

PREDICTIVE MAPPING OF TWO NEARSHORE, DEMERSAL FISH SPECIES ON  
NORTHERN CALIFORNIA REEFS USING SCUBA-BASED VISUAL SURVEYS  
AND REMOTE SENSING

By

Christopher Hiroshi Teague

A Thesis Presented to

The Faculty of Humboldt State University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

Committee Membership

Dr. Brian Tissot, Committee Chair

Dr. Timothy Mulligan, Committee Member

Dr. Tim Bean, Committee Member

Dr. Sean Craig, Committee Member

Dr. Erik Jules, Program Graduate Coordinator

May 2018

## ABSTRACT

### PREDICTIVE MAPPING OF TWO NEARSHORE, DEMERSAL FISH SPECIES ON NORTHERN CALIFORNIA REEFS USING SCUBA-BASED VISUAL SURVEYS AND REMOTE SENSING

Christopher Hiroshi Teague

California's shallow rocky reefs provide critical habitat for a diverse assemblage of fishes. Effective management strategies for these species require both accurate stock assessments as well as a spatially explicit understanding of the relationship between fishes and characteristics of their habitat. We used a generalized additive model framework to create spatially predictive maps of the abundance and biomass of two demersal fish species prevalent on northern California reefs: lingcod (*Ophiodon elongatus*) and kelp greenling (*Hexagrammos decagrammus*). These models incorporated data from SCUBA-based fish and habitat surveys at depths from 12-26 meters as well as measures of seafloor topography derived from remotely sensed bathymetric surveys. Topographic position index, a measure of a location's elevation relative to its surroundings, was an important predictor for all chosen models. Percentage of rocky substrate and rugosity, a metric describing habitat complexity, were also important predictive variables in many of the chosen models. These findings indicate that these species have complex associations with specific habitat features and that they may select these features of their environment at multiple spatial scales. The results presented here

highlight the utility of combining remotely sensed habitat data with SCUBA-based visual surveys to aid in stock assessments and marine spatial planning.

## ACKNOWLEDGEMENTS

A great number of people were involved in this work and I would not have been able to do any of this without their guidance and support. First, I would like to thank my advisor, Dr. Brian Tissot. He has been a fantastic mentor during my time at HSU and he has helped me become achieve above and beyond all of the goals I set out to reach in the graduate program here. I also want to thank the rest of my thesis committee: Drs. Tim Mulligan, Sean Craig, and Tim Bean for their excellent feedback and helpful communication.

A great many thanks to all of the scientific divers who spent many cold days on the boat with me to collect all of my data: Johnathan Centoni, Franklin Moitoza, Brett Stacy, Jeff Bernard, and Johnny Roche. Thanks as well to the HSU dive safety officer Rich Alvarez and boating safety officer Steve Monk for help with field work logistics. I'm also very thankful to Dr. Andre Buchheister for all the help working through the statistics for this project.

None of this research would have been possible without funding from the CSU Council on Ocean Affairs, Science, & Technology (COAST), the Padi Foundation, the Humboldt State MCSI, and the Malcolm Oliphant Scholarship Program. I would also like to thank the HSU Biology department for providing teaching assistantships and the North Coast MLPA Baseline project for giving me the opportunity to scout sites, cover living expenses, and see many of the wonders of the North Coast's shallow subtidal reefs.

Many thanks to everyone in the Tissot Benthic Ecology Lab for all of their support. I couldn't have asked for a better group of people to work with. In particular, I want to thank my lab twin, Jana Litt, for being a wonderful counterpart in Brian's lab in those first years when it was just the two of us. I want to thank my wife, Kristine, for her constant love, encouragement, and one-upmanship over the years that has pushed me to pursue all of my goals. I am also very thankful for my parents, Lisa and Tim, who fostered my love of science from a very young age and who have supported me mentally, emotionally, and financially throughout my academic career. Finally I want to thank my grandfather, Richard Shomura, for inspiring me to become a marine biologist. He taught me that the major problems in the world will be solved by looking to the ocean.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
LIST OF APPENDICES .....	xi
INTRODUCTION .....	1
METHODS .....	7
Study Area .....	7
Habitat Mapping .....	9
SCUBA-based Visual Surveys .....	10
Data Analysis .....	14
RESULTS .....	18
Relationships between fish and <i>in situ</i> habitat variables .....	21
Generalized additive models: Kelp greenling .....	22
Generalized additive models: Lingcod .....	28
DISCUSSION .....	34
Predictor and response variable relationships .....	35
Limitations of methodology .....	39
Management implications .....	44

CONCLUSIONS.....	46
REFERENCES .....	47
APPENDIX A.....	53
APPENDIX B .....	57

## LIST OF TABLES

Table 1: Distribution of sampling stations across sites and habitat strata. ....	12
Table 2: Final generalized additive models for kelp greenling and lingcod density and biomass .....	23



## LIST OF FIGURES

Figure 1: Study area along the Mendocino coast in Northern California. The two sites, Abalone Point and the Ten Mile State Marine Reserve are outlined in green. Purple points indicate survey locations. ....	8
Figure 2: Average per-transect densities of all species or species groups observed during SCUBA surveys .....	19
Figure 3: Size frequency distributions for kelp greenling between Abalone Point and the Ten Mile State Marine Reserve. ....	20
Figure 4: Size frequency distributions for lingcod between Abalone Point and the Ten Mile State Marine Reserve.....	20
Figure 5: Canonical Correspondence Analysis (CCA) on associations between five abundant fish species observed at the study site and two categorical habitat variables: substrate and vertical relief. Black points are sampling locations, red triangles are species, and blue vectors indicate habitat variables. ....	22
Figure 6: Generalized additive model response curves for kelp greenling density versus (a) (percent hard substrate) <sup>2</sup> and (b) fine-scale topographic position index (TPI). Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models.....	24
Figure 7: Generalized additive model response curves for kelp greenling biomass versus (a) (percent hard substrate) <sup>2</sup> and (b) fine-scale topographic position index (TPI). Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models.....	24
Figure 8: Predicted kelp greenling density. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the kelp greenling density model. ....	26
Figure 9: Predicted kelp greenling biomass. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the kelp greenling biomass model. ....	27

Figure 10: Generalized additive model response curves for lingcod density versus (a) fine-scale topographic position index (TPI) and (b) broad-scale topographic position index. Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models..... 29

Figure 11: Generalized additive model response curves for lingcod density versus (a) fine-scale rugosity (Vector Ruggedness Measure, VRM), (b) broad-scale rugosity, (c) fine-scale topographic position index (TPI), and (d) broad-scale topographic position index. Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models..... 30

Figure 12: Predicted lingcod density. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the lingcod density model..... 32

Figure 13: Predicted lingcod biomass. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the lingcod biomass model..... 33

## LIST OF APPENDICES

APPENDIX A.....	53
APPENDIX B.....	57

## INTRODUCTION

Widespread accounts of fishery declines, habitat degradation, and other anthropogenic influences on the marine realm have driven calls for innovative new strategies to better manage fisheries and preserve biodiversity (Crowder & Norse 2008, Jackson 2008, Halpern et al. 2008, Carr et al. 2011). Since marine species and the threats they face are not distributed randomly across the seascape, it is critically important that these approaches are placed in a spatially-explicit, biogeographic context (Roberts et al. 2003, Hamilton et al. 2010). Additionally, effective management strategies require a detailed understanding of the environmental factors that drive species' distributions (Pittman & Brown 2011).

Many studies have described biogeographic patterns in the distributions of temperate demersal fishes as well as associations between these species and the physical environment in which they reside (Stein et al. 1992, reviewed in Stephens et al. 2006, Love & Yoklavich 2008). These studies, however, are often subject to a trade-off between high-resolution and broad-scale. For example, visual techniques that are commonly used for reef fish community assessments, such as SCUBA and submersible surveys, provide high resolution data on species-habitat relationships, but only within the relatively narrow regions sampled. While these methods provide data at scales relevant to the ecological requirements of reef fishes, they often offer strictly qualitative information and do not allow for quantitative predictions outside of the surveyed area. Thus, these

methods are limited in their ability to make inferences across the broad geographic scales necessary for effective management (Chatfield et al. 2010, Young & Carr 2015).

Advances in seafloor mapping technology have allowed for the collection of high-resolution benthic habitat data across broad swaths of the seafloor. The California Seafloor Mapping Program (CSMP) was implemented by the state of California in order to create bathymetric digital elevation models (DEMs), often at resolutions of 2m per pixel, that span the coastline from shore to the state's three nautical mile limit. In addition to depth, a range of other habitat variables that may be important determinants of fish distributions, such as slope, topographic position, and vertical relief, can then be derived from the DEMs. Previous work has indicated that many of these variables are important determinants of fish presence, abundance, and biomass (Young et al. 2010, Wedding & Yoklavich 2015). Furthermore, by collecting these data at such fine resolutions, fish-habitat associations can be assessed at multiple spatial scales. Multiscale approaches are necessary for determining fish distributions, as organism-habitat relationships may differ between fine- and broad-scales (Wiens 1989).

Methods that link spatially-explicit species data with high-resolution seafloor maps, such as species distribution models (SDMs), can yield quantitative information on the spatial distribution of species and communities as well as the environmental factors that drive these distributions (Pittman et al. 2009, Knudby et al. 2010). An important component of SDMs is that they can be used in a predictive capacity to estimate a range of population and community metrics including probability of occurrence, abundance, biomass, and diversity at locations that have not been sampled previously (Pittman et al.

2007, Iampietro et al. 2008, Knudby et al. 2010, Wedding & Yoklavich 2015). By making predictions outside of the original sampling areas, SDMs can extend the utility of visual survey techniques to larger regional scales more relevant to management. These models thus have the potential to meet a range of management and conservation goals including the delineation of essential habitat features (Anderson & Yoklavich 2007, Ortiz & Tissot 2012), marine reserve siting (Rees et al. 2014), and accurate stock assessments (Nasby-Lucas et al. 2002, Young & Carr 2015).

SDMs can provide crucial information to fishery management efforts on shallow reefs along the west coast of the United States. Northern California's nearshore reefs are characterized by a highly complex, rocky structure which supports a high diversity of demersal fish species including rockfishes, greenlings, and surfperches (Jenkinson & Craig 2017). These species make up a substantial portion of the benthic fauna and are important components of both recreational and commercial fisheries in the region (Lea et al. 1999, Love 2006, Stephens et al. 2006). Although they are often targets of conservation efforts, there is still a lack of data on the distribution and stock size of many nearshore demersal fish species. Formal stock assessments have only been conducted for roughly half of the 87 groundfish species managed under the Pacific Fishery Management Council ([www.pcouncil.org/groundfish/stock-assessments/](http://www.pcouncil.org/groundfish/stock-assessments/)), many of which are present on California's shallow reefs. This is particularly problematic in Northern California because, historically, this region has not been surveyed as often as other parts of the state (Allen & Pondella 2006). SCUBA-based visual techniques are the primary survey method employed in these shallow habitats, however it is costly to conduct these

techniques across large areas (Jenkinson & Craig 2017). By incorporating data from visual surveys and high resolution bathymetric maps, SDMs extend the utility of visual techniques across the broad spatial scales necessary for effective marine management.

Here, we used a spatial modeling framework to create models and predictive maps of density and biomass for kelp greenling (*Hexagrammos decagrammus*) and lingcod (*Ophiodon elongatus*). Both species belong to the family *Hexagrammidae*. Kelp greenling are common from the Aleutian Islands to Central California and are found at depths up to 130m. Lingcod are found between the Gulf of Alaska and central Baja California up to depths of 475m (Love 2011). The maximum size of kelp greenling is 63cm and 2.1 kg (Love 2011). Lingcod reach a maximum size of 150cm (Cass et al. 1990) and 37.5kg (Love 2011).

Both species are bottom-oriented and adults are most often found in rocky habitats (Miller & Geibel 1973, Cope & MacCall 2005). Feeding habits and predator avoidance likely play important roles in driving this preference for reef structure (Stephens et al. 2006). Kelp greenling are opportunistic generalists that eat a wide variety of bottom-oriented foods often associated with rocky habitats including shrimp, crabs, octopuses, and amphipods (Love 1996, Howard & Silberberg 2001). Kelp greenling inhabit rocky habitats of any shape and size and are almost always found amongst algal or surfgrass cover (Cope & MacCall 2005, Love 2011), which may provide cover from predation.

Adult lingcod are primarily piscivorous though they will also feed on octopuses, squid, hermit crabs, fish eggs, and hydroids (Love 2011). In California, a substantial

portion of lingcod diets are made up of juvenile rockfishes, many species of which are in high abundance on complex, rocky reefs (Adams & Starr 2001, Stephens et al. 2006). Lingcod are often found on prominent reef features, which may be advantageous for ambushing prey (Lynn 2008, Love 2011). Structural complexity, particularly the presence of overhangs, cracks, and crevices, also appears to be highly important to lingcod as it provides ideal nesting locations and cover from predators (Stephens et al. 2006, Lynn 2008, Love 2011).

In Northern California, lingcod make up a substantial component of both the commercial and recreational fisheries (Jagiello & Wallace 2005). Kelp greenling are often targeted in the commercial live-fish fishery as well as in the recreational fishery via shore-based angling (Howard & Silberberg 2001, Berger et al. 2015). Although both species are important for fisheries, there is a paucity of data available to inform management. Cope and MacCall (2005) provided the only stock assessment of kelp greenling in California waters, however it has not been used to advise management due to insufficient information on its population structure (Berger et al. 2015). While assessments of lingcod have been conducted several times in the past thirty years, they often rely on trawl-based catch and survey data (Jagiello & Wallace 2005, Hamel et al. 2009). These data sources, however, do not encompass the full range of habitats used by lingcod, specifically rocky habitats which are untrawlable due to their high-relief nature (Zimmermann 2003).

In this study, we first use data from SCUBA-based fish and habitat surveys at two Northern California shallow rocky reef sites to examine relationships between kelp



greenling, lingcod, and the *in situ* habitat data. We then combine the survey data with remotely sensed multibeam bathymetry to develop models of density and biomass for each species. Next, we use these models to evaluate which habitat features are most highly associated with these species and determine the spatial scales at which these relationships occur. Finally, we create predictive maps of density and biomass for kelp greenling and lingcod using model outputs and assess the predictive accuracy of each map using a cross-validation procedure.

## METHODS

### Study Area

The focal sites for this study were the Ten Mile State Marine Reserve (SMR) (39.585° N, 123.786° W) and Abalone Point (39.667° N, 123.801° W). Both sites are located in Mendocino County, California several miles north of Ft. Bragg and are separated by six kilometers (Fig. 1). This area is characterized by high-relief, rocky reef structures that extend from shore to a depth of roughly 80 m. Rocky sections of the reef are separated by sand channels ranging from 10s to 100s of m in width. Adult fish assemblages are dominated by blue and black rockfish (*Sebastes mystinus* and *S. melanops*, respectively) and kelp greenling. Several other species including striped surfperch (*Embiotoca lateralis*), lingcod, and other rockfish species (genus *Sebastes*) are also present, but at lower abundances. The Ten Mile site is a no-take state marine reserve that was established in 2012. Due to the short amount of time since its implementation, there is a low likelihood of any reserve effect afforded by its protected status (Jenkinson & Craig 2017). Abalone Point remains open to recreational and commercial fishing activities.

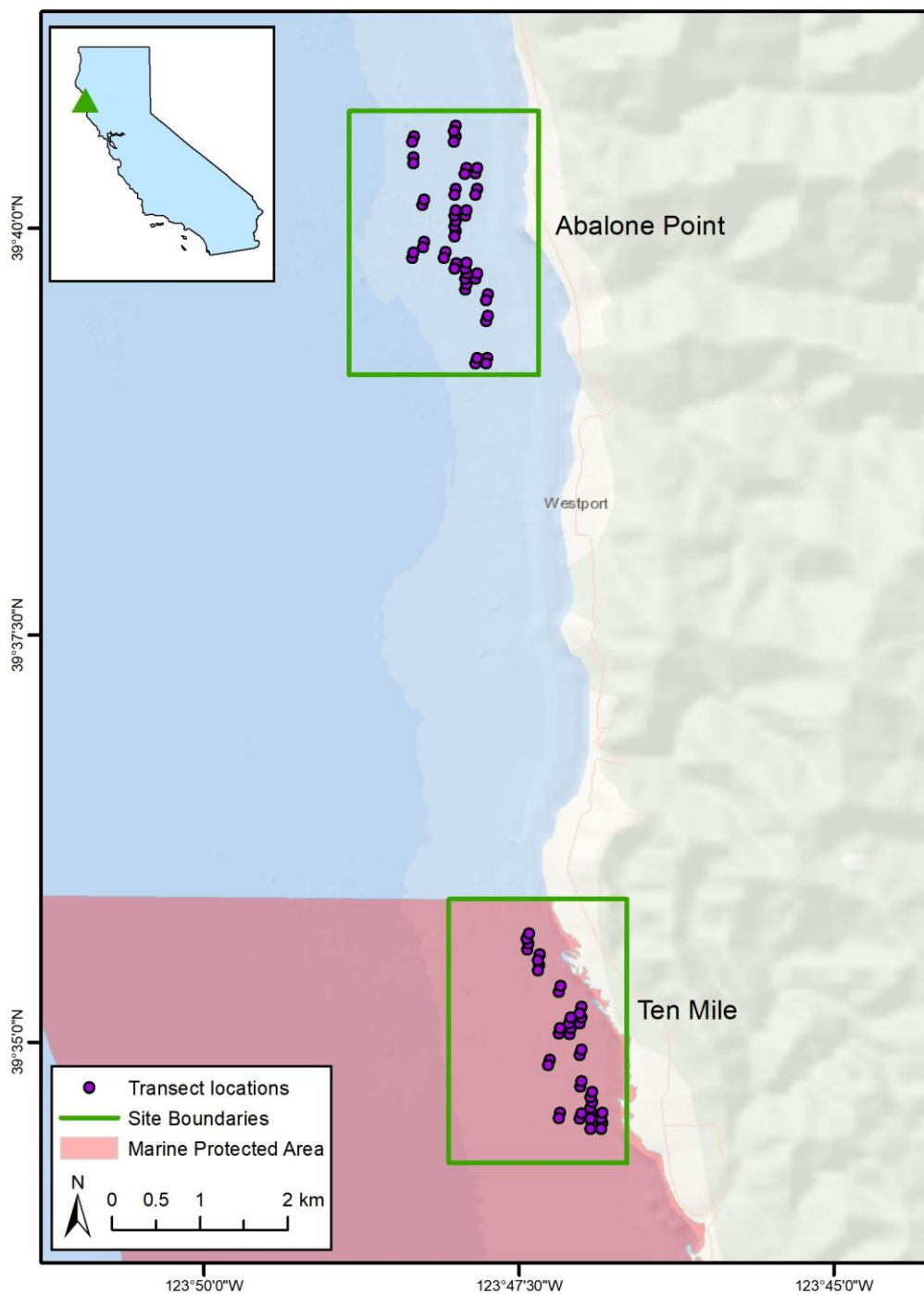


Figure 1: Study area along the Mendocino coast in Northern California. The two sites, Abalone Point and the Ten Mile State Marine Reserve are outlined in green. Purple points indicate survey locations.

## Habitat Mapping

Multibeam bathymetric data were obtained from the California Seafloor Mapping Project (CSMP 2010), a collaborative effort to create a detailed set of seafloor bathymetric maps throughout California state waters from the shoreline to the 3 nautical-mile limit. Since the project's inception in 2005, the CSMP have successfully mapped the majority of the coastline and have made their products freely available to the public. CSMP surveys of the focal sites were conducted between August 22 and October 31, 2009 using ship-based multibeam echo sounders. The CSMP post-processed the multibeam data and produced final map products. The primary product of these surveys is a series of high resolution digital elevation models (DEMs) provided in raster format. In this study, we used the highest resolution DEMs available, which have a resolution of 2m per pixel and cover a depth range of 2-84m. These maps have a reported horizontal position accuracy of  $\pm 2\text{m}$  and vertical accuracy of  $\pm 5\text{cm}$  (CSMP 2010).

A range of other habitat variables were derived from the DEM using the Spatial Analyst toolbox in ArcMap (Environmental Systems Research Institute, ESRI). Slope and slope-of-the-slope were calculated using the slope function in the Spatial Analyst toolbox. Grids of vector ruggedness measure (VRM), a rugosity metric, were created using the Benthic Terrain Modeler toolbox. VRM measures terrain ruggedness as the variation in the 3-dimensional orientation of grid cells in a DEM (Hobson 1972, Sappington et al. 2007). VRM is based on a user-specified neighborhood of cells and can thus be calculated at multiple spatial scales. Here, VRM rasters were created at six scales

ranging from fine to broad: 6, 10, 50, 118, 198, and 498 meters. These scales correspond to the neighborhood size used in raster creation (e.g. VRM6 used to a neighborhood of 6 x 6 meters).

Rasters denoting substrate type (rough/smooth classification) and topographic position index (TPI) were provided by the CSMP. The substrate raster is based on the VRM6 raster, however it has been reclassified such that values below a cutoff are considered smooth (soft sediment) and higher values are considered rough (hard substrate). TPI is a second order derivative that describes a given point based on its surroundings by comparing the elevation of each cell in a DEM to the mean elevation of neighboring cells (Weiss 2001, Lundblad et al. 2006). High TPI values indicate peaks and ridges, whereas low values indicate valleys. Like VRM, TPI can be calculated at multiple scales. The CSMP calculated TPI using an annulus neighborhood with 20m, 50m, and 250m outer radii to yield fine, medium, and broad scales, respectively (labeled hereafter as TPI20, TPI50, and TPI250).

### SCUBA-based Visual Surveys

Survey stations were selected across both sites based on a stratified random sampling design in order to ensure that a broad range of habitat types were surveyed. Strata were defined by two depth zones (10-20m and 20-30m), three substrate types (hard, soft, and mixed), and three levels of habitat complexity (low, med, and high). Using ArcMap, a grid of 120 x 120m blocks was overlaid across the two sites. Mean

depth, percent hard substrate, and habitat complexity were extracted for each block using the DEM, substrate, and VRM6 rasters, respectively. Stations were then chosen at random from each combination of variables (e.g. 10-20m, hard, high-relief). Since smooth values in the substrate raster are equivalent to low VRM6 values, blocks designated as soft substrate were not further split into varying levels of complexity. A total of 46 stations were surveyed across Abalone Point and the Ten Mile SMR (Table 1).

Dive surveys of demersal fishes and habitat characteristics were conducted from July to September in 2015 and May through August in 2016. Survey locations were based on the center point of each station. Upon arrival at a station, a weight attached to a float and line was deployed in order to minimize drift away from the actual station coordinates as divers descended. Each dive consisted of two 50 x 2 x 2m band transects. After descent to the seafloor, divers extended a 10m line from the weight location due north before beginning the transect. This was repeated to the south for the second transect. This process ensured a distance of 20m between transects in order to reduce the possibility of counting the same fish multiple times. Transects were aligned along a rough north-south orientation as a safety precaution to limit excessive changes in depth experienced by the divers.

Table 1: Distribution of sampling stations across sites and habitat strata.

Stratum	# of Stations: Abalone Point	# of Stations: Ten Mile
10-20m, hard, high relief	5	5
10-20m, hard, med relief	5	5
10-20m, hard, low relief	5	5
10-20m, soft	2	2
10-20m, mixed	3	2
20-30m, hard, high relief	1	1
20-30m, hard, med relief	1	1
20-30m, hard, low relief	1	0
20-30m, soft	1	0
20-30m, mixed	1	0

On each transect, one diver conducted fish surveys where all conspicuous fishes within the transect volume were counted and identified to the lowest possible taxon (IACUC #14/15.F.91-A, 05 May 2015). Several species in this region are too similar in appearance to distinguish between species and were therefore placed in broader umbrella groups. This applied primarily to two such groups consisting primarily of juvenile rockfishes: OYT, which is composed of olive (*S. serranoides*) and yellowtail (*S. flavidus*) rockfishes, and KGBC, which is an umbrella term used by several other fish survey organizations in California that generally includes kelp (*S. atrovirens*), gopher (*S.*

*carnatus*), black-and-yellow (*S. chrysomelas*), and copper (*S. caurinus*) rockfishes. In Northern California, however, kelp rockfish are rare and were likely not observed on our surveys. Juvenile quillback (*S. maliger*) and china (*S. nebulosus*) rockfish also may have been identified as KGBC, since smaller individuals can look similar to the other species within this grouping (Flannery 2018). The total length of each fish observed was estimated visually to the nearest centimeter. These methods are a modification of the protocols used by the Partnership for Interdisciplinary Study of Coastal Oceans (PISCO 2016). Total lengths were later converted to biomass based on length-weight relationships taken from either the literature (Lea et al. 1999, Love 2011) or recfin.org.

A second diver followed the fish surveyor along the transect line collecting habitat data using a uniform point contact (UPC) design also modified from PISCO methods. Depth was measured both at the descent weight and every two meters on the transect. Substrate type was recorded every two meters as one of four categories (sand, cobble, boulder, and bedrock) based on grain size. Vertical relief, measured as the greatest change in reef height within a 0.5 x 1m area surrounding each two meter mark, was also recorded as one of four categories: low (0-10cm), medium-low (10cm-1m), medium-high (1-2m), and high (greater than 2m). Additionally, compass headings were taken every five meters. Heading data, combined with the known transect start coordinates, allowed transects to be digitally recreated in ArcMap. Heading measurements taken in the field were corrected based on a 14° declination to account for the difference between magnetic and true north at the study location.



## Data Analysis

While data were collected on all fishes observed, analyses employed a size cutoff based on length at one year for each species in order to exclude juveniles. This was deemed necessary as many reef associated species exhibit ontogenetic shifts wherein habitat requirements differ between juveniles and adults (Love et al. 1991, Stein et al. 1992, Love 2011). Length at one year was calculated using the von Bertalanffy growth equation and published parameters for each species (Cope & MacCall 2005, Jagiello & Wallace 2005). The size cutoffs used for kelp greenling and lingcod were 22cm and 36cm, respectively. After the dive surveys were completed, each transect was digitally recreated in ArcMap using compass headings and known transect start coordinates. Distances traveled were corrected to account for changes in depth.

Raster-based habitat variables were extracted to the midpoint of each transect using a 25m radius roving window mean within ArcMap's focal statistics tool. For the substrate raster, this yielded the percent of hard substrate within the window (hereafter denoted as SUB). Additionally, *in situ* depth data were attached to each 2m mark along the transect to compare SCUBA-based depth data with the remotely sensed DEM. Accuracy of the remotely sensed data was tested by comparing DEM derived depth against *in situ* depth at both the weight location and at every 2m mark using non-parametric Spearman's rho correlation coefficient.

Canonical correspondence analysis (CCA) is an ordination technique that detects correlations between two sets of variables (Ter-Braak 1986). CCA was used to determine

relationships between fish survey data and *in situ* habitat variables (i.e. depth, substrate, and relief). Although we focus here on kelp greenling and lingcod, CCA did not produce reliable results when only two species were used in the analysis. Therefore, in addition to kelp greenling and lingcod, we included the 3 other most abundant fish species observed (blue rockfish, black rockfish, and striped surfperch). Separate CCA plots were created for fish density and biomass. Depth values used were the means of all measurements taken on each transect. Each category of substrate and relief was converted to percentage per-transect by dividing the number of observations for a given category by the number of UPC points on the transect. CCA was carried out in R using the ‘vegan’ package (Oksanen et al. 2017).

Generalized additive models (GAMs, Wood 2006) were used to create predictive models of density and biomass for kelp greenling and lingcod. GAMs are often ideal for ecological data as they apply smoothing functions across the data to allow for complex, non-linear relationships between a response variable and multiple explanatory variables (Zuur et al. 2009). Protocols in Zuur et al (2009) were used for initial data exploration prior to the model building process. Presence of outliers and the need for data transformations were determined using Cleveland dotplots. SUB was the only highly skewed variable so we applied a square transformation. Variance inflation factors (VIF) and multipanel scatterplots were used to examine multicollinearity between predictor variables. Variables with high VIF values were removed from the analysis until all values were below 3.5. After completing these procedures, six predictor variables remained for

model building: depth, SUB<sup>2</sup>, fine-scale VRM (VRM6), broad-scale VRM (VRM498), fine-scale TPI (TPI20), and broad-scale TPI (TPI250).

GAMs were produced in R with the package ‘mgcv’ (Wood 2006). Density models for both species were built using a Poisson distribution and log link. The Poisson distribution is ideal for count and density data as it assumes non-negative response values and the log link ensures that fitted values are also non-negative (Zuur et al. 2009). Biomass models were created with a Tweedie distribution and log link to allow for continuous data with a substantial zero point mass (Dunn & Smyth 2005). An upper limit was set for the degrees of freedom of smoothing functions for all variables to reduce overfitting. Degrees of freedom were limited to either 3 or 4 based on visual examination of smoothers and by comparing effective degrees of freedom with the basis dimension (Wood 2006).

Model selection was conducted by a backwards stepwise approach in which each model started with the full set of six candidate predictor variables. Variables were then removed iteratively until Akaike’s information criterion (AIC) values were minimized (Zuur et al. 2009). Final models were validated based on protocols given in Zuur et al (2009). Model deviance residuals were assessed for normality (Q-Q plot), homogeneity of variance (residuals vs fitted values), independence (residuals vs explanatory variables), and spatial autocorrelation (residuals vs coordinates).

Predictive accuracy of each model was assessed using a leave-one-out cross validation procedure. Final models were retrained on the dataset after removing a single sample and predictions were made for the removed sample. Model accuracy is given as

the correlation between these predicted values and field observations using Spearman's rho correlation coefficient. The R package 'raster' was used to create predictive maps of density and biomass for kelp greenling and lingcod across the study area (Hijmans 2016). Predictions were limited to only those values observed in the field for both predictors and response variables in order to limit any errors due to extrapolating past the values used to train the model. Standard error maps were also produced for each final model to examine the spatial distribution of model uncertainty.

## RESULTS

A total of 92 transects covering 9200 m<sup>2</sup> were conducted across Abalone Point (50 transects) and the Ten Mile SMR (42 transects). Mean transect depths based on *in situ* measurements ranged from 11.7m to 25.9m. Depths measured *in situ* were highly correlated with the DEM at both the weight location (Spearman's  $\rho = 0.855$ ,  $p < 0.001$ ) and at every two meter mark (Spearman's  $\rho = 0.911$ ,  $p < 0.001$ ). The seafloor substrate was composed of 58% bedrock, 8% boulder, 2% cobble, and 32% sand. Vertical relief on all transects was 23% low (0-10cm), 58% medium-low (10cm-1m), 14% medium-high (1-2m), and 5% high (>2m) relief.

A total of 21 species were observed during the SCUBA surveys. Fish assemblages across both sites were dominated by blue and black rockfishes (42% and 17% of total individuals observed, respectively; Fig. 2). Also in high abundance were juvenile rockfishes belonging to the OYT and KGBC groups as well as juvenile canary rockfish. The most abundant non-rockfish species were kelp greenling, striped surfperch, and lingcod. We observed 162 kelp greenling and 76 lingcod during the SCUBA surveys. Kelp greenling size frequency distributions show similar patterns between Abalone Point and the Ten Mile SMR (Fig. 3). Size frequencies for lingcod were largely similar for lingcod above 30cm total length between the two sites, however individuals smaller than 30cm were observed only at Abalone Point (Fig. 4). 158 kelp greenling and 57 lingcod were considered adults after applying size cutoffs and were included in the ordination and

modeling analyses. Adult kelp greenling ranged in size from 22 to 46cm total length and adult lingcod ranged from 36 to 120cm total length.

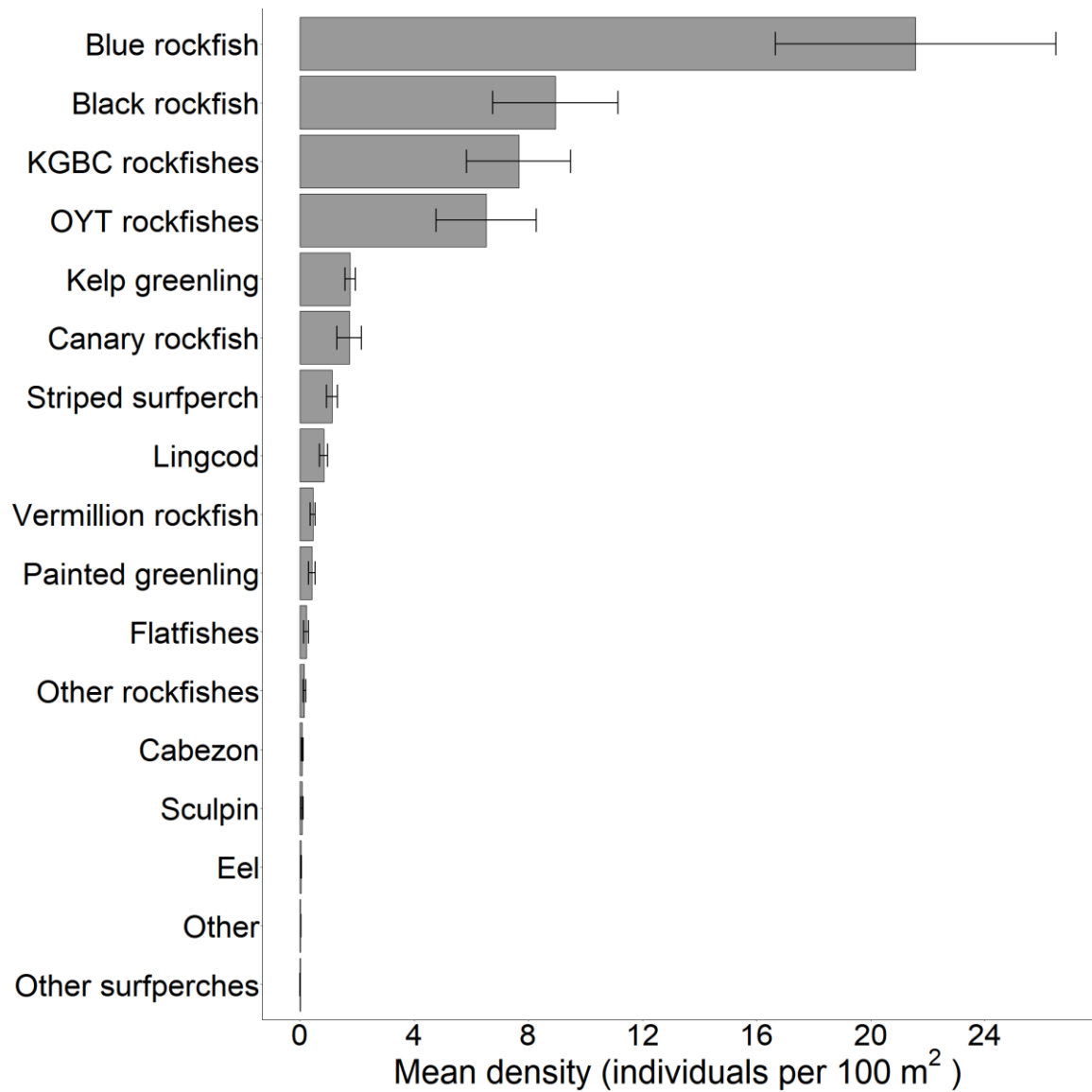


Figure 2: Average per-transect densities of all species or species groups observed during SCUBA surveys

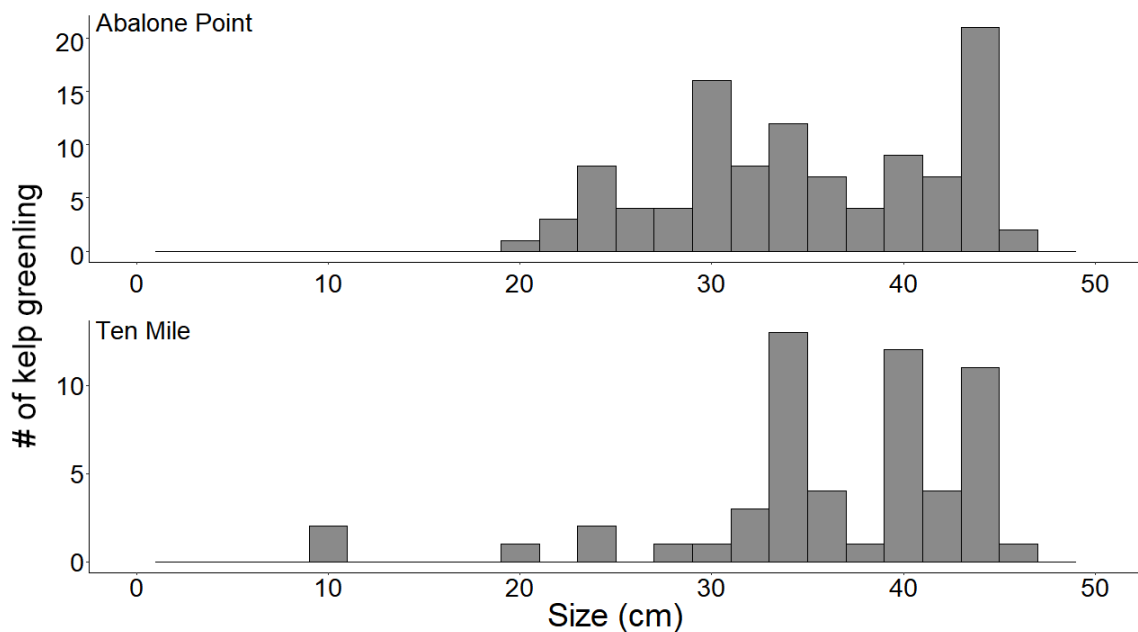


Figure 3: Size frequency distributions for kelp greenling between Abalone Point and the Ten Mile State Marine Reserve.

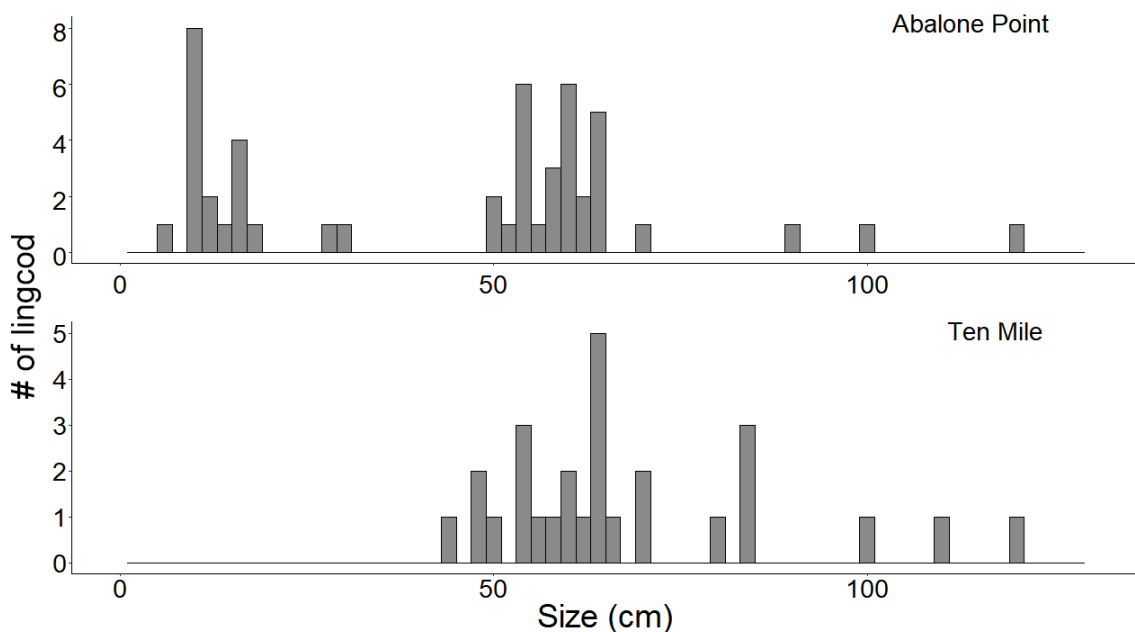


Figure 4: Size frequency distributions for lingcod between Abalone Point and the Ten Mile State Marine Reserve.

### Relationships between fish and *in situ* habitat variables

CCA showed significant associations between fish biomass and *in situ* habitat metrics (Fig. 5) (Chi-square,  $p < 0.05$ ). Overall, the CCA captured 27% of the total variation in the species data. The majority of this variation is explained by axes 1 (63%) and 2 (24%). Axis 1 was primarily associated with medium-low relief (vector length = .99,  $p < 0.05$ ), which increased along the axis. Medium-high and high relief had large vector lengths that decreased along axis 1 (-0.98 and -0.94, respectively), however these were not significant ( $p > 0.05$ ). Depth increased substantially along axis 2 (vector length = 0.97), but the relationship was only marginally significant ( $p < 0.10$ ). Kelp greenling biomass was most highly associated with boulder habitats and medium-low vertical relief (Fig. 5). Lingcod biomass was greatest in medium-high relief reef structures composed primarily of bedrock and, to a lesser extent, cobble. Neither species was associated with deep, low-relief, or sandy habitats. While CCA was able to determine relationships between fish biomass and *in situ* habitat variables, CCA based on fish abundance was not significant (Chi-square,  $p = 0.25$ ).



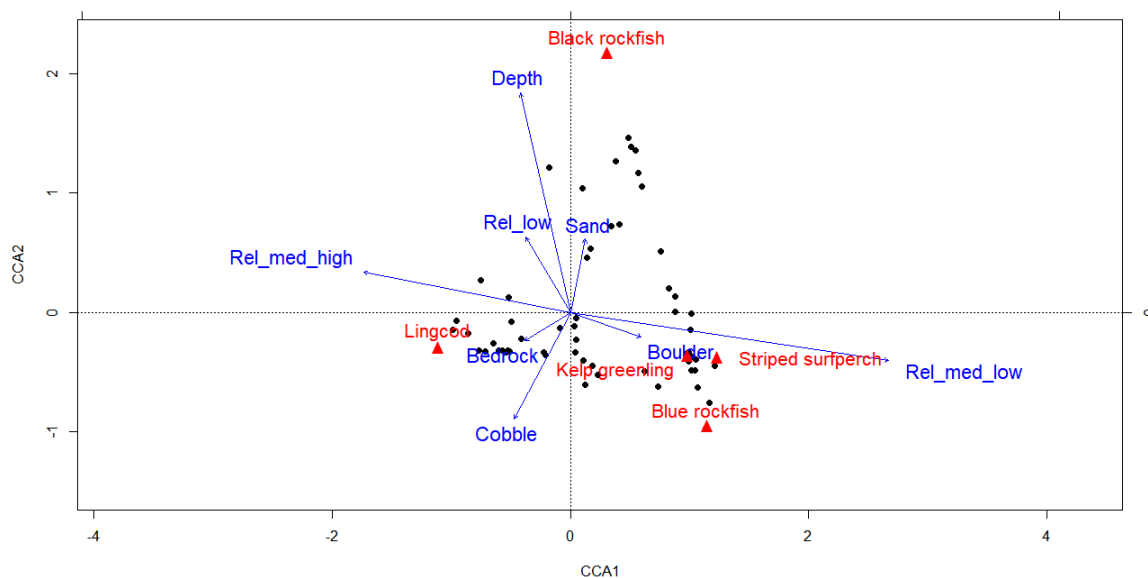


Figure 5: Canonical Correspondence Analysis (CCA) on associations between five abundant fish species observed at the study site and two categorical habitat variables: substrate and vertical relief. Black points are sampling locations, red triangles are species, and blue vectors indicate habitat variables.

### Generalized additive models: Kelp greenling

The chosen GAM for kelp greenling density identified  $SUB^2$  and fine-scale topographic position as the most important predictors (Table 2, see appendix for full model selection table). This model explained 37% of the deviance and had an adjusted  $R^2$  of 0.31. Model predictions of kelp greenling density had an accuracy of Spearman's  $\rho = 0.53$ . Response curves indicate that density increased with  $SUB^2$  initially before leveling off (Fig. 6a, see Appendix B for univariate plots of response variables versus each predictor variable included in final models). The response curve comparing density with

TPI20 showed that density increased with higher values of fine scale topographic position, corresponding to relative shallow points on the reef (Fig. 6b).

Similar to the density model, the selected GAM for kelp greenling biomass indicated that  $SUB^2$  and TPI20 were the most important predictors (Table 2). The model explained 34% of the deviance and had an adjusted  $R^2$  of 0.25. Model predictions of kelp greenling biomass had an accuracy of Spearman's  $\rho = 0.48$ . Unlike the density response curve, biomass had a nearly linear positive relationship with  $SUB^2$  throughout all values (Fig. 7a). The response curve comparing biomass with TPI20 was similar to the density response curve with biomass maximized at relative shallow parts of the reef (Fig. 7b).

Table 2: Final generalized additive models for kelp greenling and lingcod density and biomass

Species	Formula	R-sq. (adj.)	Deviance Explained (%)
Kelp greenling	Density ~ $SUB^2 + TPI20$	0.31	36.9
	Biomass ~ $SUB^2 + TPI20$	0.254	34.4
Lingcod	Density ~ $TPI20 + TPI250$	0.254	30.1
	Biomass ~ $VRM6 + VRM498 + TPI20 + TPI250$	0.409	48.9

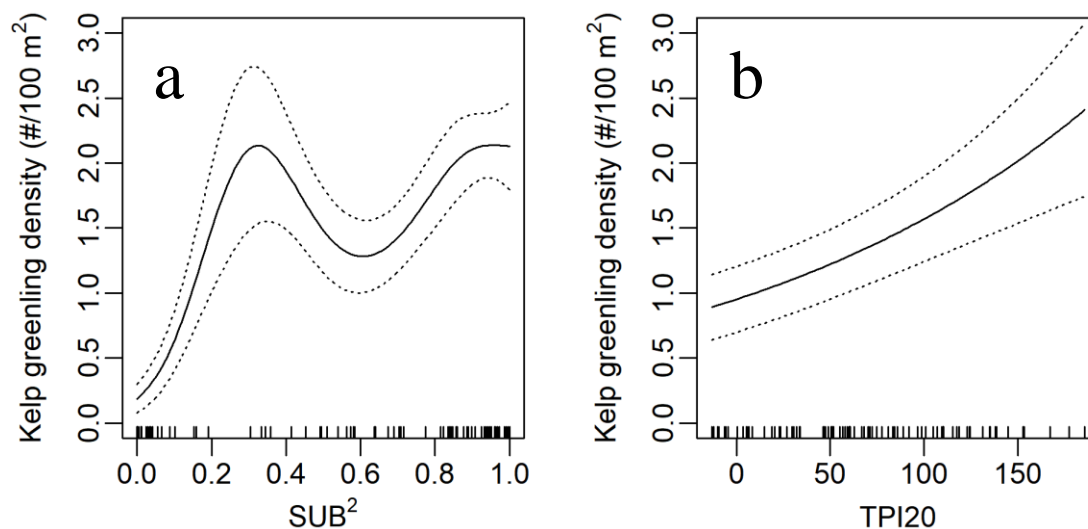


Figure 6: Generalized additive model response curves for kelp greenling density versus (a) (percent hard substrate)<sup>2</sup> and (b) fine-scale topographic position index (TPI). Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models.

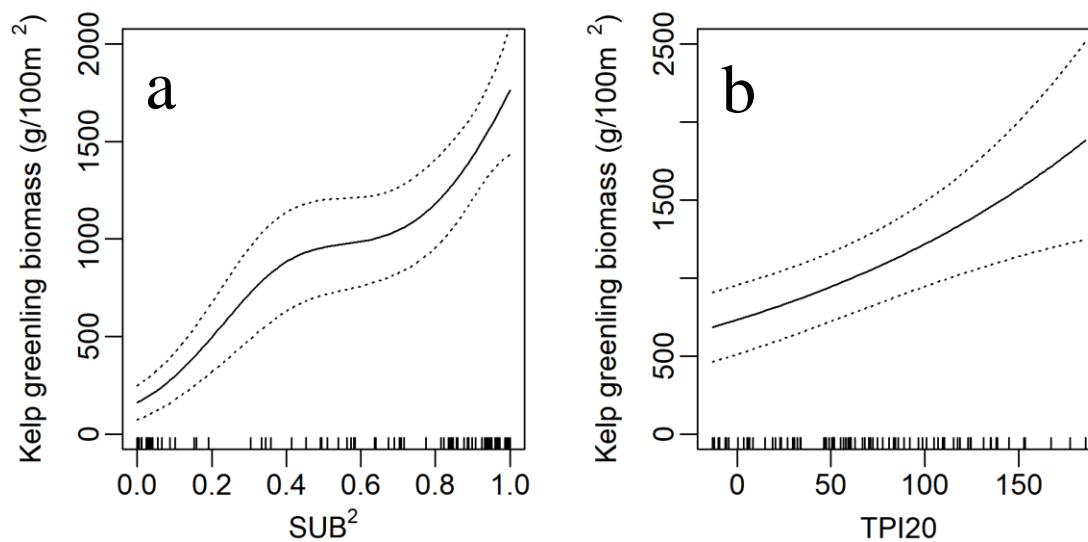


Figure 7: Generalized additive model response curves for kelp greenling biomass versus (a) (percent hard substrate)<sup>2</sup> and (b) fine-scale topographic position index (TPI). Solid lines = mean, dashed lines = +/- SE. Rug plot along x-axis indicates observed values used to train models.

Predictive maps produced nearly identical trends in the distribution of both density and biomass of kelp greenling (Figs. 8 & 9). The predicted distributions were patchy at fine spatial scales (10s of m) and this patchiness was spread evenly on reef structures across both sites. Projected values of both population metrics were highest in rocky habitats and either low or non-existent on soft sediments. Additionally, the highest predicted values were found in areas containing a large number of prominent reef outcrops. Maps of the standard error associated with both density and biomass indicated low levels of uncertainty across the study area (Figs. 8d & 9d).

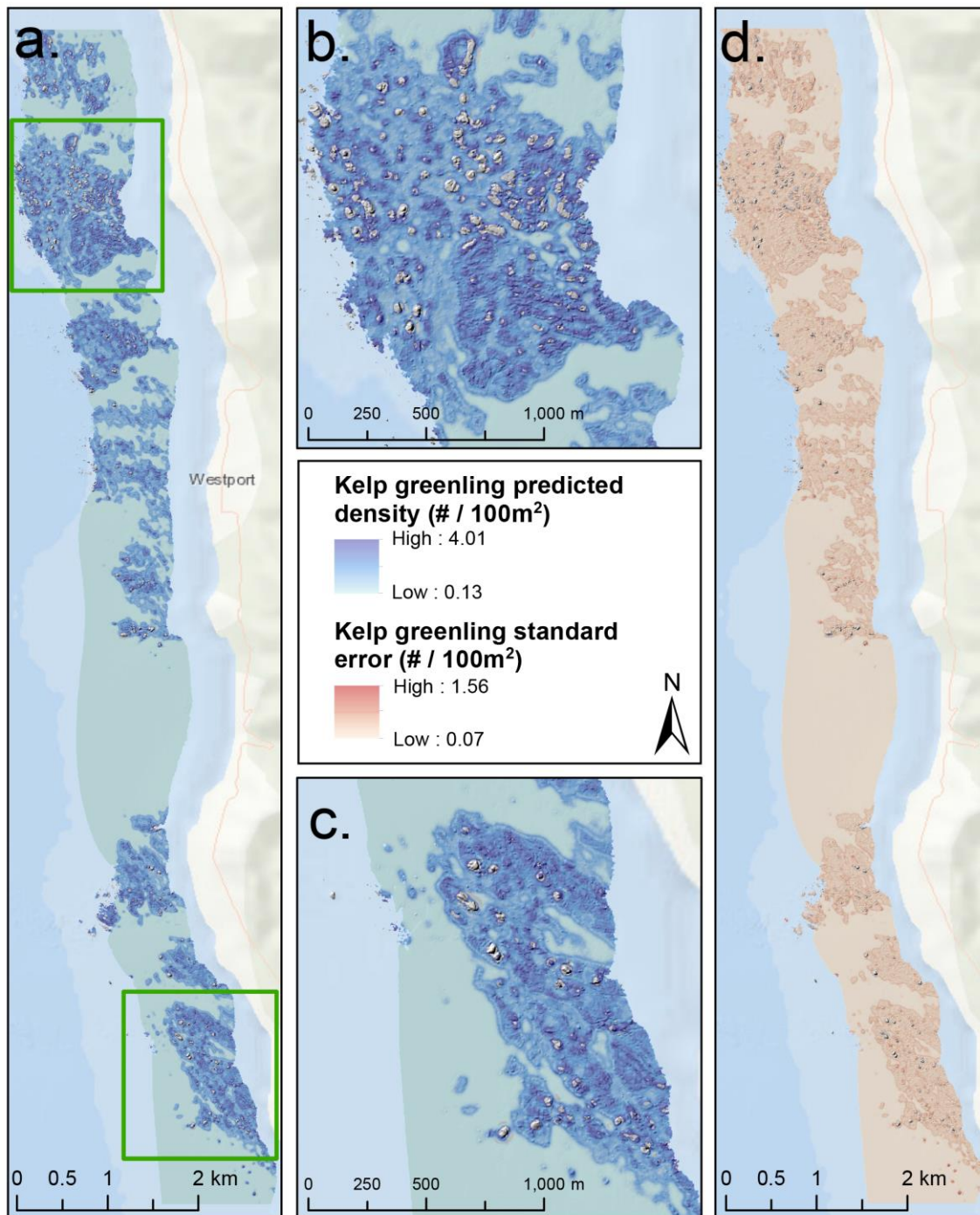


Figure 8: Predicted kelp greenling density. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the kelp greenling density model.

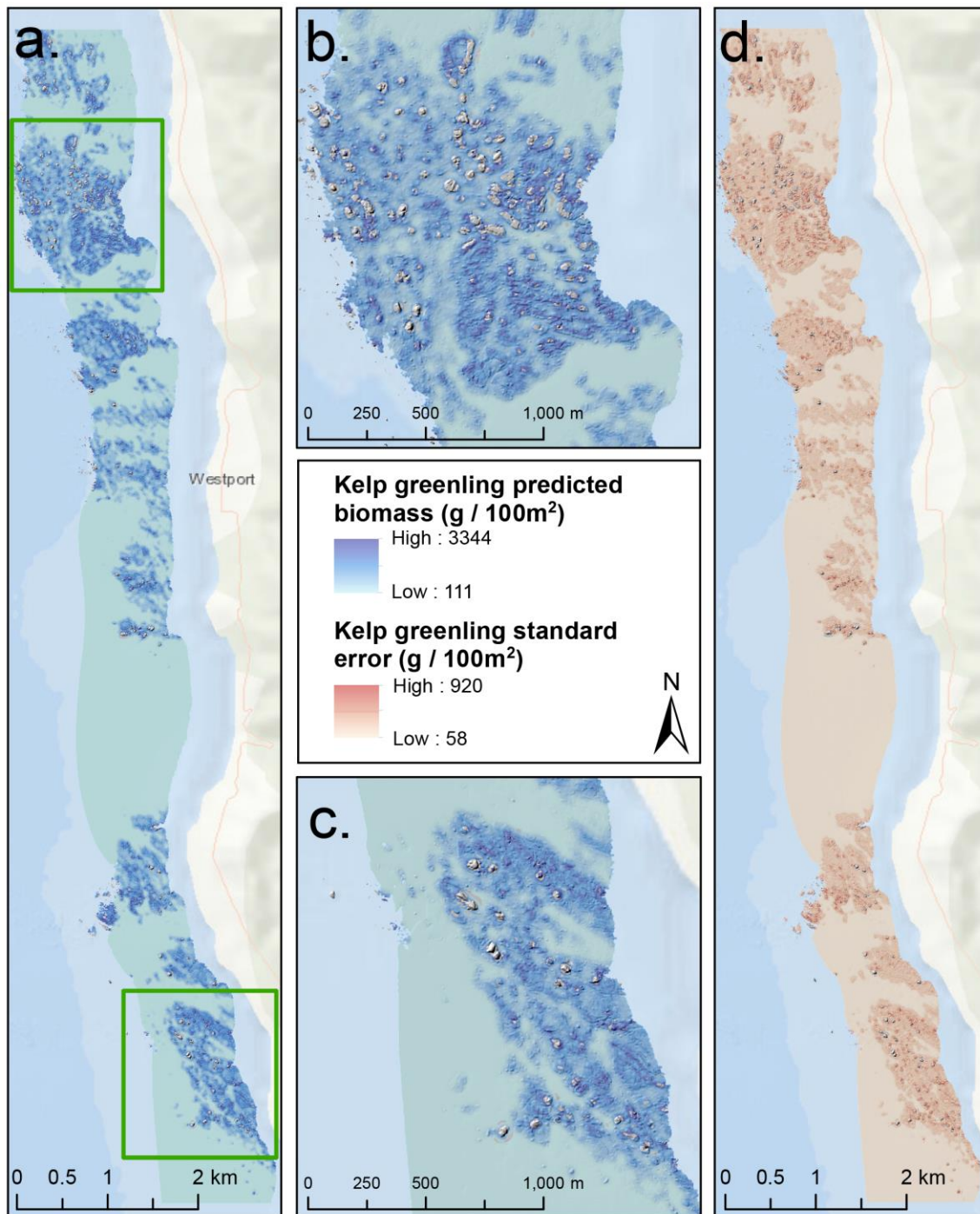


Figure 9: Predicted kelp greenling biomass. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the kelp greenling biomass model.

### Generalized additive models: Lingcod

The chosen GAM for lingcod density included TPI at both fine (20m) and broad (250m) scales as the most important predictors (Table 2). This model explained 30% of the deviance and had an adjusted  $R^2$  of 0.25. Model predictions of lingcod density had an accuracy of Spearman's  $\rho = 0.36$ . The response curve of density vs. fine-scale topographic position showed that density increased at a relatively slow rate until a TPI20 value of  $\sim 150$  at which point density increased more rapidly (Fig. 10a), suggesting that lingcod density increases were greatest in habitats characterized by peaks and ridges rather than valleys at the 20m scale. This sharp increase may be misleading, however, as the standard errors on this part of the curve are substantially higher than at other TPI20 values. The curve of density and broad-scale TPI depicted a negative relationship with the highest densities of lingcod occurring at low TPI250 values, which correspond to valleys or depressions at the 250m scale (Fig. 10b).



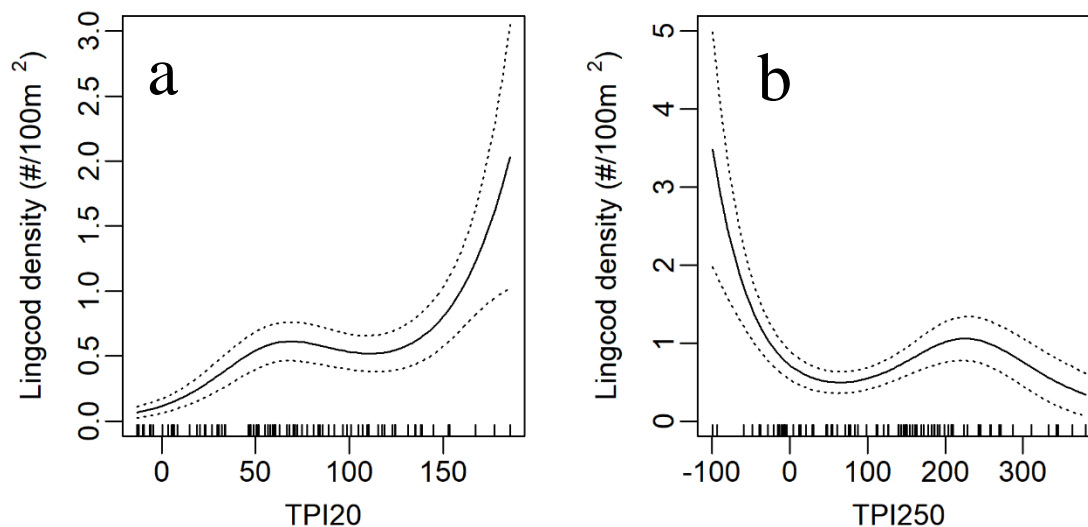


Figure 10: Generalized additive model response curves for lingcod density versus (a) fine-scale topographic position index (TPI) and (b) broad-scale topographic position index. Solid lines = mean, dashed lines =  $\pm$  SE. Rug plot along x-axis indicates observed values used to train models.

The final GAM for lingcod biomass found that VRM6, VRM498, TPI20, and TPI250 were important predictors (Table 2). The model explained 49% of the deviance and had an adjusted  $R^2$  of 0.41. Predictions based on this model had an accuracy of Spearman's  $\rho = 0.30$ . According to response curves, lingcod biomass was maximized at middling levels of fine-scale rugosity (Fig. 11a). The response curve comparing biomass with broad-scale VRM indicated that biomass increased steadily with higher VRM498 values (Fig. 11b). The curve for biomass vs fine-scale TPI was somewhat similar to the density curve with biomass showing no increase until a TPI20 value of  $\sim 150$  (Fig. 11c). Again, the standard errors surrounding this region of the curve were quite large. The response curve of lingcod biomass and broad-scale TPI showed a complex relationship with higher biomass at both relative low points (TPI250 values between -100 and 0) and



relative high points on the reef (values between 200 and 300) (Fig. 11d). Biomass was low at both moderate TPI250 values and values over 300.

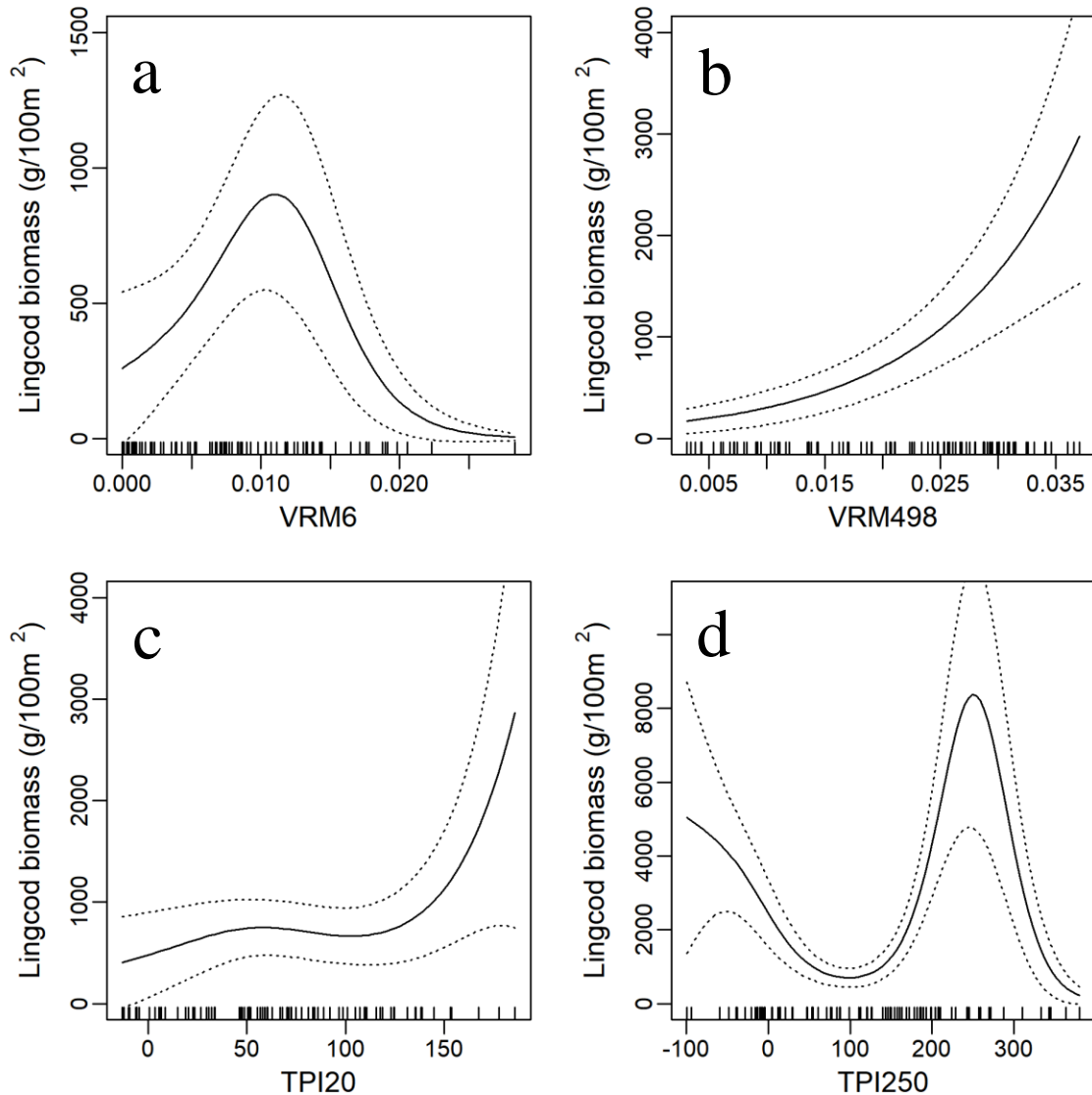


Figure 11: Generalized additive model response curves for lingcod density versus (a) fine-scale rugosity (Vector Ruggedness Measure, VRM), (b) broad-scale rugosity, (c) fine-scale topographic position index (TPI), and (d) broad-scale topographic position index. Solid lines = mean, dashed lines =  $\pm$  SE. Rug plot along x-axis indicates observed values used to train models.

Similar to that of kelp greenling, the map of predicted lingcod density was relatively patchy across the study area, however the highest densities were concentrated around several key areas (Fig. 12a). At Abalone Point, high densities were predicted at the edges of rocky habitats on the northern and southern sections of reef as well as the center of the main reef (Fig. 12b). The highest predictions in the Ten Mile SMR were along the offshore margins of the reef systems (Fig. 12c). Predictions were uniformly low in sandy, flat habitats, particularly in the large sand channels between reefs. The distribution of standard error associated with the lingcod density model largely followed the model predictions with higher predicted densities associated with higher levels of uncertainty (Fig. 12d).

The distribution of predicted lingcod biomass was more spatially concentrated than any of the previous models (Fig. 13a). These concentrated areas were centered on the northern, inshore section of the main reef and small sections of the southern reef at Abalone Point (Fig. 13b). At the Ten Mile SMR, the highest biomass predictions were located in the large reef systems at the north and south ends of the site (Fig. 13c). Again, the map of standard error revealed that the distribution of uncertainty closely followed the spatial distribution of lingcod biomass (Fig. 13d).

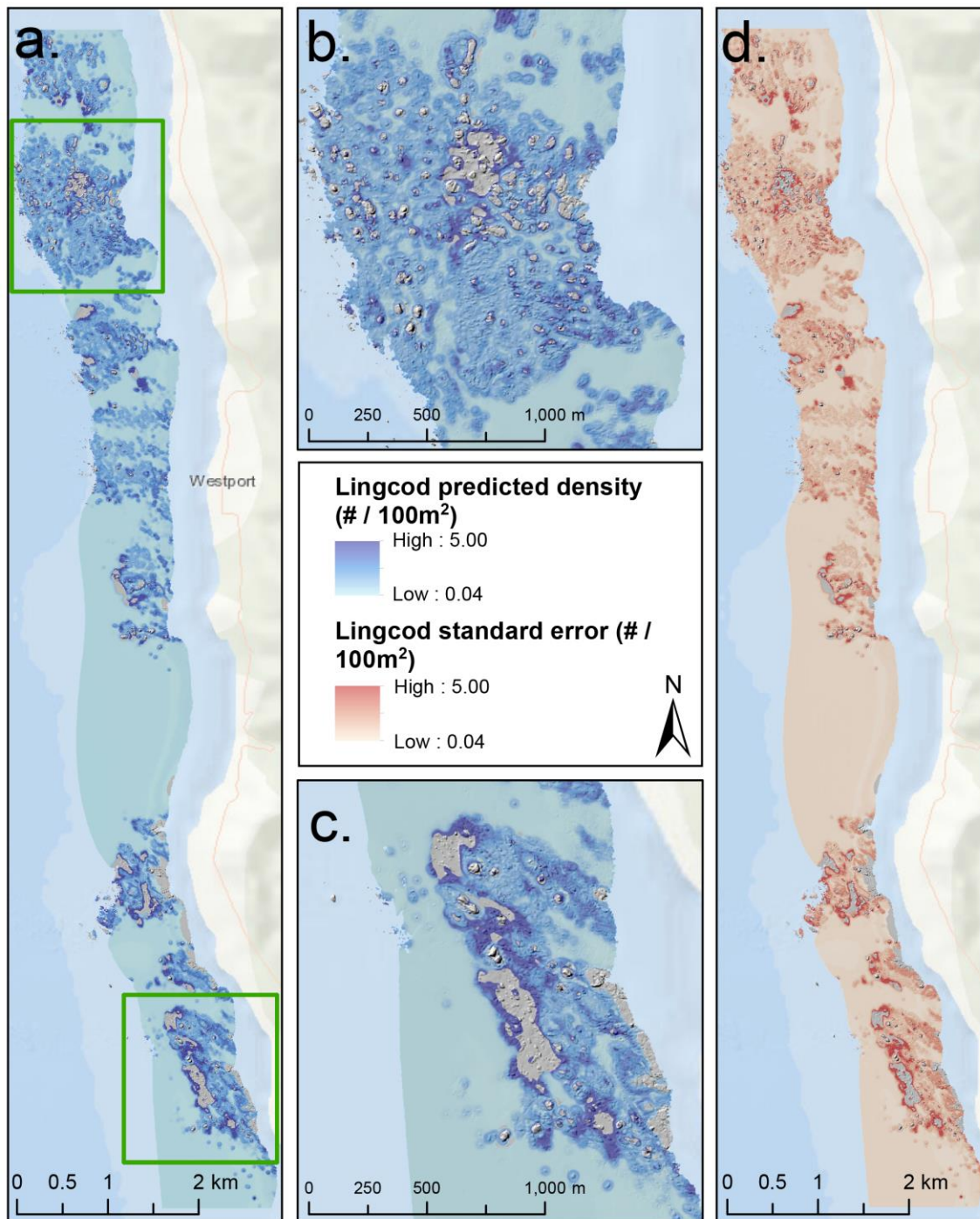


Figure 12: Predicted lingcod density. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the lingcod density model.

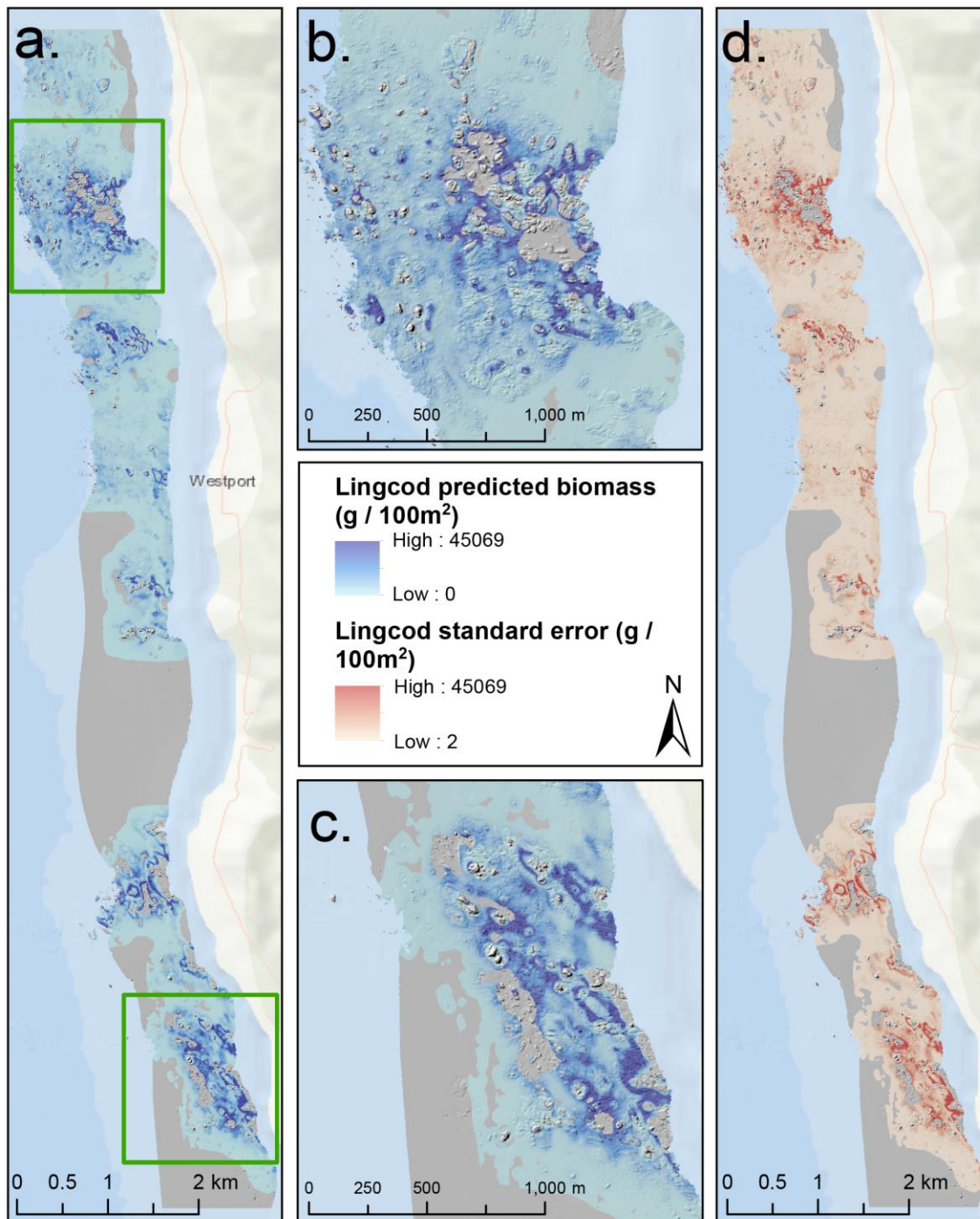


Figure 13: Predicted lingcod biomass. (a) Map of model predictions across the study area. Grey regions indicate areas where no predictions were made; (b) subset of model predictions around Abalone Point; (c) subset of model predictions in the Ten Mile State Marine Reserve; (d) distribution of standard errors associated with the lingcod biomass model.

## DISCUSSION

Previous work has shown that fishes can exhibit complex relationships with reef features in both tropical (Pittman & Brown 2011, Ortiz & Tissot 2012) and temperate (Love & Yoklavich 2008, Anderson et al. 2009) environments. In this study, we demonstrate that both kelp greenling and lingcod exhibit similarly complex associations with bathymetry-derived habitat characteristics. Additionally, we show that the nature of these relationships is often dependent on the spatial scale of the habitat features.

Our results indicate that kelp greenling density and biomass are driven by the amount of hard substrate and fine-scale topographic position. Lingcod density was primarily associated with both fine- and broad-scale topographic position. Lingcod biomass was driven by rugosity and topographic position at fine- and broad-scales. Furthermore, we highlight that GAMs are capable of using a combination of remotely sensed bathymetric data and SCUBA-based fish surveys to make accurate predictions of fish density and biomass into previously unsampled regions of the study area, however these models should be tested on an independent dataset in order to fully assess this accuracy.

### Predictor and response variable relationships

Fine-scale topographic position was an important predictor in each of the four final models. The relationships between topographic position and each response variable followed similar patterns with either density or biomass maximized at the highest values, indicating that both species share a preference for higher points on the reef relative to the immediate (~20 m) surroundings. For lingcod, this fits with Love's (2011) description wherein lingcod are often found sitting atop prominent reef features. While less information has been published on the specific habitat requirements of kelp greenling, they also seem to be found more often on tops of rocks rather than in cracks and crevices (Love 2011, Berger et al. 2015).

Depth was not a significant predictor in any of the final models. Previous work has shown that depth is an important predictor of fish presence, density, and biomass (Young et al. 2010, Wedding & Yoklavich 2015), however these studies surveyed a greater range of depths than the present study. Recent work also using SCUBA-based GAMs noted that depth contributed to models only when the species depth range was contained within the depth range surveyed (Young & Carr 2015). This phenomenon may be at play here as well, since the surveys in this study extended to depths of only 26 m while kelp greenling and lingcod can be found much deeper (130m and 475m, respectively). In order for the effects of depth on fish distributions to be fully realized, surveys must cover the full depth range of the species.

Models of density and biomass differed for both species either by the predictors present in the final models or the nature of the relationships between a given predictor and the response variables. The final model for lingcod biomass included rugosity at fine- and broad-scales as predictors, whereas the density model did not. This discrepancy between the two models shows that there are different factors driving high abundance and high biomass. As observations of higher biomass were driven by the presence of larger and likely older individuals, this suggests a potential ontogenetic shift wherein lingcod have different associations with habitat as they grow. Previous work has noted that lingcod tend to seek out higher relief bottom structure as they age (Miller & Geibel 1973, Love 2011). Our results support this, however the varied response of lingcod biomass to VRM at different spatial scales indicates that this relationship is more nuanced in that lingcod may prefer highly rugose habitats at broad scales and lower rugosity habitats at finer scales.

The response of lingcod to broad-scale topographic position was dependent on whether density or biomass was used as the response variable. Density was at its maximum at deep points on the reef relative to the surrounding 250m. The biomass model, however, showed a complex relationship with broad-scale topographic position with the highest levels of biomass occurring on relatively shallow peaks and ridges. The spike in biomass demonstrates that larger individuals are utilizing high points on the reef more than smaller fish. While this difference between the two lingcod models could also point to the presence of an ontogenetic shift in habitat preference, it may be due to larger

lingcod outcompeting smaller individuals for either preferable locations on the reef or access to food.

Competition in this instance assumes that either space in ideal habitats or prey items are limited and that lingcod will actively fend off other individuals for access to these locations or prey. Others have noted, however, that several lingcod will inhabit a single rock or promontory when conditions are favorable (Love 2011) which would suggest that the relatively low densities we observed (max 5 individuals / 100m<sup>2</sup>) are insufficient for space-based competition to be a strong contributor to these patterns. Furthermore, without more information on prey availability within the study area, it is not possible to determine the role of resource competition in driving these model differences.

The model of kelp greenling density showed a potential threshold effect in the response to the amount of available hard substrate. Density increased rapidly with hard substrate presence until 55% of the seafloor was hard substrate, at which point density then leveled off. The asymptotic nature of the response curve is indicative of a threshold where a certain percentage of rock must be present for density to be maximized and that greater amounts of hard substrate do not yield increased densities. Interestingly, no such threshold could be seen for kelp greenling biomass as biomass was maximized at the highest percentages of hard substrate. The effect of this can be seen in the predictive maps where hotspots of biomass tended to be further removed from reef edges than density hotspots.

Our results show that the spatial scale of predictors is an important factor when modeling fish distributions. Numerous studies have confirmed this, showing that fishes



interact with their environment at multiple spatial scales (Wedding et al. 2008, Anderson et al. 2009, Pittman & Brown 2011). Comparison between kelp greenling and lingcod models revealed that these species have varied responses to habitat features at fine and broad scales. Kelp greenling density and biomass were driven by fine scale topographic features whereas lingcod keyed into features at both fine and broad scales.

This pattern of lingcod responding to broader scale reef features is partially supported by differences in home range sizes between the two species. Kelp greenling have average home ranges between 500 and 1200 m<sup>2</sup> (corresponding to circles with 15 and 20m radii, respectively), whereas lingcod have average home ranges of 2800 m<sup>2</sup> (30m radius) (Freiwald 2009, Tolimieri et al. 2009). This difference in home ranges, however, is on the scale of tens of meters which does not account for the inclusion of reef features at the scale of hundreds of meters (i.e. VRM498 and TPI250) in the final lingcod models. One key behavioral difference that may lend some inference into this discrepancy is the fact that, while they have a high degree of site fidelity around a relatively small location of residence, lingcod frequently make short excursions away from their home range to feed (Starr et al. 2005, Tolimieri et al. 2009). Conversely, kelp greenling tend to stay within their home range for years at a time (Freiwald 2009). Thus, the differences in the spatial scale of predictor variables may be driven by differing foraging behaviors between the two species.

### Limitations of methodology

A potential pitfall of our survey methods is that two transects were conducted for each station, raising issues of pseudoreplication and lack of independence between samples (Hurlbert 1984). While we maintained a large (20m) distance between transects in order to increase sample independence and tested for spatial independence explicitly by plotting model residuals against transect coordinates, the possibility remains that the assumption of independence was violated. A more robust way to handle this survey design would be to utilize a generalized additive mixed model (GAMM) approach that incorporates station as a random effect (Zuur et al. 2009). We attempted to use GAMMs initially during the modeling phase of this study, however some important automated parameter selection procedures used in GAMs are not available for mixed effect models (Wood 2006). The alternative of hand-selecting these parameter values was deemed too subjective. Furthermore, attempts to utilize random effects resulted in models failing to converge. Future work should likewise attempt to use GAMMs insofar as a reliable set of parameter selection criteria can be devised. Alternatively, the use of a non-nested survey design would remove the issue entirely.

An important caveat of our modeling approach is that all of the predictor variables were derived from multibeam bathymetric sonar data. Therefore, our models only incorporate the physical structure of the habitat and do not explicitly include other potentially important drivers of fish density and biomass such as fishing pressure, nutrients, oceanographic conditions, and presence of biogenic habitat. Data for many of

these variables are either non-existent for the region or are only available at resolutions that are substantially coarser than the bathymetric data. Use of these data types as predictor variables when they become available could conceivably improve these models.

Reliance on these physical habitat characteristics also requires that survey locations overlap with the bathymetric maps. Due to difficulties in mapping shallow reefs from ship-based sonar systems, many of the shallowest portions of the coast have not yet been mapped. While there are specialized vessels capable of conducting bathymetric surveys in these nearshore regions, they have not been deployed in the vicinity of the study area (<http://seafloor.otterlabs.org/descriptions/kelpflydescrip.html>). As a result, no surveys were conducted within the 0-10 m depth zone, thereby limiting our ability to extend our models into these shallow areas. Many fish species along the north coast, including kelp greenling, utilize these shallow reefs, particularly where there is high algal and surfgrass cover (Love 2011). Incorporating shallow habitats into these models would provide valuable information on the distribution and habitat requirements of these species.

In order to effectively capture relationships between fishes and features of their habitat within an SDM framework, it is vital that the geographic locations at which fishes are observed are reproduced accurately with respect to the bathymetric maps used for modeling. Uncertainty of fish locations limits the scales of analysis (Young & Carr 2015) which could mask habitat associations at finer scales. Analysis of our data indicated a strong correlation between the DEM and *in situ* depth measurements at both the centerpoint of the station and every 2 meters along the transect. This demonstrates that

the methods used here to map transect locations (i.e. using a single known geographic location for each station) were adequate for the accurate portrayal of survey locations even at fine spatial scales.

Though our results indicated a high degree of agreement, two relatively minor changes to the survey methods could further increase the accuracy of our transect mapping. First, the use of a weighted transect line, rather than the more common plastic lines, would ensure that transects conform to the seafloor more completely. Second, placing a marker buoy at the end of each transect would yield an additional two geographic locations per station, which would allow for corrections based on the discrepancy between mapped transect endpoints and their actual locations.

Differences in historical fishing pressure between Abalone Point and Ten Mile are likely minimal given the close proximity of the two study sites as well as their similarity in seafloor topography. As such, insights into the effects of fishing on the model results are not possible without high-resolution data on past and present fishing effort within the study area. Alternatively, there may be an opportunity to examine the effects of fishing on kelp greenling and lingcod habitat use by revisiting these sites after a sufficient amount of time has elapsed since the establishment of the Ten Mile SMR for potential reserve effects to be realized. This could also yield information on how increased fish density may affect habitat associations if strong reserve effects are observed.

The temporal and spatial extent of the surveys plays an important role in the interpretation of our findings. The surveys used here were conducted in summer months between 2015 and 2016 and therefore do not incorporate seasonal and long-term changes

in fish abundance and biomass. Furthermore, these two years were peculiar for the study region as they were marked by an extreme warming event which has been linked to a number of widespread ecological changes (Kintisch 2015, OPC 2017). These surveys also spanned a relatively small length of the California coastline, thereby preventing inferences into patterns of fish distributions at larger, regional scales. Young and Carr (2015) used a similar GAM framework to model fish density across much of central California, which allowed them to incorporate broad-scale habitat variables including kelp biomass and wave orbital velocity. The scale of their survey data also enabled them to make estimates of fish density across an MPA spanning roughly nine kilometers of coastline. By including long-term datasets that span a large geographic range into this modeling framework, future work could examine any potential changes in the habitat requirements of these fishes through time as well as in the context of MPA effects and broad-scale oceanographic phenomena.

It is important to note that the models used here tend to work best for benthic-oriented species in high abundance. Young and Carr (2015) suggest that low observation rates were likely responsible for their models failing to show strong relationships between environmental variables and two of the fish species included in their study, kelp greenling and tubesnout (*Aulorhynchus flavidus*). In the current study, the reduced accuracy of the models for lingcod compared to those of kelp greenling may also be due to fewer observations, as lingcod were absent on 60% of all transects whereas kelp greenling were absent on 36% of transects. While certain distribution families used in GAMs (e.g. Poisson, Tweedie) are able to handle some amount of zeroes in the data, other model

types such as zero-inflated or hurdle models may be more suitable (Zuur et al. 2009). It is possible that such an approach could reduce uncertainty at locations where predictions indicated high habitat suitability but zero individuals were observed.

The present study focused solely on adult kelp greenling and lingcod and excluded juveniles in order to account for differences in habitat preference between juveniles and adults. Kelp greenling juveniles tend to be found in the intertidal and shallow (<10m) subtidal, whereas adults inhabit a broader range of depths from the intertidal to 130m (Love 2011, Berger et al. 2015). Juvenile lingcod settle in low complexity, sandy habitats before moving to progressively more complex environments as they grow (Miller & Geibel 1973, Love 2011). Exclusion of these individuals was prudent as their habitat preferences are distinct enough as to warrant their own separate analyses. On our surveys, observations of juveniles were too sparse to conduct such separate analyses.

A total of four juvenile kelp greenling and 19 juvenile lingcod were removed prior to analysis. The four kelp greenling were found in habitats not distinctly different from their larger counterparts: rocky habitats at depths ranging between 10 and 15m. Juvenile lingcod, however, were observed more often over either soft or mixed sediments rather than the rocky, high relief areas inhabited by adults. All 19 juvenile lingcod observations were from Abalone Point, whereas none were observed on transects in the Ten Mile SMR. This may be due, at least in part, to sampling differences as soft or mixed sediments were targeted at seven Abalone Point stations compared with four stations in the Ten Mile SMR. Future studies that explicitly model juveniles could shed further light

onto the drivers of this difference as well as provide useful information about the habitat requirements of each species at this important life stage.

Extrapolation of models to novel environments is often criticized in species distribution modeling as species interactions and distributions may change when habitats or environmental variables differ (Elith & Leathwick 2009). We sought to reduce these errors when developing our predictive maps by limiting predictor variables to only those values observed during field surveys. The effect of this can be seen as “holes” in the predictive maps where the value of at least one predictor variable fell outside the observed range. While these gaps are sparse in the kelp greenling maps, they are considerably more extensive for the lingcod models. Increased sampling effort and survey designs that specifically target these extreme values could provide more complete maps of fish distributions.

### Management implications

Spatially-explicit, quantitative data on fish distributions can help inform management and conservation efforts, particularly for species associated with complex, rocky substrates. The modeling framework used in this study is an effective method for identifying habitat features that are important for demersal fishes. Delineation of these essential fish habitats is a major stated goal of the Sustainable Fisheries Act of 1996 (SFA 1996) and is a critical component of effective marine reserve siting (Ward et al. 1999). While juveniles were excluded from analysis in this study, defining the features

that constitute ideal habitats for these young individuals will be key for determining the full extent of essential fish habitats (Rosenberg et al. 2000). These models can yield valuable insights into the life histories of demersal species by examining relationships between fishes and characteristics of their habitat. As stock assessments for reef associated species are often inadequate, SDMs may also provide a reliable, cost-effective means of estimating species density and biomass at regional scales meaningful to fisheries managers (Nasby-Lucas et al. 2002).



## CONCLUSIONS

The modeling approach used in this study combined remotely-sensed habitat data with spatially-explicit, SCUBA-based fish surveys to examine the habitat requirements of two demersal fish species in northern California. Here we demonstrate that SDMs can provide useful inferences on the relationships between fishes and characteristics of their physical habitat, the spatial scales at which these relationships occur, and how these relationships affect the distribution of each species across the study area. Our results highlight the fact that both kelp greenling and lingcod have complex and varied habitat requirements and that their relationship with habitat characteristics may change as they grow. Additionally, the nature of these requirements depend, to some extent, on the spatial scale at which they are examined. In order for these species to be properly managed, efforts must consider a variety of habitat features and spatial scales. By expanding this approach using longer datasets that span greater geographic ranges, researchers can better understand the habitat requirements of demersal fish populations, help inform the ideal siting of marine reserves, and provide accurate stock estimates of important fish species.

## REFERENCES

- Adams PB, Starr RM (2001) Lingcod. In: California's Living Marine Resources: A Status Report. California Department of Fish and Wildlife, p 191–194
- Allen LG, Pondella DJ (2006) Ecological classification. In: Allen LG, Pondella DJ, Horn MH (eds) The Ecology of Marine Fishes. p 81–113
- Anderson TJ, Syms C, Roberts DA, Howard DF (2009) Multi-scale fish-habitat associations and the use of habitat surrogates to predict the organisation and abundance of deep-water fish assemblages. *J Exp Mar Bio Ecol* 379:34–42
- Anderson TJ, Yoklavich MM (2007) Multiscale habitat associations of deepwater demersal fishes off central California. *Fish Bull* 105:168–179
- Berger AM, Arnold L, Rodomsky BT (2015) Status of Kelp Greenling (*Hexagrammos decagrammus*) along the Oregon Coast in 2015. Oregon Dep Fish Wildlife, Newport, OR
- Carr MH, Woodson CB, Cheriton OM, Malone D, Margaret A, Raimondi PT (2011) Knowledge through partnerships: integrating marine protected area monitoring and ocean observing systems. *Front Ecol Environ* 9:342–350
- Cass AJ, Beamish RJ, McFarlane GA (1990) Lingcod (*Ophiodon elongatus*). *Can Spec Publ Fish Aquat Sci* 109:40
- Chatfield BS, Niel KP Van, Kendrick GA, Harvey ES (2010) Combining environmental gradients to explain and predict the structure of demersal fish distributions. *J Biogeogr* 37:593–605
- Cope JM, MacCall AD (2005) Status of Kelp Greenling (*Hexagrammos decagrammus*) in Oregon and California Waters as assessed in 2005. Pacific Fish Manag Council Portland, OR
- Crowder LB, Norse E (2008) Essential ecological insights for marine ecosystem-based management and marine spatial planning. *Mar Policy* 32:772–778
- CSMP (2010) The California seafloor mapping project: a collaborative effort. <http://seafloor.csUMB.edu/csmp/csmp.html>.
- Dunn PK, Smyth GK (2005) Dispersion Model Densities. *Stat Comput* 15:267–280

- Elith J, Leathwick JR (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu Rev Ecol Evol Syst* 40:677–697
- Freiwald J (2009) Causes and consequences of the movement of temperate reef fishes. Dissertation. University of California, Santa Cruz, CA
- Halpern BS, Walbridge S, Selkoe KA, Kappel C V, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck R, Watson R (2008) A Global Map of Human Impact on Marine Ecosystems. *Science* 319:948–952
- Hamel OS, Sethi SA, Wadsworth TF (2009) Status and Future Prospects for Lingcod in Waters off Washington, Oregon, and California as Assessed in 2009. Pacific Fish Manag Counc Portland, OR
- Hamilton SL, Caselle JE, Malone DP, Carr MH (2010) Incorporating biogeography into evaluations of the Channel Islands marine reserve network. *Proc Natl Acad Sci* 107:18272–18277
- Hijmans RJ (2016) raster: Geographic Data Analysis and Modeling. R package version 2.5-8.
- Hobson RD (1972) Surface roughness in topography: quantitative approach. In: Chorley RJ (ed) *Spatial analysis in Geomorphology*. Harper and Row, New York, New York, USA, p 221–245
- Howard DF, Silberberg KR (2001) Kelp Greenling. In: Leet WS, Dewees CM, Klingbeil R, Larson EJ (eds) *California's Living Marine Resources: A Status Report*. University of California Department of Agriculture and Natural Resources, p 183–184
- Flannery C (2018) The effects of ocean acidification and hypoxia on juvenile rockfish behavior and physiology. Thesis. Humboldt State University, Arcata, CA
- Hurlbert SH (1984) Pseudoreplication and the Design of Ecological Field Experiments. *Ecol Monogr* 54:187–211
- Iampietro PJ, Young MA, Kvitek RG (2008) Multivariate prediction of rockfish habitat suitability in Cordell Bank National Marine sanctuary and Del Monte Shalebeds, California, USA. *Mar Geod* 31:359–371
- Jackson JBC (2008) Ecological extinction and evolution in the brave new ocean. *Proc Natl Acad Sci* 105:11458–11465

- Jagiello TH, Wallace FR (2005) Assessment of lingcod (*Ophiodon elongatus*) for the Pacific Fishery Management Council. Pacific Fish Manag Counc Portland, OR
- Jenkinson RS, Craig SF (2017) Baseline monitoring of rocky reef and kelp forest habitats of the North Coast Study Region. Calif Ocean Prot Counc Sacramento, CA
- Kintisch E (2015) “The Blob” invades Pacific, flummoxing climate experts. Science 348:17–18
- Knudby A, Brenning A, LeDrew E (2010) New approaches to modelling fish-habitat relationships. Ecol Modell 221:503–511
- Lea RN, McAllister RD, VenTresca DA (1999) Biological aspects of nearshore rockfishes of the genus *Sebastes* from central California with notes on ecologically related sport fishes. Calif Dep Fish Game, Fish Bull 177
- Love MS (1996) Probably more than you want to know about the fishes of the Pacific Coast. Really Big Press, Santa Barbara, CA
- Love MS (2006) Subsistence, Commercial, and Recreational Fisheries. In: Allen LG, Pondella DJ, Horn MH (eds) The Ecology of Marine Fishes.p 567–594
- Love MS (2011) Certainly more than you want to know about the fishes of the Pacific coast. Really Big Press, Santa Barbara, CA
- Love MS, Carr MH, Haldorson LJ (1991) The ecology of substate-associated juveniles of the genus *Sebastes*. Environ Biol Fishes 30:225–243
- Love MS, Yoklavich M (2008) Habitat characteristics of juvenile cowcod, *Sebastes levis* (Scorpaenidae), in Southern California. Environ Biol Fishes 82:195–202
- Lundblad ER, Wright DJ, Miller J, Larkin EM, Rinehart R, Naar DF, Donahue BT, Anderson SM, Battista T (2006) A benthic terrain classification scheme for American Samoa. Mar Geod 29:89–111
- Lynn K (2008) Lingcod. In: Status of the Fisheries Report: An Update Through 2008. California Department of Fish and Wildlife, p 162–174
- Miller DJ, Geibel JJ (1973) Summary of blue rockfish and lingcod life histories; a reef ecology study; and giant kelp, *Macrocystis pyrifera*, experiments in Monterey Bay, California. Calif Dep Fish Game, Fish Bull 158

- Nasby-Lucas NM, Embley BW, Hixon MA, Merle SG, Tissot BN, Wright DJ (2002) Integration of submersible transect data and high-resolution multibeam sonar imagery for a habitat-based groundfish assessment of Heceta Bank, Oregon. *Fish Bull* 100:739–751
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2017) *vegan: Community Ecology Package*. R package version 2.4-3.
- OPC (2017) *State of the California North Coast: Summary of Findings from Baseline Monitoring of Marine Protected Areas, 2013–2017*. Calif Dep Fish Wildlife, Calif Ocean Sci Trust, Calif Ocean Prot Counc California, USA
- Ortiz DM, Tissot BN (2012) Evaluating ontogenetic patterns of habitat use by reef fish in relation to the effectiveness of marine protected areas in West Hawaii. *J Exp Mar Bio Ecol* 432–433:83–93
- PISCO (2016) *The Partnership for Interdisciplinary Study of Coastal Oceans*. <http://www.piscoweb.org/>.
- Pittman SJ, Brown KA (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PLoS One* 6:e20583
- Pittman SJ, Christensen JD, Caldow C, Menza C, Monaco ME (2007) Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecol Modell* 204:9–21
- Pittman SJ, Costa BM, Battista TA (2009) Using Lidar Bathymetry and Boosted Regression Trees to Predict the Diversity and Abundance of Fish and Corals. *J Coast Res* 10053:27–38
- Rees MJ, Jordan A, Price OF, Coleman MA, Davis AR (2014) Abiotic surrogates for temperate rocky reef biodiversity: Implications for marine protected areas. *Divers Distrib* 20:284–296
- Roberts CM, Andelman S, Branch G, Bustamante RH, Castilla C, Dugan J, Halpern BS, Lafferty KD, Leslie H, Mcardle D, Possingham HP, Ruckelshaus M, Warner RR, Roberts M, Lubchenco J (2003) Ecological Criteria for Evaluating Candidate Sites for Marine Reserves. *Ecol Applications* 13:199–214
- Rosenberg A, Bigford TE, Leathery S, Hill RL, Bickers K (2000) Ecosystem approaches to fishery management through essential fish habitat. *Bull Mar Sci* 66:535–542

- Sappington MJ, Longshore KM, Thompson DB (2007) Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. *J Wildl Manage* 71:1419–1426
- SFA (1996) Sustainable Fisheries Act. US Public law:104–297
- Starr RM, O’Connell V, Ralston S, Breaker L (2005) Use of Acoustic Tags to Estimate Natural Mortality, Spillover, and Movements of Lingcod (*Ophiodon elongatus*) in a Marine Reserve. *Mar Technol Soc J* 39:19–30
- Stein DL, Tissot BN, Hixon MA, Barss W (1992) Fish-habitat associations on a deep reef at the edge of the Oregon continental shelf. *Fish Bull* 90:540–551
- Stephens JS, Larson RJ, Pondella DJ (2006) Rocky Reefs and Kelp Beds. In: Allen LG, Pondella DJ, Horn MH (eds) *The Ecology of Marine Fishes*.p 227–252
- Ter Braak CJF (1986) Canonical Correspondence Analysis : A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology* 67:1167–1179
- Tolimieri N, Andrews K, Williams G, Katz S, Levin PS (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
- Ward TJ, Vanderklift MA, Nicholls AO, Kenchington RA (1999) Selecting Marine Reserves Using Habitats and Species Assemblages as Surrogates for Biological Diversity. *Ecol Appl* 9:691–698
- Wedding LM, Friedlander AM, McGranaghan M, Yost RS, Monaco ME (2008) Using bathymetric lidar to define nearshore benthic habitat complexity: Implications for management of reef fish assemblages in Hawaii. *Remote Sens Environ* 112:4159–4165
- Wedding L, Yoklavich MM (2015) Habitat-based predictive mapping of rockfish density and biomass off the central California coast. *Mar Ecol Prog Ser* 540:235–250
- Weiss AD (2001) Topographic position and landforms analysis. In: Poster presentation, ESRI User Conference.p July 9-13
- Wiens JA (1989) Spatial Scaling in Ecology. *Funct Ecol* 3:385–397
- Wood SN (2006) *Generalized Additive Models: An introduction with R*. Chapman and Hall/CRC, New York, NY

- Young M, Carr MH (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
- Young MA, Iampietro PJ, Kvitek RG, Garza CD (2010) Multivariate bathymetry-derived generalized linear model accurately predicts rockfish distribution on Cordell Bank, California, USA. *Mar Ecol Prog Ser* 415:247–261
- Zimmermann M (2003) Calculation of untrawlable areas within the boundaries of a bottom trawl survey. *Can J Fish Aquat Sci* 60:657–669
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York, New York

## APPENDIX A

Appendix A. Backwards stepwise model selection tables used to select final models for each of the four response variables: (A) Kelp greenling density, (B) Kelp greenling biomass, (C) Lingcod density, and (D) Lingcod biomass.

## A.

<b>Formula</b>	<b>R-sq. (adj.)</b>	<b>Deviance Explained (%)</b>	<b>AIC</b>
Kelp greenling density ~ Depth + SUB <sup>2</sup> + VRM6 + VRM498 + TPI20 + TPI250	.337	43.8	294.1895
Kelp greenling density ~ Depth + SUB <sup>2</sup> + VRM498 + TPI20 + TPI250	.319	40.7	292.6760
Kelp greenling density ~ Depth + SUB <sup>2</sup> + TPI20 + TPI250	.326	40.3	291.2203
Kelp greenling density ~ Depth + SUB <sup>2</sup> + TPI20	.319	38.7	291.8041
<b>Kelp greenling density ~ SUB<sup>2</sup> + TPI20</b>	<b>.31</b>	<b>36.9</b>	<b>290.5798</b>
Kelp greenling density ~ SUB <sup>2</sup>	.244	33.1	295.5677



**B.**

<b>Formula</b>	<b>R-sq. (adj.)</b>	<b>Deviance Explained (%)</b>	<b>AIC</b>
Kelp greenling biomass ~ Depth + SUB <sup>2</sup> + VRM6 + VRM498 + TPI20 + TPI250	.27	38.4	1090.957
Kelp greenling biomass ~ Depth + SUB <sup>2</sup> + VRM498 + TPI20 + TPI250	.279	38.3	1089.046
Kelp greenling biomass ~ Depth + SUB <sup>2</sup> + VRM498 + TPI20	.274	36.6	1088.708
Kelp greenling biomass ~ SUB <sup>2</sup> + VRM498 + TPI20	.26	34.9	1087.392
<b>Kelp greenling biomass ~ SUB<sup>2</sup> + TPI20</b>	<b>.254</b>	<b>34.4</b>	<b>1086.941</b>
Kelp greenling biomass ~ SUB <sup>2</sup>	.205	32.2	1088.733

## C.

Formula	R-sq. (adj.)	Deviance Explained (%)	AIC
Lingcod density ~ Depth + SUB <sup>2</sup> + VRM6 + VRM498 + TPI20 + TPI250	.313	37.0	186.2230
Lingcod density ~ Depth + VRM6 + VRM498 + TPI20 + TPI250	.315	37.1	184.5080
Lingcod density ~ Depth + VRM6 + TPI20 + TPI250	.317	36.3	183.9734
Lingcod density ~ Depth + TPI20 + TPI250	.282	33.0	183.9876
<b>Lingcod density ~ TPI20 + TPI250</b>	<b>.254</b>	<b>30.1</b>	<b>183.8107</b>
Lingcod density ~ TPI20	.127	20.0	191.0412

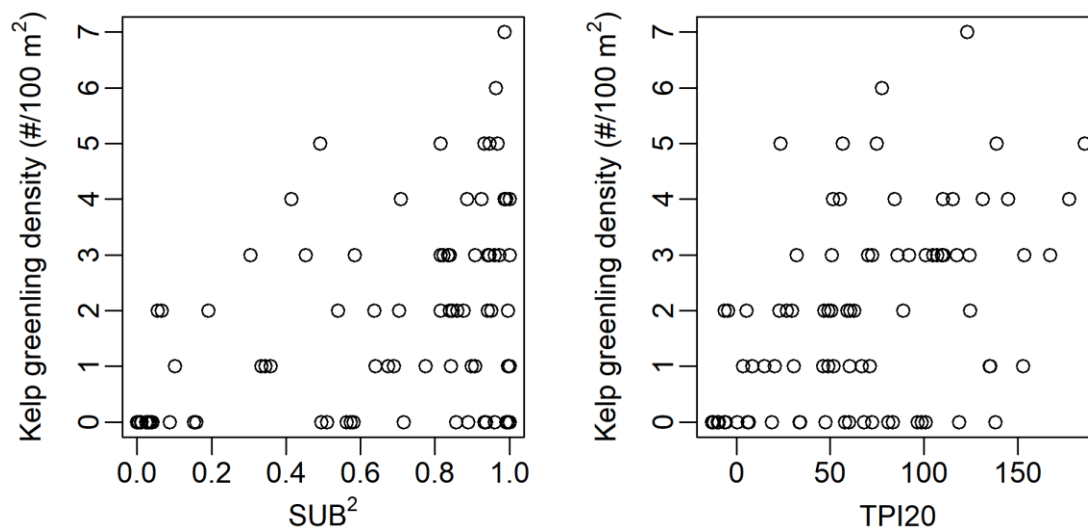
**D.**

<b>Formula</b>	<b>R-sq. (adj.)</b>	<b>Deviance Explained (%)</b>	<b>AIC</b>
Lingcod biomass ~ Depth + SUB <sup>2</sup> + VRM6 + VRM498 + TPI20 + TPI250	.408	60.5	816.2784
Lingcod biomass ~ Depth + VRM6 + VRM498 + TPI20 + TPI250	.422	51.1	813.7402
<b>Lingcod biomass ~ VRM6 + VRM498 + TPI20 + TPI250</b>	<b>.409</b>	<b>48.9</b>	<b>812.6507</b>
Lingcod biomass ~ VRM6 + VRM498 + TPI250	.379	43.9	814.3767
Lingcod biomass ~ VRM6 + TPI250	.229	37.2	821.5283
Lingcod biomass ~ VRM6	.063	21.4	831.4932

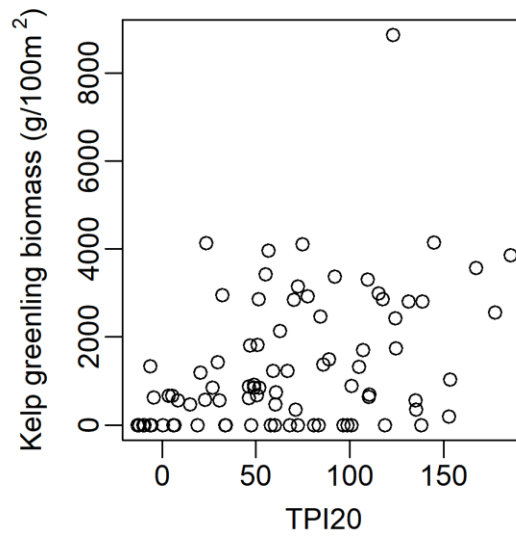
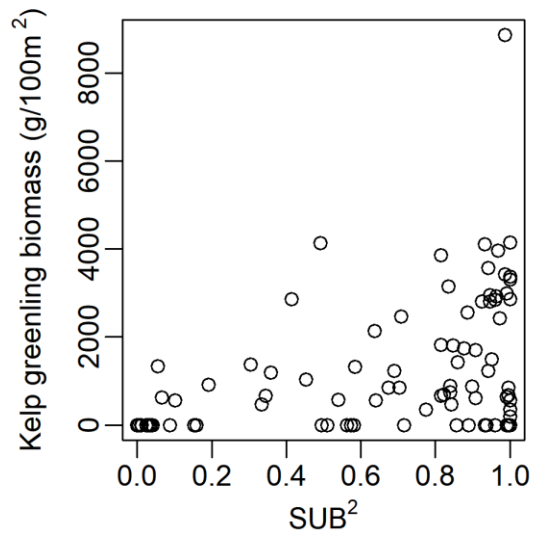
## APPENDIX B

Appendix B. Scatterplots of each of the four response variables: (A) Kelp greenling density, (B) Kelp greenling biomass, (C) Lingcod density, and (D) Lingcod biomass versus each predictor variable included in the final models.

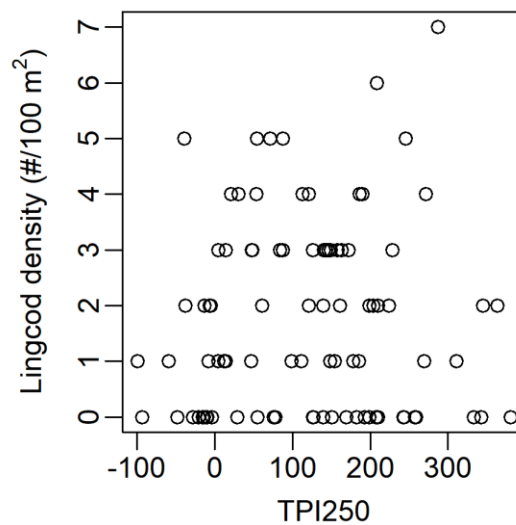
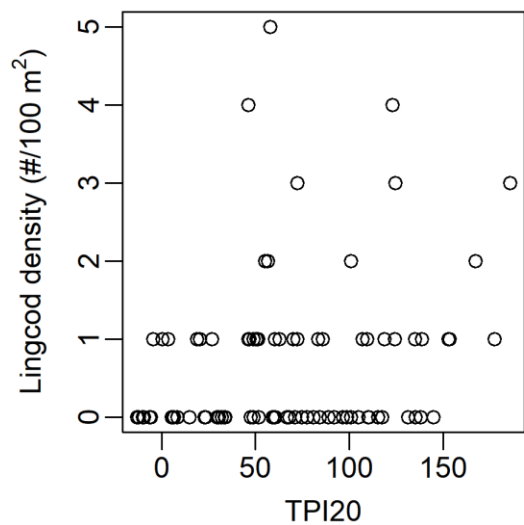
A.



**B.**



**C.**



D.

