysis of variance (ANOVA) examining interactions between shared socioeconomic pathway (SSP) simulations and projections from generalized additive models (GAMs) trained during different ecological and climatic regimes.

Term	Sum of squares	d.f.	Mean squares	*F*	*p*
PDO regime shifts – *E. mordax*					
GAM	174.0	1	174.0	0.8	0.3823
SSP	98,191.9	2	49,096.0	217.1	<0.0001
GAM*SSP	47.7	2	23.8	0.1	0.9001
PDO regime shifts – *S. sagax*					
GAM	11,829.7	1	11,829.7	366.9	<0.0001
SSP	9,735.6	2	4,867.8	151.0	<0.0001
GAM*SSP	4,363.3	2	2,181.6	67.7	<0.0001
PDO regime shifts – *S. japonicus*					
GAM	1,723.5	1	1,723.5	265.5	<0.0001
SSP	681.7	2	340.8	52.5	<0.0001
GAM*SSP	361.7	2	180.9	27.9	<0.0001
PDO regime shifts – *T. symmetricus*					
GAM	21,853.1	1	21,853.1	393.3	<0.0001
SSP	6,135.2	2	3,067.6	55.2	<0.0001
GAM*SSP	723.9	2	361.9	6.5	0.0021
Mesozooplankton volume regime shifts – *E. mordax*					
GAM	5,140.4	2	2,570.2	14.5	<0.0001
SSP	112,616.6	2	56,308.3	316.6	<0.0001
GAM*SSP	7,600.6	4	1,900.2	10.7	<0.0001
Mesozooplankton volume regime shifts – *S. sagax*					
GAM	9,920.9	2	4,960.4	943.7	<0.0001
SSP	2,697.0	2	1,348.5	256.5	<0.0001
GAM*SSP	2,011.0	4	502.8	95.6	<0.0001
Mesozooplankton volume regime shifts – *S. japonicus*					
GAM	9,842.1	2	4,921.1	225.0	<0.0001
SSP	2,637.9	2	1,318.9	60.3	<0.0001
GAM*SSP	2,943.3	4	735.8	33.6	<0.0001
Mesozooplankton volume regime shifts – *T. symmetricus*					
GAM	47,321.5	2	23,660.7	411.3	<0.0001
SSP	6,621.6	2	3,310.8	57.6	<0.0001
GAM*SSP	3,343.8	4	835.9	14.5	<0.0001
SSB regime shifts – *E. mordax*					
GAM	445.5	1	445.5	2.7	0.1025
SSP	68,597.3	2	34,298.7	208.6	<0.0001
GAM*SSP	1,194.7	2	597.3	3.6	0.0295
SSB regime shifts – *S. sagax*					
GAM	1,337.2	2	668.6	33.9	<0.0001
SSP	8,009.1	2	4,004.6	202.9	<0.0001
GAM*SSP	670.3	4	167.6	8.5	<0.0001

The ANOVA response variable is the habitat suitability for larval fish species integrated over each year of the 20-year period examined by each SSP simulation.

and Byler, 2015), suggesting highly non-linear and unstable dynamics. Fernandes et al. (2020) showed that SDMs have a reduced capacity to predict the normalized biomass of pelagic species compared to benthic species. However, the mechanism behind this observation is unclear and could be due to either greater non-stationarity among pelagic fishes or differences in sampling efficacy.

4.1.1 Non-stationarity in the California Current System

Within the CCS, evidence has previously suggested that non-stationarity may be common among *S. sagax*, but much less research has investigated dynamics of other SPF. One early publication indicating that *S. sagax* has a variable relationship with environmental conditions is Lynn (2003) who found that

TABLE 4 Mean coefficient of variation (CV) for GAM model projections of annual integrated habitat suitability (IHS) under the historical and SSP5-8.5 climate scenarios.

(A) Species

	E. mordax	*S. sagax*	*S. japonicus*	*T. symmetricus*
Historical scenario				
Mean	0.24	0.67	0.96	0.54
CV				
SSP5-8.5				
Mean	0.61	0.88	1.04	0.50
CV				

(B) Change point type

	PDO	Zooplankton volume	SSB
Historical scenario			
Mean	0.59	0.69	0.29
CV			
SSP5-8.5			
Mean	0.71	0.88	0.58
CV			

(C) Environmental variables

	Temperature	Salinity	Dissolved oxygen	Zooplankton volume
Historical scenario				
Mean	0.55	0.68	0.51	0.65
CV				
SSP5-8.5				
Mean	0.71	0.83	0.72	0.87
CV				

(D) Non-stationarity metric

	Metric 1 (variables in GAM)	Metric 2 (linearity)	Metric 3 (rank order)	Metric 4 (response curve shape)	Metric 5 (response curve range)	Metric 6 (peak value)
Historical scenario						
Mean	0.56	0.66	0.62	0.58	0.58	0.55
CV						
SSP5-8.5						
Mean	0.79	0.80	0.78	0.75	0.82	0.71
CV						

In (C) and (D), only models for which there is some evidence of non-stationary are included in the means.
Mean CVs are presented by (A) species, (B) change point type, (C) environmental variable, and (D) non-stationary metric.

SST delimits the northern extent of *S. sagax* spawning habitat, but that the specific limit differs between years. Several studies have documented that the relationship between temperature and *S. sagax* recruits per spawner is sensitive to time period and source of temperature data (Jacobson and MacCall, 1995; McClatchie et al., 2010; Lindegren and Checkley, 2012; Zwolinski and Demer, 2019). Muhling et al. (2020) found indications of non-stationarity for *S. sagax* during the 2014-2017 MHW when fish occurred at temperatures warmer than projected by SDMs. Our results expand upon Muhling et al. (2020) by identifying changes in the sensitivity of *S. sagax* to environmental variables during earlier periods, indicating that non-stationarity during the MHW was not solely due to the inability of *S. sagax* to avoid unfavorable habitats during rapid change. Our results also confirm that non-stationarity among *S. sagax* can occur in absence of novel environmental conditions,

such as those associated with an MHW. Instead, non-stationarity likely emerges due to interplay between multiple factors (e.g., variations in population size, prey availability, interactions between oceanic conditions, shifts in where and when fish spawn).

Our results help explain some contradictions between earlier publications on SPF spawning habitat. There is generally a consensus that the northern stock of *S. sagax* spawns at 12-16° C, with several publications indicating peak spawning at temperatures around 13-14° C (Checkley Jr et al., 2000; Lynn, 2003; Reiss et al., 2008; Zwolinski et al., 2011; Asch and Checkley Jr, 2013). Our results are consistent with this consensus, although there are variations between periods in how quickly optimal spawning habitat declines at temperatures moving away from this peak. *E. mordax* generally spawn at 12-18° C, but the exact range of temperatures occupied by this species varies between studies, which may reflect variations in the rate at which response curves decline moving away from peak temperatures (Fiedler, 1983; Lluch-Belda et al., 1991; Checkley Jr et al, 2000; Reiss et al., 2008; Weber and McClatchie, 2010; Asch and Checkley Jr, 2013). Checkley Jr et al. (2000) and Asch and Checkley Jr (2013) found that *S. sagax* eggs were most frequently observed at intermediate salinities of 33.0-33.4 psu, whereas Weber and McClatchie (2010) identified a monotonically decreasing relationship between *S. sagax* larvae and salinity. This contradiction likely reflects the fact that each study considered a different period since the shape of salinity response curves is sensitive to the years used to parameterize SDMs. In contrast, all previous research including ours indicate that *E. mordax* spawn at higher salinities in the southern CCS (Checkley Jr et al., 2000; Weber and McClatchie, 2010; Asch and Checkley Jr, 2013). However, given that this species resides in the Columbia River freshwater plume in the northern CCS (Kaltenberg et al., 2010), phenotypic plasticity or local adaptation might influence *E. mordax* larval occurrence with regard to salinity. Different studies have identified positive and negative relationships between *S. sagax* and zooplankton concentration (Checkley Jr et al., 2000; Lynn, 2003; Agostini et al., 2007). While this might reflect differences in the life stage of *S. sagax* studied, variations in zooplankton species composition, or spurious correlations, non-stationary relationships provide an alternative explanation.

Less research has been conducted on the relationship between SPF and DO in the southern CCS. Koslow et al. (2013) suggested that there was a positive relationship between DO and *S. sagax* larvae, which is consistent with our results across the majority, but not all, regimes. Howard et al. (2020) indicated that the distribution of *E. mordax* is sensitive to DO, especially at high temperatures, which is comparable to our results from recent years, although other patterns are seen early in the CalCOFI time series. These two papers mainly focused mid-water column depths because projected declines in DO concentration under climate change are maximized across this range (Dussin et al., 2019). Our research focused on environmental conditions in the upper 50 m of the water column coincident with the peak vertical distribution of SPF eggs and larvae. Since hypoxic conditions at these depths only occur during extreme upwelling, the reaction of SPF larvae to DO in our study is more representative of the influence of DO as an indicator of water mass characteristics rather than as a physiological stressor.

Less research has been conducted on environmental influences on the species distribution of *S. japonicus* and *T. symmetricus* in the southern CCS. Our results are consistent with prior studies of the influence of temperature and salinity on their spawning distribution (Weber and McClatchie, 2012; Asch and Checkley Jr, 2013). However, this is less so for ZDV. For *S. japonicus*, Weber and McClatchie (2012) found that larvae were most likely to be present at intermediate ZDVs of ~5-7 log cm^3 1,000 m^{-3}. While we observed a similar relationship between ZDV and *S. japonicus* larvae during 1951-1968 (Figure S9), this pattern was not apparent in other periods. Asch and Checkley Jr (2013) identified the highest probability of *T. symmetricus* eggs at low ZDV. The current study identified a similar pattern during 1984-2015, which coincides with years examined by Asch and Checkley Jr (2013). However, differing relationships between *T. symmetricus* distribution and ZDV were observed during earlier periods.

T. symmetricus was the only species to experience a projected increase in IHS under SSP1-2.6 and SSP5-8.5. We hypothesize that this increase in suitable habitat is related to a shift in spawning phenology of *T. symmetricus* under climate change. Future projections were made for March-May since an empirical formula for converting between mesozooplankton carbon biomass from CNRM-ESM2 to ZDV was only available for this season (Supplementary Material 1.2). While these months coincided with the seasonal peak in larval concentration for *E. mordax*, *S. sagax*, and *S. japonicus*, maximum concentrations of *T. symmetricus* are observed in June (Moser et al., 2001). Asch (2015) identified *T. symmetricus* as belonging to a group of fishes whose phenology has become earlier in recent decades in response to warming. The projected future increase in habitat suitability for *T. symmetricus* during March-May likely represents a continuation of this shift towards earlier spawning phenology.

Since fish-environmental relationships change over time, this emphasizes the importance of accurately detecting timing of regime shifts. Our study analyzed change points associated with the 1976/1977 PDO phase change, 1968/1969 and 1983/1984 shifts in ZDV, changes in *S. sagax* and *E. mordax* SSB in 1963/1964, and a second shift in *S. sagax* SSB in 1997/1998. No change points were detected in time series of temperature, salinity, and DO, which may reflect that biological time series often have more non-linear dynamics than physicochemical variables (Hsieh et al., 2005). The change points detected were well supported by other studies of the southern CCS. The 1976/

1977 PDO transition was associated with reduced survival young-of-year *E. mordax* (Nishikawa et al., 2019). The presence of a mid-1960s regime shift was consistent with an analysis of 35 species of CCS ichthyoplankton (Peabody et al., 2018). Other ichthyoplankton studies have identified faunal shifts during 1983/1984 and the late 1990s (Miller and McGowan, 2013; Peabody et al., 2018; Thompson et al., 2019a), which approximately coincide with our change points in ZDV and *S. sagax* SSB, respectively. Unlike previous studies, we did not detect a 1989/1990 regime shift (Miller and McGowan, 2013; Koslow et al., 2015; Peabody et al., 2018). This might reflect that this change point seems to be principally associated with shifts among a few highly abundant taxa in the southern CCS (Peabody et al., 2018). Our Bayesian change point algorithm indicated that there was some uncertainty in the exact year of transitions (Figures 2, 3). This uncertainty may reflect gaps in CalCOFI time series coverage, discontinuities in stock assessments, the decision to log-transform SSB prior to change point detection, and uncertainty related to parameter choice during change point detection (Overland et al., 2008; Peabody et al., 2018). For instance, the choice of minimum regime length affects detection of recent ecological shifts, such as the crash and subsequent recovery of *E. mordax* (Thayer et al., 2017; Thompson et al., 2019b).

4.2 Mechanisms responsible for non-stationary dynamics

Currently there is limited capacity for predicting the occurrence of non-linear ecosystem regime shifts. A meta-analysis of 4,600 global change impacts concluded that such shifts were rarely detectable in advance (Hildebrand et al., 2020). While many regime shifts are characterized by increased time series variance (Lenton, 2011), this signal can be obscured by small variations in organismal responses (Hildebrand et al., 2020). Similarly, Field et al. (2009) concluded that fluctuations in SPF abundance in paleo-ecological time series were characterized by red noise that was not predictable. When combined with novel environmental conditions and changes in how fish react to oceanic variables across regimes, these factors challenge the ability of empirically derived models to make accurate future projections needed for management. However, models that incorporate physiological principles and mechanistic ecological understanding may fare better.

While our study did not directly investigate mechanisms responsibility for non-stationarity, some insights can be attained and may help generate hypotheses for future research. Given the greater amount of literature on *S. sagax* and *E. mordax*, more hypotheses exist to explain non-stationary dynamics among these species. Previous studies suggested that the relationships between these fishes and SST may be a proxy for other environmental factors (e.g., prey availability) that more directly influence population dynamics (Fiedler, 1983; Jacobson and MacCall, 1995). This could lead to non-stationarity if relationships between SST and the direct influences on a species become decoupled. However, this seems unlikely to explain the non-stationarity observed here because the relationship between temperature and larval habitat exhibited greater stationarity than other variables. Previous studies have indicated that DO in the CCS is correlated with variations in nutrient and chlorophyll concentration, water mass characteristics, and geostrophic flow (Weber and McClatchie, 2010; Koslow et al., 2013). Since the relationship between DO and larval presence/absence was subject to greater non-stationarity, changes in the strength of these correlations could be possibly responsible for this non-stationarity.

Changes in modes of climate variability and trophodynamic relationships have also been hypothesized to be mechanisms responsible for non-stationarity in SDMs (Litzow et al., 2019). We observed slightly more non-stationarity across zooplankton regime changes than PDO shifts, suggesting support for trophodynamic changes as an underlying cause of non-stationarity. Related to this point, it must be noted that an environmental variable needs to exceed an organism's tolerance range to affect its distribution. Under modes of climate variability that are favorable to an organism, this tolerance range might not be exceeded. However, values outside of their tolerance may be experienced by fishes during the opposite phase of climate variability or as the climate continues to change. This mechanism could lead to the appearance of non-stationarity when using SDMs parameterized with data from different periods.

Additional mechanisms for explaining non-stationarity are related to migration and dispersal. Since larvae are subject to advection, they do not have complete control over habitats occupied, which could increase the likelihood of non-stationarity (Brun et al., 2016). Conversely, movement by adults can help fishes track favored environmental conditions whereas less migratory species may be unable to follow such conditions (Reglero et al., 2012). This would imply that less migratory species may be subject to greater non-stationarity. However, migratory species may be equipped to face a greater variety of conditions encountered along migration pathways, implying that their distribution may be less tightly coupled with oceanic conditions. *S. sagax* displays greater seasonal migratory behavior than *E. mordax* (Zwolinski et al., 2011) and exhibited a greater incidence of non-stationarity. This suggests the latter idea (i.e., migratory behavior is associated with fewer environmental distribution constraints) has more support based on our data. Our results are also consistent with Planque et al. (2007); Weber and McClatchie (2010), and Muhling et al. (2020) who found that *E. mordax* distribution could be better fit by SDMs than *S. sagax*. *S. sagax* tends to exhibit greater variability in distribution than *E. mordax* at interannual-to-decadal scales, expanding its distribution

offshore and northward when abundant (MacCall, 1990). This expansion, hypothesized to be driven by density-dependent habitat use, may be responsible for greater non-stationarity among *S. sagax*.

Beyond migratory behavior, there are at least two other hypotheses that could explain the high degree of non-stationarity among *S. sagax*. This species is known to undergo demographic changes as its abundance fluctuates. *S. sagax* reaches maturity at age 1 under low biomass and matures at age 2 at high biomass (Hill et al., 2008). Such demographic changes can increase the sensitivity of species to environmental variability (Anderson et al., 2008), which could generate non-stationarity. Another potential explanation could be related to intermixing between the U.S. and southern Baja California stocks of *S. sagax*, which use distinct thermal habitats (Lynn, 2003; Dorval et al., 2011). Nonetheless, the thermal history of habitat occupancy recorded in *S. sagax* otoliths from the southern CCS suggests intermixing of stocks is somewhat rare (Dorval et al., 2011).

4.3 Non-stationarity among oceanic variables

Climate change projections for marine organisms may be improved by focusing on oceanic variables less likely to exhibit non-stationarity. Of the variables considered, temperature most frequently exhibited stable relationships with larvae distribution (Table 2). This reflects that temperature has a direct influence on biological processes as diverse as gene expression, enzyme kinetics, metabolism, consumption, and growth in poikilotherms (Hare et al., 2012). Most marine fishes do not change their mean temperature of occurrence over time (Nye et al., 2009) and track climate velocity by shifting their distribution and depth to reflect changing temperatures (Pinsky et al., 2013). Rates of evolution of thermal niches are projected to be much slower than rates of future environmental change, leading to niche conservatism (Jezkova and Wiens, 2016). Consequently, SDMs driven by thermal preferences may be more reliable for making future projections than those with substantial influences from other variables. Nonetheless, multivariate SDMs generally are better at predicting historical distribution than univariate models (McHenry et al., 2019).

Salinity and ZDV exhibited an intermediate-to-high amount of non-stationarity. Species were often less responsive to these variables during recent regimes as indicated by exclusion of these variables from models, flattened response curves, or decreases in their ranking (e.g., Figures 4, 5). For ZDV, in some cases, fishes were less likely to display a unimodal response curve in recent years. Some non-stationarity observed among these variables may be related to the fact that their response curves had wider confident intervals near the minima and maxima of observed conditions. Due to wide confidence intervals, it was not always possible to determine whether changes in response curves between regimes represented changes in larval occurrence or solely a lack of capacity to precisely quantify responses to infrequently observed states. Brun et al. (2016) obtained similar results where SDMs displayed decreased skill near the edges of a species range where conditions were more extreme. It is important to understand how species react to such extremes since they are projected to occur more frequent under climate change (Frölicher et al., 2018). Laboratory experiments may be useful since they allow for replication of extremes observed infrequently in nature.

DO often exhibited a greater influence on SPF during recent regimes (Figures 5, 6, S2). Under climate change, DO in the CCS is projected to decline due to reduced solubility of oxygen in warmer water, increased stratification, changes in deep-water circulation causing reduced ventilation, and changes in upwelling strength (Rykaczewski and Dunne, 2010; Dussin et al., 2019). These changes have been documented to influence the historical abundance of mesopelagic fishes in the southern CCS (Koslow et al., 2011) and are projected to affect the future persistence of *E. mordax* in the region (Howard et al., 2020). Our findings are consistent with these patterns.

4.4 Projection uncertainty

For climate change impacts to be considered in fisheries management, uncertainty in future projections must be quantified. This is because managers will need to contemplate both best- and worst-case scenarios in the planning process (Cheung et al., 2016a). In ecological models, uncertainty can result from incomplete observational records, different approaches to conceptual and numerical model formulation, parameter estimation, model selection, choice of spatiotemporal scale, and adaptability of living systems (Planque et al., 2011). Future research should consider non-stationarity in fish-environmental relationships as another source of model uncertainty. Here we showed that the period used to parameterize SDMs can have a substantial impact on future projections due to non-stationarity, with the magnitude of this effect sometimes exceeding the effect of different climate scenarios. One understudied area with respect to climate change uncertainty is whether there might be interactions between different sources of uncertainty. We found that an interaction exists between uncertainty due to non-stationarity and SSP scenario, with an increasing effect of non-stationarity at higher emissions.

As with most SDMs, there are a number of qualifications that may affect our results. To take advantage of the multi-decadal CalCOFI time series, our analysis focused on the southern CCS, which does not encompass the full range of target species. Nonetheless, given the pronounced onshore-offshore gradients sampled by CalCOFI, this dataset covers

several oceanic water masses exhibiting different conditions (McClatchie, 2013). Also, previous research has used CalCOFI to understand how environmental change affects fish distribution despite the dataset's limited spatial extent (Hsieh et al., 2008; Hsieh et al., 2009; Howard et al., 2020; Muhling et al., 2020). A second qualification is that some of the changes in how fishes respond to the environment could be related to interactions between multiple variables influencing fish distribution. Changes in response curve shape may reflect the fact that partial responses from GAMs depend on the partial response of a species to other variables. For example, the extent to which DO is a stressor depends on temperature (Howard et al., 2020). GAMs often do not account for such interactions, but other SDMs do. We evaluated non-stationarity across periods with change points in *S. sagax* SSB using a second model that accounts for such interactions (the non-parametric probabilistic ecological niche model; Beaugrand et al., 2011; R.G. Asch unpublished data). Since non-stationarity was also common when using this alternative SDM, the high incidence of non-stationarity in the GAMs cannot be explained solely by multivariate interactions. Our models purposely did not include SSB as an independent variable because it is unlikely that future SSB would be precisely known when projecting climate change impacts. However, SSB can influence *S. sagax* and *S. japonicus* larval distribution (Weber and McClatchie, 2010; Weber and McClatchie, 2012). Models may display fewer incidences of non-stationarity due to density dependence if different SSB scenarios are included in long-range projections. Another critique of SDMs is that they do not typically allow for acclimation or adaptation to changing conditions. However, it is also unclear how important these processes are for fishes since thermal niches evolve slowly (Jezkova and Wiens, 2016). Also, fishes may migrate towards preferred conditions prior to acclimation (Habary et al., 2016).

4.5 Recommendations for improving SDM projections for marine fishes

Moving forward, it is important to determine if the high incidence of non-stationarity detected here is widespread or mainly a characteristic among SPF larvae in upwelling systems. For populations likely subject to non-stationary environmental relationships, we recommend validating SDMs with independent datasets whenever possible. Cross-validation with a subset of the original dataset can result in potential overestimation of model skill due to temporal and spatial autocorrelation or overfitting (Araújo et al., 2005; Planque et al., 2011). Some measures of model skill, such as the true skill statistic, perform similarly regardless of the time lag between datasets used for model development and testing (Brun et al., 2016). Wider use of the true skill statistic could help realistically assess model skill when an independent dataset is unavailable for validation. Since variables exhibiting indications of non-stationarity were more likely to have SDM response curves with wide confidence intervals, we recommend that response curve confidence intervals be more frequently reported. Nonetheless, some climate-envelope models may underestimate confidence intervals associated with the centroid of species distribution (Thorson, 2018).

Another suggestion for guarding against non-stationarity and improving confidence in SDM projections is to compare model-derived environmental niches against those from physiological experiments (Asch and Erisman, 2018; Muhling et al., 2020). Alternatively, physiologically based thermal tolerances can be used to parameterize SDMs (Hare et al., 2012). However, it is not unusual to see discrepancies between laboratory-derived and field-based estimates of thermal niche due to differences between fundamental and realized niches (Henderson, 2019). Related to this, fishes may not fully occupy suitable habitat within their realized niche during low abundance (Planque et al., 2007), which can lead to non-stationary relationships. Using thresholds GAMs where a threshold is prescribed based on fish biomass is a common way to mitigate against such dynamics (Lindegren and Eero, 2013; Beggs et al., 2014; van der Sleen et al., 2018).

Obtaining reliable projections of fish species distribution, phenology, and population dynamics is important, because it allows fisheries managers to better engage in adaptive management. Networks of marine protected areas and the timing of seasonal fishing closures may need adjustment as fishes undergo range shifts or phenological changes (McLeod et al., 2009; Peer and Miller, 2014). Fisheries independent surveys can be made more efficient when relationships between fish distribution and the environment are used to adaptively adjust sampling (Zwolinski et al., 2011). Most stock assessments assume population processes affecting fisheries are stationary, which can create retrospective bias in estimates of population parameters if there has been a change in fishery productivity (Szuwalski and Hollowed, 2016). Stock assessments may be improved by incorporating environmentally variable recruitment, growth, mortality, or catchability into assessments (Adams et al., 2015; Pershing et al., 2015; Tommasi et al., 2017). If the productivity of stocks changes as a function of climate, it may be necessary to adjust acceptable biological catch to meet management objectives (Vert-pre et al., 2013). Alternative approaches to dealing with non-stationarity when setting management targets include adopting targets that harvest a constant fraction of the stock and only considering the most recent regime when parameterizing stock assessments (Vert-pre et al., 2013; Szuwalski and Hollowed, 2016). Management strategy evaluation also relies on robust assessments of climate change impacts on fishes when assessing which strategies produce resilient fisheries (Szuwalski and Hollowed, 2016). Non-stationary relationships that create greater uncertainty in future projections may reduce the reliability of these

management strategies for adapting to change. However, this challenge only further underscores the importance of adaptive management to account for the non-stationary reactions of fishes.

In conclusion, we determined that non-stationary relationships between larval occurrence and environmental variables were nearly ubiquitous in the CCS, occurring across multiple types of indicators, regime shifts, oceanic variables, and species. This has implications for the robustness of future projections of species distribution changes since most projections rely on statistical models that assume stationary relationships. Differences between alternative projections became amplified under climate change, suggesting this source of uncertainty may become increasingly important in the future. Nonetheless, the relationship between temperature and larval occurrence was more stable than other variables, likely due to effects of temperature on fish physiology. Non-stationarity was especially pronounced when examining regime shifts defined by biological changes, such as shifts in SSB and ZDV. This suggests that density dependence and prey availability may play key roles modulating how fishes react to oceanic conditions.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: NOAA ERDDAP server: https://coastwatch.pfeg.noaa.gov/erddap/index.html; CMIP6: https://esgf-node.llnl.gov/projects/cmip6/.

Ethics statement

Ethical review and approval was not required for animal use in this study because this manuscript solely uses historical data that were not gathered by the authors and were archived in online databases.

Author contributions

RGA designed the research. RGA, JS, and KC performed the research and analyzed the data. RGA wrote the first draft of the manuscript. All authors contributed to the manuscript revision, read, and approved the final version.

Funding

RGA was supported by Nippon Foundation-Nereus Program, Alfred P. Sloan Foundation Research Fellowship Program, and NSF OCE award number 2049624. JS and KC received support from the High Meadows Environmental Institute.

Acknowledgments

We thank the CalCOFI program and CMIP6 for making available observational data and earth system model output. We would also like to thank the two reviewers and associate editor whose suggestions helped to improve this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2022.711522/full#supplementary-material

References

Adams, C. F., Miller, T. J., Manderson, J. P., Richardson, D. E., and Smith, B. E. (2015). *Butterfish 2014 stock assessment* (Woods Hole, MA: National Marine Fisheries Service, Northeast Fisheries Science Center).

Agostini, V. N., Bakun, A., and Francis, R. C. (2007). Larval stage controls on Pacific sardine recruitment variability: High zooplankton abundance linked to poor reproductive success. *Mar. Ecol. Prog. Ser.* 345, 237–244. doi: 10.3354/meps06992

Anderson, J. J., Gurarie, E., Bracis, C., Burke, B. J., and Laidre, K. L. (2013). Modeling climate change impacts on phenology and population dynamics of migratory marine species. *Ecol. Model.* 264, 83–97. doi: 10.106/j.ecolmodel.2013.03.009

Anderson, C. N. K., Hsieh, C. H., Sandin, S. A., Hewitt, R., Hollowed, A., Beddington, J., et al. (2008). Why fishing magnifies fluctuations in fish abundance. *Nature* 452, 835–839. doi: 10.1038/nature06851

Araújo, M. B., Pearson, R. G., Thuiller, W., and Erhard, M. (2005). Validation of species-climate impact models under climate change. *Glob. Change Biol.* 11, 1504–1513. doi: 10.1111/j.1365-2486.2005.01000.x

Asch, R. G. (2015). Climate change and decadal shifts in the phenology of larval fishes in the California Current Ecosystem. *Proc. Nat. Acad. Sci. U.S.A.* 112 (30), E4065–E4074. doi: 10.1073/pnas.1421946112

Asch, R. G., and Checkley, ,. D. M.Jr. (2013). Dynamic height: A key variable for identifying the spawning habitat of small pelagic fishes. *Deep-Sea Res. I* 71, 79–91. doi: 10.1016/j.dsr.2012.08.006

Asch, R. G., and Erisman, B. (2018). Spawning aggregations act as a bottleneck influencing climate change impacts on a critically endangered reef fish. *Divers. Distrib.* 24 (12), 1712–1728. doi: 10.1111/ddi.12809

Asch, R. G., Stock, C. A., and Sarmiento, J. L. (2019). Climate change impacts on mismatches between phytoplankton blooms and fish spawning phenology. *Glob. Change Biol.* 25, 2544–2559. doi: 10.1111/gcb.14650

Barange, M., Coetzee, J., Takasuka, A., Hill, K., Gutierrez, M., Oozeki, Y., et al. (2009). Habitat expansion and contraction in anchovy and sardine populations. *Prog. Oceanogr.* 83, 251–260. doi: 10.1016/j.pocean.2009.07.027

Beaugrand, G., Lenoir, S., Ibañez, F., and Manté, C. (2011). A new model to assess the probability of occurrence of a species, based on presence-only data. *Mar. Ecol. Prog. Ser.* 424, 175–190. doi: 10.3354/meps08939

Beggs, S. E., Cardinale, M., Gowen, R. J., and Bartolino, V. (2014). Linking cod (*Gadus morhua*) and climate: investigating variability in Irish Sea cod recruitment. *Fish. Oceanogr.* 23, 54–64. doi: 10.1111/fog.12043

Bell, R. J., Richardson, D. E., Hare, J. A., Lynch, P. D., and Fratantoni, (2015). Disentangling the effects of climate, abundance, and size on the distribution of marine fish: an example based on four stocks from the northeast US shelf. *ICES J. Mar. Sci.* 72, 1311–1322. doi: 10.1093/icesjms/fsu217

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., et al. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science* 366, 339–345. doi: 10.1126/science.aaw1620

Brun, P., Kiørboe, T., Licandro, P., and Payne, M. R. (2016). The predictive skill of species distribution models for plankton in a changing climate. *Global Change Biol.* 22, 3170–3181. doi: 10.1111/gcb.13274

Burnham, K. P., and Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (New York, NY: Springer).

Chan, F., Barth, J. A., Kirincich, A., Weeks, H., Peterson, W. T., and Menge, B. A. (2008). Emergence of anoxia in the California Current large marine ecosystem. *Science* 319, 920. doi: 10.1126/science.1149016

Chavez, F. P., Ryan, J., Lluch-Cota, S. E., and Ñiquen, C. M. (2003). From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science* 299, 217–221. doi: 10.1126/science.1075880

Checkley, D. M., Asch, R. G., and Rykaczewski, R. R. (2017). Climate, anchovy, and sardine. *Ann. Rev. Mar. Sci.* 9, 469–493. doi: 10.1146/annurev-marine-122414-033819

Checkley, D. M.Jr., Dotson, R. C., and Griffin, D. A. (2000). Continuous, underway sampling of eggs of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) in spring 1996 and 1997 off southern and central California. *Deep-Sea Res. II* 47, 1139–1155. doi: 10.1016/S0967-0645(99)00139-3

Chen, L. C., Hill, J. K., Ohlemüller, R. D. B., and Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science* 333, 1024–1026. doi: 10.1126/science.1206432

Cheung, W. W. L., Frölicher, T. L., Asch, R. G., Jones, M. G., Pinsky, M. L., Reygondeau, G., et al. (2016a). Building confidence in projections of the responses of living marine resources to climate change. *ICES J. Mar. Sci.* 73, 1283–1296. doi: 10.1093/icesjms/fsv250

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. doi: 10.1111/j.1467-2979.2008.00315.x

Cheung, W. W. L., Reygondeau, G., and Frölicher, F. L. (2016b). Large benefits to marine fisheries of meeting the 1.5°C global warming target. *Science* 354, 1591–1594. doi: 10.1126/science.aag2331

Ciannelli, L., Bailey, K. M., Chan, K. S., and Stenset, N. C. (2007). Phenological and geographical patterns of walleye pollock (*Theragra chalcogramma*) spawning in the western Gulf of Alaska. *Can. J. Fish. Aquat. Sci.* 64, 713–722. doi: 10.1139/f07-049

Collie, J. S., Richardson, K., and Steele, J. H. (2004). Regime shifts: can ecological theory illuminate the mechanisms? *Prog. Oceanogr.* 60, 281–302. doi: 10.1016/j.pocean.2004.02.013

Crone, P. R., and Hill, K. T. (2015). *Pacific mackerel (Scomber japonicus) stock assessment for USA management in the 2015-2016 fishing year* (Portland, OR: Pacific Fishery Management Council).

Curtis, K. A., Checkley, D. M., and Pepin, P. (2007). Predicting the vertical profiles of anchovy (*Engraulis mordax*) and sardine (*Sardinops sagax*) eggs in the California Current System. *Fish. Oceanogr.* 16, 68–84. doi: 10.1111/j.1365-2419.2006.00414.x

Cury, P. M., Boyd, I. L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R. J., Furness, R. W., et al. (2011). Global seabird response to forage fish depletion–one-third for the birds. *Science* 334, 1703–1706. doi: 10.1126/science.1212928

Di Lorenzo, E., and Mantua, N. (2016). Multi-year persistence of the 2014-2015 North Pacific marine heatwave. *Nat. Clim. Change* 6, 1042–1048. doi: 10.1038/nclimate3082

Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P. J. S., Chhak, K., Miller, A. J., et al. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.* 35, L08607. doi: 10.1029/2007GL032838

Donelson, J. M., Munday, P. M., McCormick, M. I., and Pitcher, C. R. (2012). Rapid transgenerational acclimation of a tropical reef fish to climate change. *Nat. Clim. Change* 2, 30–32. doi: 10.1038/nclimate1323

Dorval, E., Piner, K., Robertson, L., Reiss, C. S., Javor, B., and Vetter, R. (2011). Temperature record in the oxygen stable isotopes of Pacific sardine otoliths: Experimental vs. wild stocks from the Southern California Bight. *J. Exp. Mar. Biol. Ecol.* 397, 136–143. doi: 10.1016/j.jembe.2010.11.024

Dussin, R., Curchitser, E. N., Stock, C. A., and van Oostende, N. (2019). Biogeochemical drivers of changing hypoxia in the California Current Ecosystem. *Deep-Sea Res. II*, 169–170, 105490. doi: 10.1016/j.dsr2.2019.05.013

Efron, B., and Tibshirani, R. J. (1998). *An introduction to the bootstrap* (Boca Raton, F: Chapman & Hall/CRC Press).

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. doi: 10.1146/annrev.ecolsys.110308.120159

Fernandes, J. A., Cheung, W. W. L., Jennings, S., Butenschön, M., de Mora, L., Frölicher, T. L., et al. (2013). Modelling the effects of climate change on the distribution and production of marine fishes: accounting for trophic interactions in a dynamic bioclimate envelope model. *Glob. Change Biol.* 19, 2596–2607. doi: 10.1111/gcb.12231

Fernandes, J. A., Rutterford, L., Simpson, S. D., Buttenschon, M., Frölicher, T. L., Yool, A., et al. (2020). Can we predict changes in fish abundance in response to decadal-scale climate scenarios? *Glob. Change Biol.* 26, 3891–3905. doi: 10.1111/gcb.15081

Fiedler, P. C. (1983). Satellite remote sensing of the habitat of spawning anchovy in the Southern California Bight. *Cal. Coop. Ocean Fish. Invest. Rep.* 24, 202–209.

Field, D. B., Baumgartner, T. R., Ferreira, V., Gutierrez, D., Lozano-Montes, H., Salvatteci, R., et al. (2009). "Variability in small pelagic fishes from scales in marine sediments and other historical records," in *Climate change and small pelagic fish*. Eds. D. M. Checkley, C. Roy, J. Alheit and Y. Oozeki (Cambridge, UK: Cambridge University Press), 45–63.

Free, C. M., Thorson, J. T., Pinsky, M. L., Oken, K. L., Wiedenmann, J., and Jensen, O. P. (2019). Impacts of historical warming on marine fisheries production. *Science* 363, 979–983. doi: 10.1126/science.aau1758

Frölicher, T. L., Fischer, E. M., and Gruber, N. (2018). Marine heatwaves under global warming. *Nature* 560, 360–366. doi: 10.1038/s41586-018-0383-9

Golden, C. D., Allison, E. H., Cheung, W. W. L., Dey, M. M., Halpern, B. S., McCauley, D. J., et al. (2016). Fall in fish catch threatens human health. *Nature* 534, 317–320. doi: 10.1038/534317a

Habary, A., Johansen, J., Nay, T. J., Steffensen, J. F., and Rummer, J. L. (2016). Adapt, move or die – how will tropical coral reef fishes cope with ocean warming? *Glob. Change Biol.* 23, 566–577. doi: 10.1111/gcb.13488

Hare, S. R., Mantua, N. J., and Francis, R. C. (1999). Inverse production regimes: Alaska and West Coast Pacific salmon. *Fisheries* 24, 6–14. doi: 10.1577/1548-8446(1999)024<0006:IPR>2.0.CO;2

Hare, J. A., Wuenschel, M. J., and Kimball, M. E. (2012). Projecting range limits with coupled thermal tolerance – climate change models: an example based on gray snapper (*Lutjanus griseus*) along the U.S East Coast. *PloS One* 7 (12), e52294. doi: 10.1371/journal.pone.0052294

Hastie, T. J. (1991)."Generalized additive models," in *Statistical Models in S*. Eds. J. M. Chambers and T. J. Hastie (Boca Raton, FL: Chapman and Hall/CRC), 249–308.

Henderson, M. E. (2019). *Direct and indirect effects of temperature on marine fish distributions along the Northeast United States continental shelf* (Dissertation. Stony Brook (NY: Stony Brook University).

Hildebrand, H., Donohue, I., Harpole, W. S., Hodapp, D., Kucera, M., Lewandowska, A. M., et al. (2020). Thresholds for ecological responses to global change do not emerge from empirical data. *Nat. Ecol. Evol.* 4, 1502–1509. doi: 10.1038/s41559-020-1256-9

Hill, K. T., Crone, P. R., and Zwolinski, J. P. (2018). *Assessment of the Pacific sardine resource in 2018 for U.S. management in 2018-2019. NOAA technical memorandum NMFS-SWFSC-600* (La Jolla, CA: US Department of Commerce).

Hill, K. T., Dorval, E., Lo, N. C. H., Macewicz, B. J., Show, C., and Felix-Uraga, R. (2008). "Assessment of the Pacific sardine resource in 2008 for U.S. management in 2009," (La Jolla, CA: NOAA National Marine Fisheries Service, Southwest Fisheries Science Center).

Howard, E. M., Penn, J. L., Frenzel, H., Seibel, B. A., Bianchi, D., Renault, L., et al. (2020). Climate-driven aerobic habitat loss in the California Current System. *Sci. Adv.* 6, eaay3188. doi: 10.1126/sciadv.aay3188

Hsieh, C. H., Glaser, S. M., Lucas, A. J., and Sugihara, G. (2005). Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* 435, 336–340. doi: 10.1038/nature03553

Hsieh, C. H., Kim, H. J., Watson, W., Di Lorenzo, E., and Sugihara, G. (2009). Climate-driven changes in abundance and distribution of larvae of oceanic fishes in the Southern California region. *Glob. Change Biol.* 15, 2137–2152. doi: 10.1111/j.1365-2486.2009.01875.x

Hsieh, C. H., Reiss, C. S., Hewitt, R. P., and Sugihara, G. (2008). Spatial analysis shows that fishing enhances the climatic sensitivity of marine fishes. *Can. J. Fish. Aquat. Sci.* 65, 947–961. doi: 10.1139/f08-017

Jacobson, L. D., and MacCall, A. D. (1995). Stock-recruitment models for Pacific sardine (*Sardinops sagax*). *Can. J. Fish. Aquat. Sci.* 52, 566–577. doi: 10.1139/f95-057

Jezkova, T., and Wiens, J. J. (2016). Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. *Proc. R. Soc B* 282, 20162104. doi: 10.1098/rspb.2016.2104

Joyce, T. (2002). One hundred plus years of wintertime climate variability in the Eastern United States. *J. Clim.* 15, 1076–1086. doi: 10.1175/1520-0442(2002)015<1076:OHPYOW>2.0.CO;2

Kaltenberg, A. M., Emmett, R. L., and Benoit-Bird, K. J. (2010). Timing of forage fish seasonal appearance in the Columbia river plume and link to ocean conditions. *Mar. Ecol. Prog. Ser.* 419, 171–184. doi: 10.3354/meps08848

Kaplan, I. C., Koehn, L. E., Hodgson, E. E., Marshall, K. N., and Essington, T. E. (2017). Modeling food web effects of low sardine and anchovy abundance in the California Current. *Ecol. Model.* 359, 1–24. doi: 10.1016/j.ecolmodel.2017.05.007

Koslow, J. A., Goericke, R., Lara-Lopez, A., and Watson, W. (2011). Impact of declining intermediate-water oxygen on deepwater fishes in the California Current. *Mar. Ecol. Prog. Ser.* 436, 207–218. doi: 10.3354/meps09270

Koslow, J. A., Goericke, R., and Watson, W. (2013). Fish assemblages in the Southern California Current: relationships with climate 1951-2008. *Fish. Oceanogr.* 22, 207–219. doi: 10.1111/fog.12018

Koslow, J. A., Miller, E. F., and McGowan, J. A. (2015). Dramatic declines in coastal and oceanic fish communities off California. *Mar. Ecol. Prog. Ser.* 538, 221–227. doi: 10.3354/meps.11444

Kramer, D., Kalin, M. J., Stevens, E. G., Thrailkill, J. R., and Zweifel, J. R. (1972). *Collecting and Processing Data on Fish Eggs and Larvae in the California Current Region. NOAA Technical Report NMFS CIRC-370*. Seattle: National Marine Fisheries Service.

Lavaniegos, B. E., and Ohman, M. D. (2007). Coherence of long-term variations of zooplankton in two sectors of the California Current System. *Prog. Oceanogr.* 75, 42–69. doi: 10.1016/j.pocean.2007.07.002

Lenton, L. M. (2011). Early warning of climate tipping points. *Nat. Clim. Change* 1, 201–209. doi: 10.1038/nclimate1143

Lindegren, M., and Checkley, D. M. (2012). Temperature dependence of Pacific sardine (*Sardinops sagax*) recruitment in the California Current Ecosystem revisited and revised. *Can. J. Fish. Aquat. Sci.* 70, 245–252. doi: 10.1139/cjfas-2012-0211

Lindegren, M., and Eero, M. (2013). Threshold-dependent climate effects and high mortality limit recruitment and recovery of the Kattegat cod. *Mar. Ecol. Prog. Ser.* 490, 223–232. doi: 10.3354/meps10437

Litzow, M. A., Ciannelli, L., Puerta, P., Wettstein, J. J., Rykaczewski, R. R., and Opiekun, M. (2018). Non-stationary climate-salmon relationships in the Gulf of Alaska. *Proc. R. Soc B* 285, 20181855. doi: 10.1098/rspb.2018.1855

Litzow, M. A., Ciannelli, L., Puerta, P., Wettstein, J. J., Rykaczewski, R. R., and Opiekun, M. (2019). Nonstationary environmental and community relationships in the North Pacific Ocean. *Ecology* 100 (8), e02760. doi: 10.1002/ecy.2760

Litzow, M. A., Hunsicker, M. E., Bond, N. A., Burke, B. J., Cunningham, C. J., Gosselin, J. L., et al. (2020). The changing physical and ecological meanings of North Pacific Ocean climate indices. *Proc. Nat. Acad. Sci. U.S.A.* 117, 7665–7671. doi: 10.1073/pnas.1921266117

Lluch-Belda, D., Lluch-Cota, D. B., Hernandez-Vazquez, S., Salinas-Zavala, C. A., and Schwartlose, R. A. (1991). Sardine and anchovy spawning as related to temperature and upwelling in the California Current System. *Cal. Coop. Ocean Fish. Invest. Rep.* 32, 105–111.

Lynn, R. (2003). Variability in the spawning habitat of Pacific sardine (*Sardinops sagax*) off southern and central California. *Fish. Oceanogr.* 12, 554–568. doi: 10.1046/j.1365-2419.2003.00232.x

MacCall, A. D. (1990). *Dynamic geography of marine fish populations* (Seattle: University of Washington Press).

Mair, L., Thomas, C. D., Anderson, B. J., Fox, R., Botham, M., and Hill, J. K. (2012). Temporal variation in responses of species to four decades of climate warming. *Glob. Change Biol.* 18, 2439–2447. doi: 10.1111/j.1365-2486.2012.02730.x

McClatchie, S. (2013). *Regional fisheries oceanography of the California Current System: The CalCOFI program* (New York: Springer).

McClatchie, S., Goericke, R., Auad, G., and Hill, K. (2010). Re-assessment of the stock recruit and temperature-recruit relationships for Pacific sardine (*Sardinops sagax*). *Can. J. Fish. Aquat. Sci.* 67, 1782–1790. doi: 10.1139/F10-101

McClatchie, S., Hendy, I. L., Thompson, A. R., and Watson, W. (2017). Collapse and recovery of forage fish populations prior to commercial exploitation. *Geophys. Res. Lett.* 44, 1877–1885. doi: 10.1002/2016GL071751

McGowan, J. A., Bograd, S. J., Lynn, R. J., and Miller, A. J. (2003). The biological response to the 1977 regime shift in the California Current. *Deep-Sea Res. II* 50, 2567–2582. doi: 10.1016/S0967-0645(03)00135-8

McHenry, J., Welch, H., Lester, S. E., and Saba, V. (2019). Projecting marine species range shifts from only temperature can mask climate vulnerability. *Glob. Change Biol.* 25, 4208–4221. doi: 10.1111/gcb.14828

McLeod, E., Salm, R., Green, A., and Almany, J. (2009). Designing marine protected area networks to address the impacts of climate change. *Front. Ecol. Environ.* 7, 362–370. doi: 10.1890/070211

Mhlongo, N., Yemane, D., Hendricks, M., and van der Lingen, C. D. (2015). Have the spawning habitat preferences of anchovy (*Engraulis encrasicolus*) and sardine (*Sardinops sagax*) in the southern Benguela changed in recent years? *Fish. Oceanogr.* 24, 1–14. doi: 10.1111/fog.12061

Miller, E. F., and McGowan, J. A. (2013). Faunal shift in Southern California's coastal fishes: A new assemblage and trophic structure takes hold. *Estuar. Coast. Shelf Sci.* 127, 29–36. doi: 10.1016/j.ecss.2013.04.014

Morley, J. W., Selden, R. L., Latour, R. J., Frölicher, T. L., Seagraves, R. J., and Pinsky, M. L. (2018). Projecting shifts in thermal habitat for 686 species on the North American continental shelf. *PloS One* 13 (5), e0196127. doi: 10.1371/journal.pone.0196127

Moser, H. G., Charter, R. L., Smith, P. E., Ambrose, D. A., Watson, W., Charter, S. R., et al. (2001). *Distributional atlas of fish larvae and eggs in the Southern California Bight region: 1951-1998. Atlas 34* (La Jolla, CA: National Marine Fisheries Service, Southeast Fisheries Science Center).

Muhling, B. A., Brodie, S., Smith, J. A., Tommasi, D., Gaitan, C. F., Hazen, E. L., et al. (2020). Predictability of species distributions deteriorates under novel environmental conditions in the California Current System. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00589

Murawski, S. A. (1993). Climate change and marine fish distributions: forecasting from historical analogy. *Trans. Am. Fish. Soc* 122, 647–658. doi: 10.1577/1548-8659(1993)122<0647:CCAMFD>2.3.CO;2

Nishikawa, H., Curchitser, E. N., Fiechter, J., Rose, K. A., and Hedstrom, K. (2019). Using a climate-to-fishery model to simulate the influence of the 1976-1977 regime shift on anchovy and sardine in the California Current System. *Prog. Earth Planet. Sci.* 6, 9. doi: 10.1186/s40645-019-0257-2

Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Mar. Ecol. Prog. Ser.* 393, 111–129. doi: 10.3354/meps08220

Ohman, M. D., and Smith, P. E. (1995). A comparison of zooplankton sampling methods in the CalCOFI time series. *Calif. Coop. Ocean. Fish. Invest. Rep.* 36, 153–158.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci. Model. Dev.* 9, 3461–3482. doi: 10.5194/gmd-9-3461-2106

Overland, J., Rodionov, S., Minobe, S., and Bond, N. (2008). North Pacific regime shifts: Definitions, issues and recent transitions. *Prog. Oceanogr.* 77, 92–102. doi: 10.1016/j.pocean.2008.03.016

Peabody, C. E., Thompson, A. R., Sax, D. F., Morse, R. E., and Perretti, C. T. (2018). Decadal regime shifts in Southern California's ichthyoplankton assemblage. *Mar. Ecol. Prog. Ser.* 607, 71–83. doi: 10.3354/meps12787

Peer, A. C., and Miller, T. J. (2014). Climate change, migration phenology, and fisheries management interact with unanticipated consequences. *N. Am. J. Fish. Manage.* 34, 94–110. doi: 10.1080/02755947.2013.847877

Perry, A. L., Low, P. J., Ellis, J. R., and Reynolds, J. D. (2005). Climate change and distribution in marine fishes. *Science* 308, 1912–1915. doi: 10.1126/science.1111322

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., et al. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science* 350, 809–812. doi: 10.1126/science.aac9819

Peterson, W. (2009). Copepod species richness as an indicator of long-term changes in the coastal ecosystem of the northern California Current. *Calif. Coop. Ocean Fish. Invest. Rep.* 50, 73–81.

PFMC (Pacific Fishery Management Council) (2019) *Coastal pelagic species fishery management plan* (Portland, OR: Pacific Fishery Management Council). Available at: https://www.pcouncil.org/documents/2019/06/cps-fmp-as-amended-through-amendment-17.pdf/ (Accessed September 26, 2021).

Pikitch, E. K., Rountos, K. J., Essington, T. E., Santora, C., Pauly, D., Watson, R., et al. (2014). The global contribution of forage fish to marine fisheries and ecosystems. *Fish Fish.* 15, 43–64. doi: 10.1111/faf.12004

Pinsky, M. L., and Byler, D. (2015). Fishing, fast growth and climate variability increase the risk of collapse. *Proc. R. Soc B.* 282, 20150153. doi: 10.1098/rspb.2015.1053

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L., and Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* 569, 108–111. doi: 10.1038/s41586-019-1132-4

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science* 341, 1239–1242. doi: 10.1126/science.1239352

Planque, B., Bellier, E., and Lazure, P. (2007). Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. *Fish. Oceanogr.* 16, 16–30. doi: 10.1111/j.1365-2419.2006.00411.x

Planque, B., Bellier, E., and Loots, C. (2011). Uncertainties in projecting spatial distributions of marine populations. *ICES J. Mar. Sci.* 68, 1045–1050. doi: 10.1093/icesjms/fsr007

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., et al. (2013). Global imprint of climate change on marine life. *Nat. Clim. Change* 3, 919–925. doi: 10.1038/nclimate1958

Puerta, P., Ciannelli, L., Rykaczewski, R. R., Opiekun, M., and Litzow, M. A. (2019). Do Gulf of Alaska fish and crustacean poplations show synchronous non-stationary responses to climate? *Prog. Oceanogr.* 175, 161–170. doi: 10.1016/j.pocean.2019.04.002

Reglero, P., Ciannelli, L., Alvarez-Berastegui, D., Balbín, R., López-Jurado, J. L., and Alemany, F. (2012). Geographically and environmentally driven spawning distributions of tuna species in the western Mediterranean Sea. *Mar. Ecol. Prog. Ser.* 463, 273–284. doi: 10.3354/meps09800

Reiss, C. S., Checkley, D. M., and Bograd, S. J. (2008). Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) within the California Current. *Fish. Oceanogr.* 17, 126–136. doi: 10.1111/j.1365-2419.2008.00469.x

Roberts, S. M., Boustany, A. M., Halpin, P. N., and Rykaczewski, R. R. (2019). Cyclical climate oscillation alters species statistical relationships with local habitat. *Mar. Ecol. Prog. Ser.* 614, 159–171. doi: 10.3354/meps12890

Roemmich, D., and McGowan, J. (1995). Climate warming and the decline of zooplankton in the California Current. *Science* 267, 1324–1326. doi: 10.1126/science.267.5202.1324

Ruggieri, E. (2013). A Bayesian approach to detecting change points in climatic records. *Int. J. Climatol.* 33, 520–528. doi: 10.1002/joc.3447

Rykaczewski, R. R., and Checkley, D. M.Jr. (2008). Influence of ocean winds on the pelagic ecosystem in upwelling regions. *Proc. Nat. Acad. Sci. U.S.A.* 105, 1965–1970. doi: 10.1073/pnas.071777105

Rykaczewski, R. R., and Dunne, J. P. (2010). Enhanced nutrient supply to the California Current Ecosystem with global warming and increased stratification in an earth system model. *Geophys. Res. Lett.* 37, L21606. doi: 10.1029/2010GL045019

Schwartzlose, R. A., Alheit, J., Bakun, A., Baumgartner, T. R., Cloete, R., Crawford, ,. R. J. M., et al. (1999). Worldwide large-scale fluctuations of sardine and anchovy populations. *S. Afr. J. Mar. Sci.* 21, 289–347. doi: 10.2989/025776199784125962

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM earth system model, CNRM-ESM2-1: Role of earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.* 11. doi: 10.1029/2019MS001791

Selden, R. L., Batt, R. D., Saba, V. S., and Pinsky, M. L. (2018). Diversity in thermal affinity among key piscivores buffers impacts of ocean warming on predator-prey interactions. *Glob. Change Biol.* 24, 117–131. doi: 10.1111/gcb.13838

Smith, A. D. M., Brown, C. J., Bulman, C. M., Fulton, E. Z., Johnson, P., Kaplan, I. C., et al. (2011). Impacts of fishing low-trophic level species in marine ecosystems. *Science* 333, 1147–1150. doi: 10.1126/science.1209395

Stock, C. A., Alexander, M. A., Bond, N. A., Brander, K. M., Cheung, W. W. L., Curchitser, E. N., et al. (2011). On the use of IPCC-class models to assess the impact of climate on living marine resources. *Prog. Oceanogr.* 88, 1–27. doi: 10.1016/j.pocean.2010.09.001

Sunday, J. M., Bates, A. E., and Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nat. Clim. Change* 2, 686–690. doi: 10.1038/nclimate1539

Szuwalski, C. S., and Hollowed, A. B. (2016). Climate change and non-stationary population processes in fisheries management. *ICES J. Mar. Sci.* 73, 1297–1305. doi: 10.1093/icesjms/fsv229

Takasuka, A., Oozeki, Y., Kubota, H., and Lluch-Cota, S. E. (2008). Contrasting spawning temperature optima: Why are anchovy and sardine regime shifts synchronous across the North Pacific? *Prog. Oceanogr.* 77, 225–232. doi: 10.1016/j.pocean.2008.03.008

Thayer, J. A., MacCall, A. D., Sydeman, W. J., and Davison, D. (2017). California anchovy population remains low 2012-2016. *Cal. Coop. Ocean. Fish. Invest. Rep.* 58, 69–76.

Thompson, A. R., Harvey, C. J., Sydeman, W. J., Barceló, C., Bograd, S. J., Brodeur, R. D., et al. (2019a). Indicators of pelagic forage community shifts in the California Current Large Marine Ecosystem 1998-2016. *Ecol. Indic.* 105, 215–228. doi: 10.1016/j.ecolind.2019.05.057

Thompson, A. R., McClatchie, S., Weber, E. D., Watson, W., and Lennert-Cody, C. E. (2017). Correcting for bias in CalCOFI ichthyoplankton abundance estimates associated with the 1977 transition from ring to bongo net sampling. *Calif. Coop. Ocean. Fish. Invest. Rep.* 58, 113–123.

Thompson, A. R., Schroeder, I. D., Bograd, S. J., Hazen, E. L., Jacox, M. G., Leising, A., et al. (2019b). State of the California Current 2018-2019: A novel anchovy regime and a new marine heat wave? *Cal. Coop. Ocean. Fish. Invest.* 60, 1–65.

Thorson, J. (2018). Forecast skill for predicting distribution shifts: A retrospective experiment for marine fishes in the Eastern Bering Sea. *Fish Fish.* 20, 159–173. doi: 10.1111/faf.12330

Tommasi, D., Nye, J., Stock, C., Hare, J. A., Alexander, M., and Drew, K. (2015). Effect of environmental conditions on juvenile recruitment of alewife (*Alosa pseudoharengus*) and blueback herring (*Alosa aestivalis*) in fresh water: a coastwide perspective. *Can. J. Fish. Aquat. Sci.* 72, 1037–1047. doi: 10.1139/cjfas-2014-0259

Tommasi, D., Stock, C. A., Pegion, K., Vecchi, G. A., Methot, R. D., Alexander, M. A., et al. (2017). Improved management of small pelagic fisheries through seasonal climate prediction. *Ecol. Appl.* 27, 378–388. doi: 10.1002/eap.1458

van der Sleen, P., Rykaczewski, R. R., Turley, B. D., Sydeman, W. J., García-Reyes, M., Bograd, S. J., et al. (2018). Non-stationary responses in anchovy (*Engraulis encrasicolus*) recruitment to coastal upwelling in the Southern Benguela. *Mar. Ecol. Prog. Ser.* 596, 155–164. doi: 10.3354/meps12567

van Oostende, N., Dussin, R., Stock, C. A., Barton, A. D., Curchitser, E., Dunne, J. P., et al. (2018). Simulating the ocean's chlorophyll dynamic range from coastal upwelling to oligotrophy. *Prog. Oceanogr.* 168, 232–247. doi: 10.1016/j.pocean.2018.10.009

Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., and Hilborn, R. (2013). Frequency and intensity of productivity regime shifts in marine fish stocks. *Proc. Nat. Acad. Sci. U.S.A.* 110, 1779–1784. doi: 10.1073/pnas.1214879110

Walsh, H. J., Richardson, D. E., Marancik, K. E., and Hare, J. A. (2015). Long-term changes in the distribution of larval and adult fish in the Northeast U.S. Shelf Ecosystem. *PloS One* 10 (9), e0137382. doi: 10.1371/journal.pone.0137382

Weber, E. D., Chao, Y., and Chai, F. (2018). Performance of fish-habitat classifiers based on derived predictors from a coupled biophysical model. *J. Mar. Sys.* 186, 105–114. doi: 10.1016/j.jmarsys.2018.06.012

Weber, E. D., and McClatchie, S. (2010). Predictive models of northern anchovy *Engraulis mordax* and Pacific sardine *Sardinops sagax* spawning habitat in the California Current. *Mar. Ecol. Prog. Ser.* 406, 251–263. doi: 10.3354/meps08544

Weber, E. D., and McClatchie, S. (2012). Effect of environmental conditions on the distribution of Pacific mackerel (*Scomber japonicus*) larvae in the California Current System. *Fish. Bull.* 110, 85–97.

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

Zwolinski, J. P., and Demer, D. A. (2019). Re-evaluation of the environmental dependence of Pacific sardine recruitment. *Fish. Res.* 216, 120–125. doi: 10.1016/j.fishres.2019.03.022

Zwolinski, J. P., Emmett, R. L., and Demer, D. A. (2011). Predicting habitat to optimize sampling of Pacific sardine (*Sardinops sagax*). *ICES J. Mar. Sci.* 68, 867–879. doi: 10.1093/icesjms/fsr038

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership