UNDERSTANDING THE HABITAT ASSOCAITIONS OF NEARSHORE ROCKFISHES INSIDE AND OUTSIDE OF MARINE PROTECTED AREAS USING HOOK-AND-LINE DATA

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Biological Science

by

Kaila Raeanne Fritch

June 2025

COMMITTEE MEMBERSHIP

TITLE:	Understanding the Habitat Associations of Nearshore Rockfishes Inside and Outside of Marine Protected Areas Using Hook-and-Line Data
AUTHOR:	Kaila Raeanne Fritch
DATE SUBMITTED:	June 2025
COMMITTEE CO-CHAIR:	Benjamin Ruttenberg, PhD Director of Center for Coastal Marine Sciences Professor of Biology
COMMITTEE CO-CHAIR:	Dean Wendt, PhD Dean of Bailey College of Science and Mathematics Professor of Biology
COMMITTEE MEMBER:	Melissa Monk, PhD Research Mathematic Statistician NOAA Fisheries

ABSTRACT

Understanding the Habitat Associations of Nearshore Rockfishes Inside and Outside of

Marine Protected Areas Using Hook-and-Line Data

Kaila Raeanne Fritch

Habitat is an important driver in shaping the distribution of a range of species. For demersal fish species, benthic habitat characteristics can drive distribution and community structure. Understanding the relationship between habitat and species distribution in important, especially for exploited species such as rockfishes (*Sebastes spp.*). Increased data on the relationship between rockfish and benthic habitat characteristics could enhance marine protected areas (MPAs) and management practices which are used to protect rockfishes along the U.S. West Coast. To address this, we used hook-and-line, catch-and-release data that were collected inside two MPAs on the central coast of California, and two nearby reference sites that are open to fishing. We overlayed this catch data onto seafloor habitat data from the Seafloor Mapping Lab at California State University, Monterey Bay to create species distribution models for Vermilion (*Sebastes miniatus*), Copper (*S. caurinus*) and Olive Rockfish (*S. serranoides*). We found bathymetry and site to be an important predictor for Vermilion, Copper, and Olive rockfish, and substrate to be important for Copper and Olive rockfish. Bathymetry was also a spatially varying coefficient for Vermilion and Copper rockfish, while substrate was spatially varying for Olive rockfish. Understanding these habitat preferences for nearshore groundfishes could allow scientists to better predict species assemblages, understand how habitat influences species distribution and better design MPAs to manage species and conserve marine resources.

Keywords: Marine Protected Area, Rockfish, Species Distribution, Habitat Associations

ACKNOWLEDGMENTS

This work was supported by the Ocean Protection Council, NOAA Fisheries and the CSU Council on Ocean Affairs, Science & Technology (COAST) – thank you. I would like to thank my family and friends that supported me throughout my academic career. My parents, Tom and Denise Fritch, my brother Noah Fritch and sister Aubree Fritch, for their continued encouragement and help. My grandparents Pat and Dick Fritch and Susan and Pat Darrah for making this all possible. Alessandra Gardner and Landon Lujan, who pushed me to keep going when I wanted to quit and listened to all the highs and lows of graduate school. Erin Johnston who is a great friend and an amazing mentor. She supported me every step of the way, encouraged me when I needed it, picked me up when I made mistakes and celebrated when I succeeded – thank you this work would not have been possible without you. In addition, I would like to thank my committee, Dean Wendt, Ben Ruttenberg and Melissa Monk for their continued guidance and support. My mentors, Eric Ward, for all the coding help, and Grant Waltz for getting me started in the DEW lab and opening the door to this opportunity. Thank you to all of the DEW lab staff and undergraduates who have and continue to collect and enter the data that goes into this project. In particular, Erin Schneider who spent hours sorting through old datasheets with me. Finally, I would like to thank all of the volunteer anglers, past, present and future, that have spent countless hours fishing for us and collecting this data. As well as Virg's Landing, Morro Bay Landing and Patriot Sportfishing for allowing us to conduct our research with them and their continued knowledge and expertise. I would especially like to thank the captains and crew of the F/V Patriot, F/V Fiesta, F/V Flying Fish, F/V Rita G., and F/V Avenger.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Chapter 1. Understanding the Habitat Associations of Nearshore Rockfishes Inside and Outside of Marine Protected Areas Using Hook-and-Line Data.

1.1. INTRODUCTION

Habitats are critical to shaping species distributions in all ecosystems (Odum 1963). In many marine systems, benthic composition and other habitat variables can drive demersal fish distributions and community structure (Anderson et al. 2009; Bozec et al. 2005; Chittaro, 2004). Seafloor complexity (Pittman et al. 2007), depth (Anderson et al. 2009; Friedlander et al. 2007) and substrate type (Love et al. 2009) are all variables commonly used to study the relationship between habitat and fish distributions. These variables influence fish distributions differently for every fish species across a range of habitats and depth zones (Love et al. 2009). The relationship between these many habitat variables and species abundance and distributions are often complex and vary even between closely related congenerics [or species]. The fine-scale differences in the distributions of fish within the same genera may result in part from differences in life history traits such as feeding preferences (Bozec et al. 2005), site fidelity (Hannah & Rankin 2011), recruitment, spawning (Friedlander et al. 2007), ontogenetic movements (Love et al. 2009), growth, and many other factors that impact population dynamics. Because these habitat associations can have profound impacts on species abundance across a range of ontogenetic stages, understanding species-specific habitat associations, and the factors that influence those associations, are important. This is particularly true for fisheries management, where a better understanding of species-specific habitat associations could help inform species-specific management practices.

On the West Coast of the United States there are 61 species of managed rockfishes (Sebastes spp), many of which are important to both commercial and recreational fisheries in California (Stephens et al. 2006). These rockfishes have diverse life history traits and occupy a wide range of complex habitats. Rockfishes can be found from the intertidal zone to over 700 meters in depth (Love et al. 1990), and have lifespans that can range from 11 years to over 200 years (Love et al. 2002). Additionally, these species vary greatly in abundances and distributions, making them a complex group to manage. Having a better understanding of how habitat influences species specific distributions improves our ability to determine areas of essential fish habitat and estimate habitat availability for individual species, information that is critical for fisheries management (Anderson & Yoklavich 2007).

While rockfishes use a diversity of habitats, many important commercial and recreational species (are associated with specific benthic habitat characteristics, a number of which are quantifiable from seafloor mapping efforts. These include bathymetry, substrate type, substrate complexity, and the structural composition of nearby habitats (Anderson & Yoklavich 2007; Carr 1991; Caselle et al. 2002). Additionally, not all hard habitat provides the same ecosystem function, and some rockfish species associate with mixed substrate over hard substrate alone (Anderson & Yoklavich 2007; Love & York 2006). Certain seafloor characteristics, such as habitat structure and depth, can predict species distributions and assemblages (Love et al. 2022; Firedlander et al. 2007; Pittman et al. 2007; Love & York 2006), which improve our understanding of rockfish population dynamics and fisheries models used to manage those species.

In the late 20th century, declining populations of rockfish species led to more intensive fisheries management and additional conservation measures to restore and recover these populations, including the California Marine Life Protection Act of 1999

(MLPA). The MLPA stipulated the implementation of a network of marine protected areas (MPAs) off the coast of California, designed to protect marine ecosystems and essential habitats, as well as rebuild declining marine life populations (Gleason et al. 2013). There is a rich literature demonstrating that MPAs–and no-take marine reserves in particular–generally lead to increases in fish size, fish abundance and species richness (Ziegler et al. 2024; Hamilton et al. 2021; Hamilton et al. 2010, Chirico et al. 2017; Guidetti et al. 2014). However, there has been very little work examining how MPAs and the reduction of fishing pressure they create, affect the way fish utilize habitat and whether the relationship between fish abundance and habitat differs with the removal of fishing pressure inside of MPAs. MPAs have become a critical part of many fisheries management strategies, making it important to have an understanding of how they influence fish habitat associations. Having a better understanding of fish distribution amongst habitat inside and outside MPAs is the best way to develop better informed management solutions, such MPAs that consider the habitat requirements of specific species (Friedlander et al. 2007). This could potentially increase the effectiveness and efficiency of MPAs as a management tool, by allowing for the designation of habitat-specific MPAs to incorporate habitat that is ecologically important to the managed species (Friedlander et al. 2007; Lindholm et al. 2001). Rockfishes in particular will benefit from a better understanding of species-specific habitat usage to enhance spatially-informed management practices.

Fortunately, there are several long-term monitoring programs designed to evaluate the impacts of MPAs in California. One of these is the California Collaborative Fisheries Research Program (CCFRP). CCFRP is a hook-and-line, catch-and-release survey program that focuses on tracking the abundance of recreational fish species inside and outside of many MPAs across California (Hamilton et al. 2021; Wendt & Starr 2009). The CCFRP program captures a wide range of recreationally targeted species, the vast majority of which are rockfishes. In Central California, three of the most commonly caught species by recreational and commercial fisheries and the CCFRP program are Vermilion (S. miniatus), Copper (S. caurinus) and Olive rockfish (S. serranoides), all of which are generally associated with benthic habitat features. All three of these species are economically and socially important and thus would benefit from additional data regarding habitat associations to help inform effective management strategies (Cope et al., 2015; Love, 2011; Love et al., 1998; Monk et al., 2021).

The CCFRP dataset presents a unique opportunity to use hook-and-line data to model the habitat associations via species distribution models. Most previous work regarding rockfish habitat associations has been done using submersibles, remote operated vehicles or on SCUBA (Love et al. 2022; Love et al. 2009; Caselle et al. 2002; Yoklavich et al. 2002; Yoklavich et al. 2000; ect.) However, using hook and line data to create species distribution models may be a valuable way to increase the availability of habitat usage information, particularly for nearshore rockfish because of the ease and relative low cost of obtaining this data. Understanding these habitat associations could improve our ability to determine areas of essential fish habitat, estimate habitat availability, and elucidate how reduction in fishing pressure in MPAs influence the distribution of rockfish species (Anderson & Yoklavich, 2007). To address this, we created species distribution models to determine how seafloor characteristics influence the distribution of Vermilion, Copper, and Olive rockfish across the two MPAs and nearby reference sites that CCFRP surveys on the central coast of California.

1.2. Methods

1.2.1. Fisheries Data

We used data from the California Polytechnic State University, San Luis Obispo (Cal Poly SLO) CCFRP. CCFRP is a collaboration between scientists and recreational anglers, designed to collect data on marine fishes using standardized hook-and-line surveys. This catch and release survey monitors the effects of MPAs throughout California by collecting rockfish length and abundance data from inside and outside of MPAs (Hamilton et al. 2021; Wendt & Starr 2009). The Cal Poly SLO chapter of CCFRP is one of six institutions that survey 12 MPAs across the coast of California. Every summer, each institution conducts six trips in each of the areas they survey, and three at each site (MPA or reference site), within that area. Reference sites (REF) are areas adjacent to the MPAs that are open to fishing, represent similar habitat as within the nearby MPA, and used as a control. The CCFRP sampling design stratifies each site into a series of 500m x 500m grid cells (Figure 1). On each sampling day, four grid cells in a site are randomly selected and each cell is fished three times in fifteen-minute intervals, called drifts, during which the boat is not in gear and is freely drifting. Before each new drift, captains are instructed to locate suitable rockfish habitats using their knowledge of rockfish catchability to maximize catch. Within a drift, the number of volunteer anglers (between nine and twelve), the time fishing (fifteen minutes), the start and end location (GPS coordinates), depth, seafloor relief, and other environmental conditions are recorded. Every fish that is caught by a volunteer angler is brought on board, identified to the species, measured to the nearest millimeter, checked for any outstanding body damage, and then released/descended. GPS coordinates, called waypoints, corresponding with a caught fish are taken opportunistically throughout the drift. The Cal Poly SLO chapter of CCFRP surveys two MPAs, Point Buchon State Marine Reserve (SMR; a type of MPA where no commercial or recreational take is allowed) and Piedras Blancas SMR, as well as the associated REFs (Figure 1). These two areas have historically experienced different levels of fishing pressure since the implementation of the MPAs (Ivens-Duran, 2014). Point Buchon has been more heavily targeted compared to Piedras Blancas due to its location between two fishing ports and relatively short commute to the fishing grounds. This historical and current fishing effort is the most influential factor in driving variation in rockfish biomass response between MPAs and areas open to fishing (Ziegler et al., 2022). CCFRP has sampled the Central Coast of California using these methods since 2007. However, this project focused on data collected from 2018-2022, which had more intermediate waypoints than data taken prior to 2018 and therefore provided higher spatial resolution. Because waypoints are taken opportunistically, most individual fish brought on board do not have unique locations associated with them. However, we estimated the position of fish without waypoints as follows. First, fish were recorded on the data sheets in the order they were caught. Second, we assumed the boat was drifting at a constant rate between starting waypoints, intermediate waypoints, and ending waypoints. Third, we assumed that fish were caught evenly along the drift path of the boat. With this information, we interpolated locations for all fish caught between the starting waypoint and first intermediate waypoint, between all intermediate waypoints, and between the last intermediate waypoint and the ending waypoint. This meant each fish was then placed along the drift path and assigned a GPS coordinate that was evenly spaced between the start or end coordinates and the waypoints taken intermittently throughout the drift. While this approach relies on several assumptions about the location of individual fish, it was the best way to get spatial information for each caught fish using the available data. To reduce the error associated with this method, we checked the data to ensure the fish were entered in the database in the same order as the original field datasheets. We excluded a drift from the analysis if there were no intermediate waypoints associated with caught fish, or if it was not possible to determine the original order of the field datasheets. In addition, if the first or last fish recorded in a drift did not have a waypoint associated with it, we assigned the start or end coordinates respectively, so there was a start and end coordinate between which to divide the fish. These data cleaning steps were done using R (R Core Development Team 2023), tidyverse (Wickham et al. 2019) and zoo (Zeileis & Grothendieck 2005).

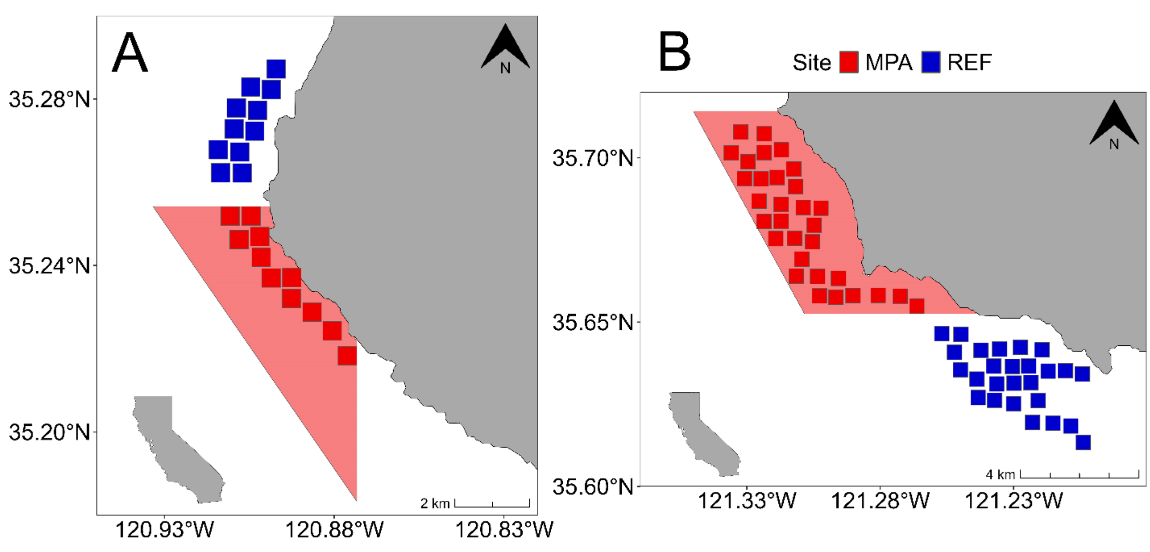

Figure 1. Map of the study region in San Luis Obispo County, California. Showing the California Collaborative Fisheries Research Program cells of the (A) Point Buchon region and the (B) Piedras Blancas region with the state marine reserve outlined in red and the marine protected area (MPA) cells in red and the reference area (REF) cells in blue.

The coordinates assigned to each fish are approximations for several reasons. First, interpolating points necessarily introduces some unknown error. Second, there is a lag in the time from which a fish is hooked to when it is brought up to the boat and measured, and if multiple fish are brought up at the same time, they would have been assigned slightly different locations. In order to account for this approximation, we calculated the average distance the vessel moved during the average amount of time of a drift (5 meters), and used this value to create a buffer, a circular polygon that extends the area of each point to a specified distance, around each fish. This step was done in ArcGIS Pro (vers 3.1.2).

1.2.2. Focal Species

We chose to focus our modeling efforts on three species commonly caught on the Central Coast of California by recreational and commercial fisheries and the CCFRP program and generally associate with benthic variables. Vermilion rockfish are most commonly found in depths of 50 to 150 meters but typically reside in the shallower part of their depth range in areas north of Point Conception (Love et al. 2002). This species can live upwards of 60 years, and tends to move deeper in the water with age (Love et al. 2002). Copper rockfish are found in subtidal waters up to 183 meters in depth, and are known to recruit to nearshore habitat sooner than other species (Love et al. 2002). They are thought to be relatively resident with small home ranges, and tend to reside near the seafloor (Love et al. 2002). Olive rockfish are a shallower species, ranging in depths from subtidal to 172 meters and tend to reside in the water column as well as near the seafloor (Love et al. 2002). Olive rockfish are a more mobile species and have been recaptured up to 510 km away from where they were tagged (Hanan & Curry 2012).

1.2.3. Habitat Characteristics

We obtained all seafloor habitat data for this project from the Seafloor Mapping Laboratory at California State University, Monterey Bay. The data included bathymetry and derived products including slope, vector ruggedness measure (VRM) and substrate. Slope and VRM are both measures of habitat complexity, which can be important predictors of occurrence for some rockfishes (Wedding & Yoklavich, 2015; Young et al., 2010). Bathymetry (depth) is also an important predictor of rockfish species distributions and abundance patterns (Williams & Ralston 2002). Rockfish species have a variety of depth distributions ranging from nearshore environments to deep sea (200+ m) (Love et al. 2002). Substrate is a binomial variable, where one means hard substrate and zero means soft substrate. We included substrate because substrate composition can be an important predictor of fish density (Anderson & Yoklavich, 2007). We projected each raster to the WGS 1984 UTM Zone 10N coordinate system to match the coordinate system of the fish locations, and merged rasters of the same habitat variable. Data processing for habitat data utilized the Raster Toolset in ArcGIS Pro (vers 3.1.2).

To determine the habitat characteristics at each of the caught fish locations, we took the mean value of each continuous habitat variable within the 5 meter buffer around each caught fish, substrate became the proportion of hard versus soft where values closer to one indicate more hard substrate and values closer to zero indicate soft substrate. Other summary statistics were also considered, including the minimum and maximum value, but we ultimately decided that the mean best represented habitat features in the 5 meter buffer.

1.2.4. Species Distribution Modeling

To determine how habitat characteristics, such as slope, bathymetry, substrate, and VRM influence the distribution of Vermilion, Copper, and Olive rockfish, we created a series of spatial models extending generalized mixed effects models (GLMMs), implemented in the R package sdmTMB (Anderson et al., 2024). Because the CCFRP represents presence only data (catches are recorded, but effort and associated metadata when fish are not caught is not), we used a point process model with downweighted Poisson regression (DWPR; Renner et al., 2015). The DWPR method works by weighting the presence points (locations of catches in CCFRP) by a very small number (1×10^{-6}), while the quadrature weights are set to the area of the study region divided by the number of quadrature points (Renner et al., 2015).

We did this to allow for the intercept term and the log-likelihood to be estimable, and to create a way to choose how many quadrature points are necessary to ensure model convergence (Renner et al., 2015). This method was implemented using the R package sdmTMB (Anderson et al., 2024) which combines spatial approximation using the stochastic partial differential equation approach (Lindgren et al. 2011) with fast marginal maximum likelihood estimation in Template Model Builder (Kristensen et al. 2016). The structure of the spatial GLMM can be represented as

$$\boldsymbol{\mu}_s = f^{-1}(\mathrm{X}^{main}\boldsymbol{\beta} + \mathbf{X}^{SVC}\boldsymbol{\zeta} + \boldsymbol{\omega})$$

Where $\boldsymbol{\mu}_\mathbf{s}$ represents a vector of predicted occurrences at locations $\mathbf{s}$, $f^{-1}()$ is the inverse link function, here it is a log link. The variable $\mathbf{X}^{main}$ represents a matrix of main fixed-effects coefficients (bathymetry, substrate, slope, VRM), with estimated coefficient vector $\boldsymbol{\beta}$. The spatially varying coefficient at a point in space is represented by $\boldsymbol{\zeta} \sim$ MVN($\mathbf{0}$, Σ_ζ), and $\mathbf{X}^{SVC}$ represents the matrix of the spatially varying coefficients. Spatially varying coefficients represents a flexible approach that allows the effect of each habitat value (bathymetry, substrate, etc) to vary spatially (Barnett et al., Thorson et al. 2023). The spatial random field $\boldsymbol{\omega}$ is included to represent processes not explained by our covariates, and is included as a vector of random effects $\boldsymbol{\omega} \sim$ MVN($\mathbf{0}$, Σ_ω). In SPDE approach, the covariance matrix Σ_ω uses a Matérn covariance function with estimated spatial range and variance. We estimated the spatial random effect using a triangular mesh with a cutoff distance set for each species (Lindgren et al. 2011); one kilometer for Vermilion and Copper rockfish and four kilometers for Olive rockfish.

1.2.5. Quadrature Points

To create quadrature points for each species, we generated a 12.5-meter buffer around each of the drift lines over which the boat drifted to create the survey area. We chose a buffer of 12.5 meters because we believed it was representative of the area reasonably surveyed based on the drift path of the boat, the casting behavior of anglers and the movement of fish to the fishing gear. Then, the five meter buffer area of each caught fish was removed to prevent duplication of a quadrature point with the location of a caught fish. We then created a uniform grid of points across this survey area. We repeated this step four times for each species, generating 2,000, 5,000, 10,000 and 15,000 quadrature points. We chose these quantities to determine the point at which the model best converged, our maximum level of quadrature points was at approximately 15,000 because this was the largest number that could fit within the drift buffers. We then buffered each of these points to a radius of five meters to match that of the fish, and took the average value for the bathymetry, slope, substrate, and VRM within the buffer. We removed any quadrature points that had overlapping buffers or no longer fell fully within the survey area once they were buffered.

To evaluate the effect of the number of quadrature points on model performance, we ran four identical models, as described above, with the four different quantities of quadrature points. We then used the area under the ROC curve (AUC) and true skill statistic (TSS) metrics to determine the quantity of quadrature points that maximized model predictive power (El-Gabbas & Dormann 2018). These steps were repeated for each of the three species.

1.2.6. Habitat Characteristic Selection

We used a binomial generalized linear model with a logit link for seafloor variable selection. We ran the model using the dataset with the number of quadrature points as determined above and included all of the habitat characteristics; bathymetry, slope, substrate, and VRM. We included the habitat characteristics that were statistically significant in the binomial generalized linear model as standardized variables and as spatially varying coefficients in the final model fitted in the sdmTMB package. Spatially varying coefficients are an optional model component of sdmTMB models. They can be included in the model to determine whether the effect of the variable varies across space (Anderson et al., 2024). We included the statistically significant habitat characteristics as spatially varying coefficients as well to test if the effect the variable had on the distribution of fish varied across the study areas or within the areas between the MPAs and areas open to fishing.

1.3. Results

1.3.1. Vermilion Rockfish

We found that for Vermilion rockfish (n = 1,000), the model predictive power maximized at 1,153 quadrature points (Figure 2A-B) and a mesh cutoff distance of 1 km. The final model included standardized bathymetry (coefficient = -0.33, estimate = exp(-0.33) = 0.71, p = 0.12) and site (coefficient = -2.48, estimate = exp(-2.48) = 0.08, p<0.0001), with standardized bathymetry also included as a spatially varying coefficient (σ= 0.97) (Figures 3, 4C-D). The estimate of standardized bathymetry was 0.71, such that for every one standard deviation decrease in depth (i.e., shallower) (5.7 m) from the mean depth (26.2 m) there is a 71% decrease in the relative intensity of vermilion rockfish sightings. The parameter estimate for site was 0.08, (1-0.08 = 0.92) with the reference as the main factor, so the relative intensity of Vermilion rockfish sightings decreased by 92% in the reference area relative to the MPA. The model had a Matérn range of 1.92 km (representing the distance at which 2 locations are functionally independent), and a spatial standard deviation of 0.75 (Figure 5). The magnitude of variability between the spatially varying effect of bathymetry and the latent spatial field can be compared directly, which suggests that the effect of bathymetry is a larger source of variation.

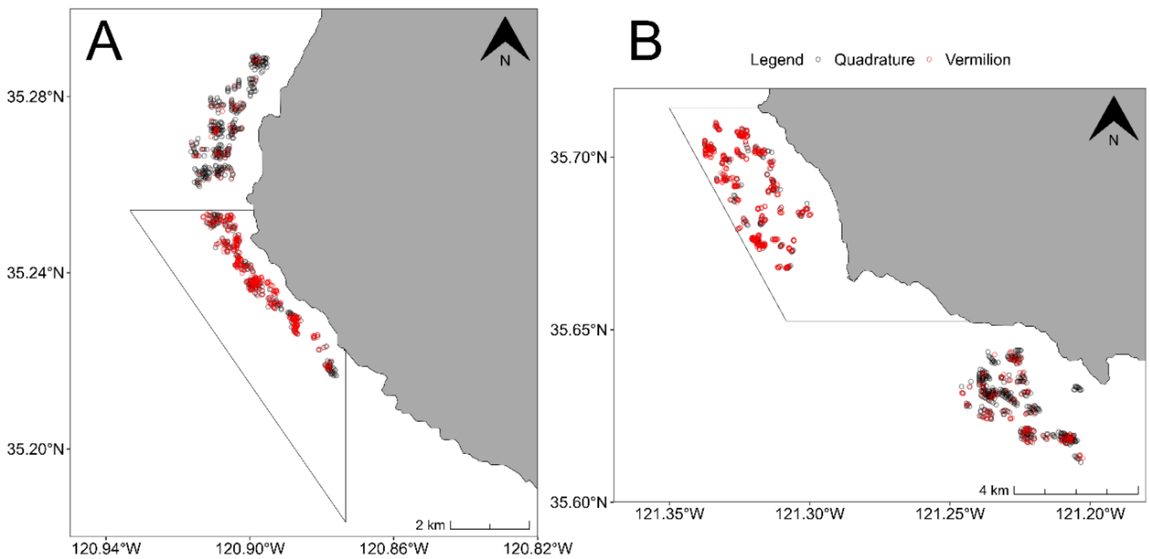

Figure 2. Map of the caught Vermilion rockfish (*Sebastes miniatus*) and the quadrature points at (A) Point Buchon and (B) Piedras Blancas in the study region of San Luis Obispo County, California.

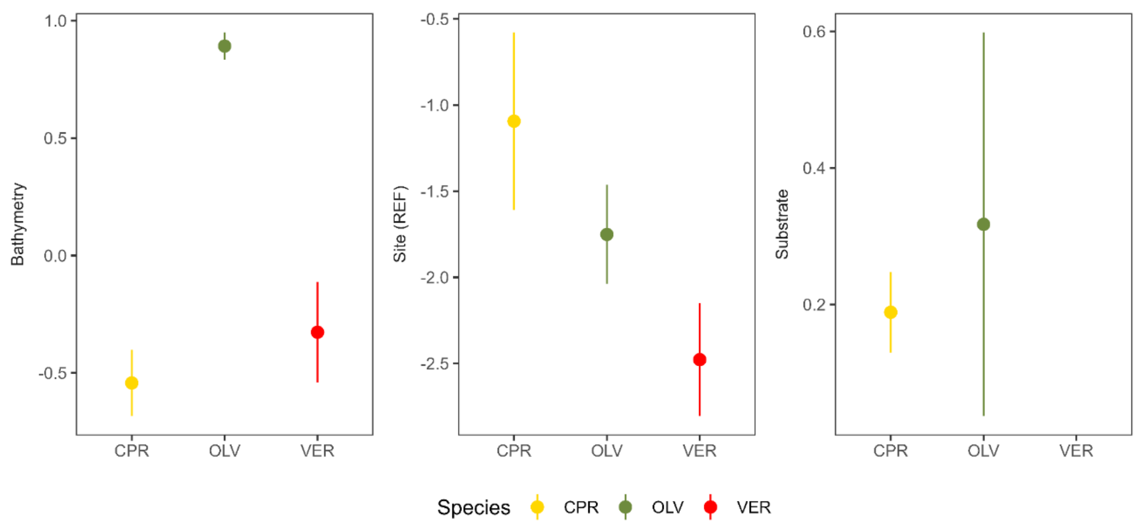

Figure 3. Results from the down-weighted spatial generalized mixed effects model with a Poisson distribution and a log link for Copper (*Sebastes caurinus*), Olive (*S. serranoides*), and Vermilion rockfish (*S. miniatus*). The fish data for these models were collected from June 2018 to September 2022 on the central coast of California, the habitat data was collected from the Seafloor Mapping Laboratory at California State University, Monterey Bay. While the coefficients output from the model are on a logarithmic scale, we include the exponentiated coefficients in this figure. The coefficients for bathymetry and substrate were standardized using z-score. Substrate is a continuous variable between 1 and - which expresses the proportion of the seafloor that is hard vs soft, where 1 is completely hard and 0 is completely soft. The variable site is a factor with two levels, MPA and REF where MPA is a marine protected area closed to

fishing and REF is a nearby area that is open to fishing. Error bars represent the standard error.

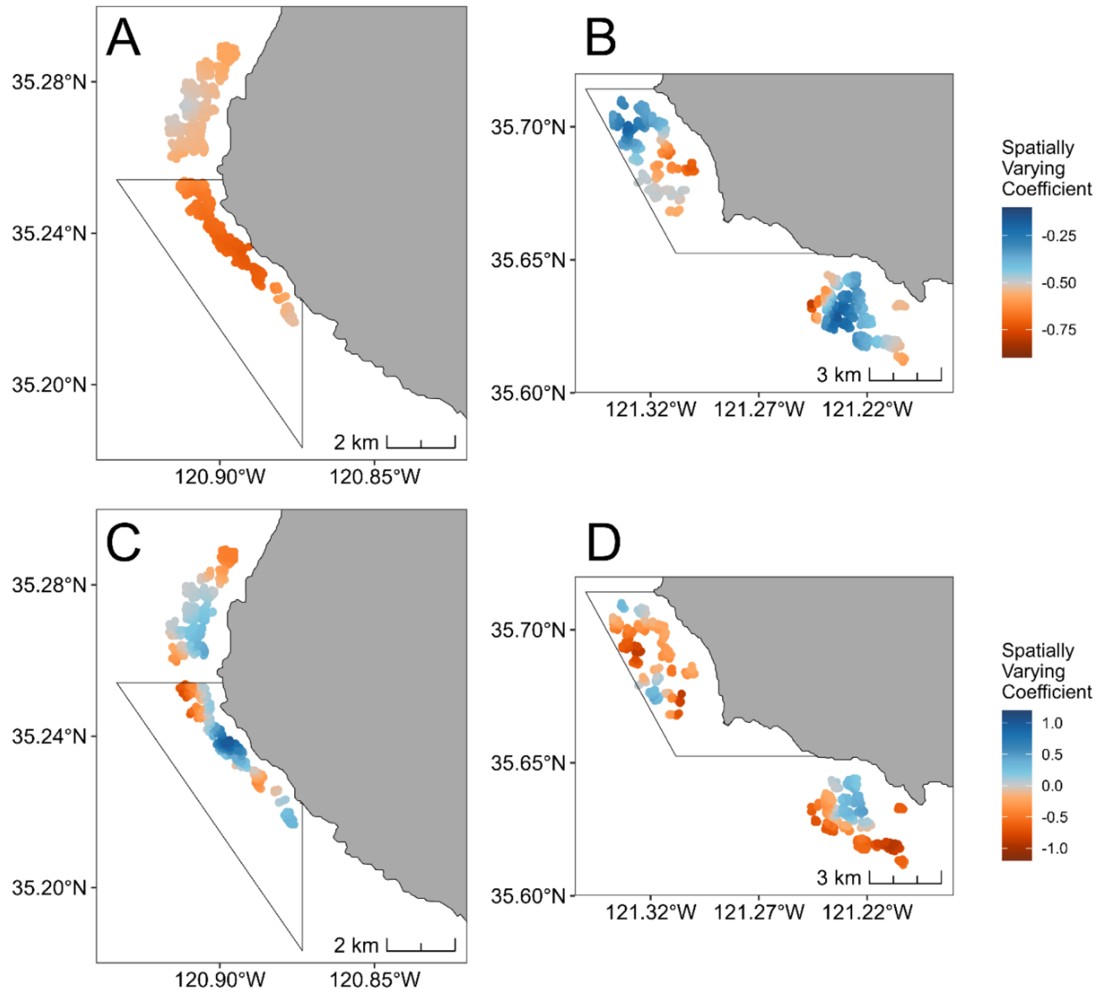

Figure 4. Map of the spatially varying coefficient of bathymetry for Copper rockfish (*Sebastes caurinus*) at (A) Point Buchon and (B) Piedras Blancas and for Vermilion rockfish (*Sebastes miniatus)* at (C) Point Buchon and (D) Piedras Blancas.

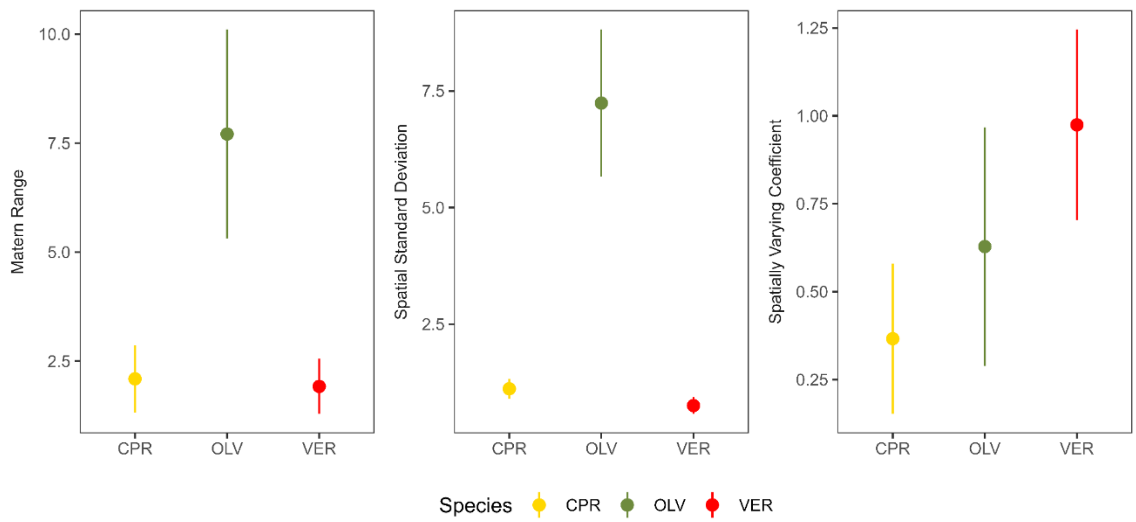

Figure 5. Results from the down-weighted spatial generalized mixed effects model with a Poisson distribution and a log link for Copper (*Sebastes caurinus*), Olive (*S. serranoides*), and Vermilion rockfish (*S. miniatus*). The fish data for these models were collected from June 2018 to September 2022 on the central coast of California, the habitat data was collected from the Seafloor Mapping Laboratory at California State University, Monterey Bay. The fixed effects from the model are included. The matérn range, the distance at which two locations in the modeled area are no longer correlated, the spatial standard deviation, a measure of how clustered the data are, and the spatially varying coefficient are included for each of the three species. Error bars represent standard error.

1.3.2. Copper Rockfish

We found that for Copper rockfish (n = 448), the model predictive power maximized at 9,091 quadrature points (Figure 6A-B) and a mesh cutoff distance of 1 km. The final model included standardized bathymetry (coefficient = -0.54, estimate = exp(-0.54) = 0.58, p = 0.00013), standardized substrate (coefficient = 0.19, estimate = exp(0.19) = 1.2, p = 0.0014) and site (coefficient = -1.09, estimate = exp(-1.09) = 0.34,p = 0.034), with standardized bathymetry also included as a spatially varying coefficient (σ= 0.37) (Figure 3, 4A-B). The estimate of standardized bathymetry was 0.58, which means for every one standard deviation decrease in depth (shallower; 5.8 m) from the mean depth of (25.8 m) there is a 58% decrease in the relative intensity of Copper rockfish sightings. The estimate of standardized substrate was 1.2, which means for every one standard deviation increase in the proportion of hard bottom (0.41) from the mean proportion of hard bottom (0.59) the relative intensity of Copper rockfish sightings increases by 1.2 times. The estimate of site was 0.34, (1-0.34 = 0.66) with the reference as the main factor, which means the relative intensity of Copper rockfish sightings decreases by 66% in the reference area relative to the MPA. The model had a Matérn range of 2.09 km (representing the distance at which 2 locations are functionally independent) and a spatial standard deviation of 1.11 (Figure 5).

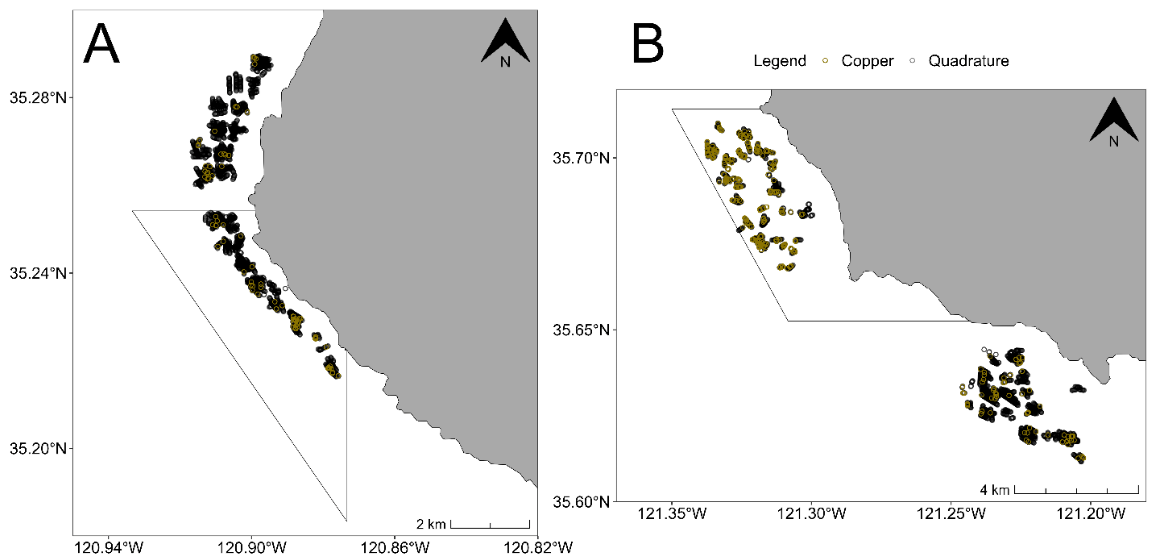

Figure 6. Map of the caught Copper rockfish (*Sebastes caurinus*) and the quadrature points at (A) Point Buchon and (B) Piedras Blancas in the study region of San Luis Obispo County, California.

1.3.3. Olive Rockfish

While the model for Olive rockfish converged, the residuals are much larger than the other two species. We report and discuss the results of the Olive rockfish model, however, it is important to note that these may be unreliable due to modeling difficulties.

Olive rockfish are a mobile, mid-water column dwelling species (Love et al. 2002). Perhaps the habitat characteristics were not able to capture all of the factors influencing the distribution of Olive rockfish.

We found that for Olive rockfish (n = 1,865), the model predictive power maximized at 8,466 quadrature points (Figure 7A-B) and a mesh cutoff distance of 4 km. The final model include standardized bathymetry (coefficient = 0.89, estimate = exp(0.89) = 2.43, $p < 0.0001$), standardized substrate (coefficient = 0.32, estimate = exp(0.32) = 1.38, $p = 0.26$), and site (coefficient = -1.75, estimate = exp(-1.75) = 0.17, $p < 0.0001$), with standardized substrate also included as a spatially varying coefficient (σ= 0.63) (Figures 3, 8). The estimate of standardized bathymetry was 2.43, which means for every one standard deviation decrease in depth (shallower; 5.7 m) from the mean depth (25.8 m) the relative intensity of Olive rockfish sightings increases by 2.43 times. The estimate of standardized substrate was 1.38, which means that for every one standard deviation increase in the proportion of hard bottom (0.41) from the mean proportion of hard bottom (0.59) the relative intensity of Olive rockfish sightings increases by 1.38 times. The estimate of site was 0.17, (1-0.17 = 0.83) with the reference site as the main factor, which means the relative intensity of Olive rockfish sightings decreases by 83% in the reference area relative to the MPA. The model had a Matérn range of 7.71 km, and a spatial standard deviation of 7.24 (Figure 5).

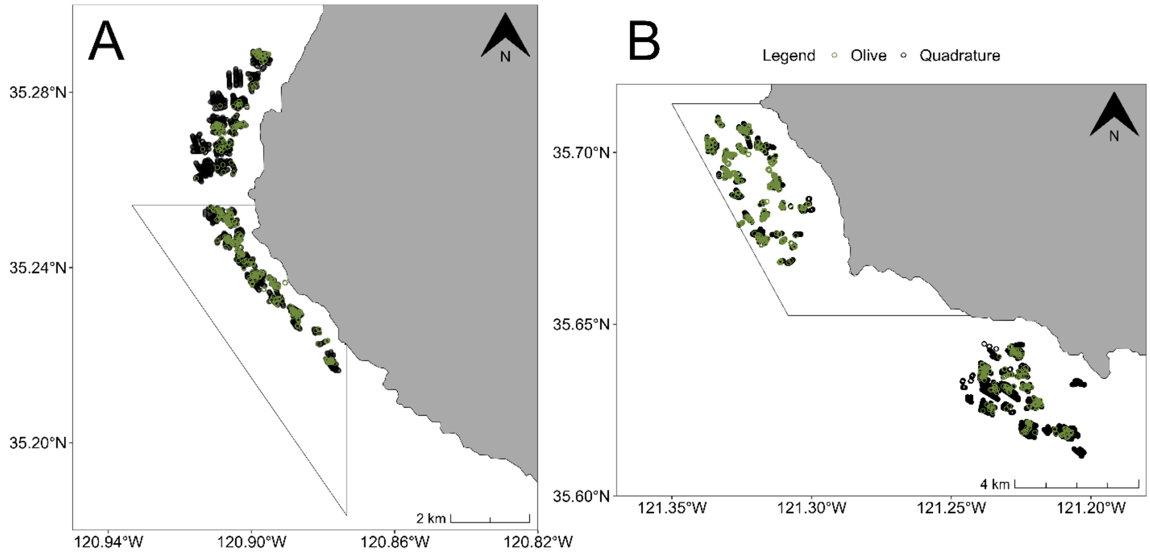

Figure 7. Map of the caught Olive rockfish (*Sebastes serranoides*) and the quadrature points at (A) Point Buchon and (B) Piedras Blancas in the study region of San Luis Obispo County, California.

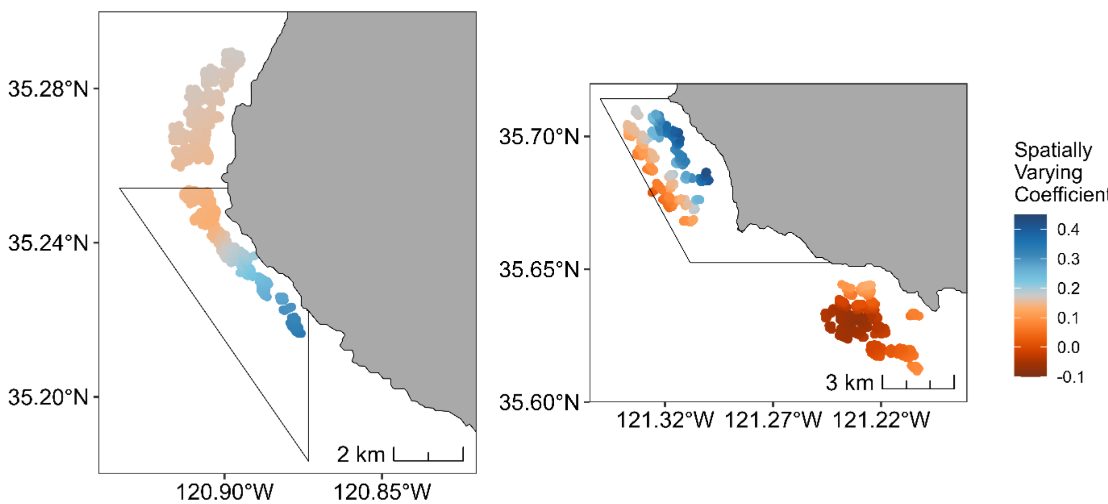

Figure 8. Map of the spatially varying coefficient of substrate for Olive rockfish (*Sebastes serranoides*) at (A) Point Buchon and (B) Piedras Blancas.

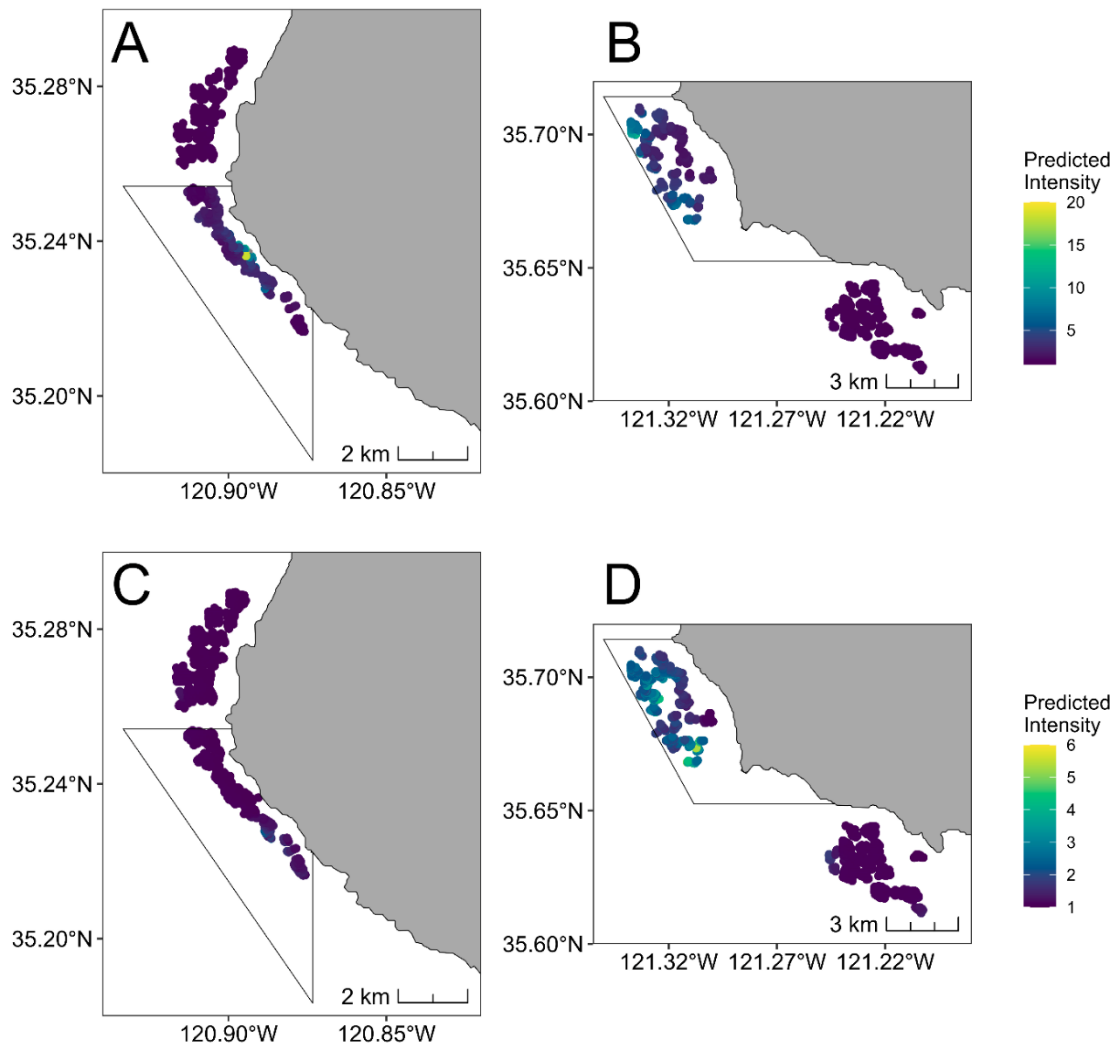

Figure 9. Map of the predicted relative intensity of sightings for Vermilion rockfish (*Sebastes miniatus)* at (A) Point Buchon and (B) Piedras Blancas and for Copper rockfish (*Sebastes caurinus*) at (C) Point Buchon and (D) Piedras Blancas.

1.4. Discussion

We sought to explore the effects of MPA protection on habitat use in three commercial and recreationally important rockfish species, Vermilion, Copper and Olive rockfish. To accomplish this, we used hook-and-line, catch-and-release fishing data as a way to create species distribution models across our study areas. Many rockfishes utilize habitats differently, but the factors that drive those associations are less well known.

However, there is little information on how MPA protection and fishing pressure may alter habitat use. Having an understanding of how fishing pressure may influence habitat use could improve our ability to account for habitat and fishing pressure in fisheries models.

Bathymetry is often a strong predictor of fish assemblages and species-specific occurrences because many species occupy specific depth ranges (Anderson et al. 2009; Friedlander et al. 2007; Williams & Ralston 2002). For this reason, depth is commonly used as a way to understand the relationship between fish distributions and habitat features; this is particularly true for rockfishes (Williams & Ralston 2002). Our results were consistent with these previous findings; bathymetry was an important predictor for all three of our study species (Figure 3). We found that for Vermilion and Copper rockfish, the relative intensity of sightings increased with depth, while Olive rockfish showed the opposite pattern. The effect of depth on relative intensity of sightings was larger for Vermilion rockfish compared to Copper rockfish, suggesting that while both species are more commonly found in deeper waters, depth may be more important for Vermilion than Copper rockfish. This is consistent with other work suggesting that while Olive rockfish are more abundant in shallower waters (Love et al. 2009), Vermilion and Copper rockfish are more commonly associated with deeper waters and tend to migrate deeper with increased size and age (Casselle et al., 2002; Love et al., 2009; Love & Schroeder 2007). Vermilion rockfish rarely inhabit the nearshore environment once they reach adulthood, due to their ontogenetic shift into deeper waters. This shift could explain the larger effect of depth for Vermilion rockfish compared to Copper rockfish, which often inhabit shallower waters (Love et al. 2009).

Seafloor composition is another habitat feature that influences fish distributions and diversity (Anderson & Yoklavich 2007; Pittman et al. 2007). Hard or rock substrate is typically positively associated with higher fish populations and diversity, while less complex soft or sand substrates contain fewer fish and lower diversity (Anderson et al. 2009; Anderson & Yoklavich 2007; Friedlander et al. 2007). Rockfishes, as the name suggests, are more commonly found in areas with hard or mixed substrate (Anderson & Yoklavich 2007). Copper rockfish in particular are typically found over complex bottom and the intersection of sand and reef, but will use low relief rock substrate (Johnson et al. 2003; Love et al. 2002). While Olive rockfish typically reside higher in the water column, they are found over complex, high relief areas (Love et al. 2011; Love et al. 2002). We were unable to directly account for complexity or relief, however, these studies are consistent with our findings that the relative intensity of sightings of Copper and Olive rockfish increase with a higher proportion of hard substrate (Figure 3). Vermilion rockfish are known to inhabit areas with shelter and deep crevices (Caselle et al. 2002; Love et al. 2006). Thus, we were surprised that substrate was not an important predictor for Vermilion rockfish. It is possible the habitat measure of proportion of hard versus soft substrate alone was not able to capture the actual structure to which Vermilion rockfish are responding. Higher relief mapping or additional bottom characteristics might be better able to account for available shelter space and crevice size.

MPAs are an important and effective fisheries management tool. They lead to increases in fish biomass, abundance, species richness and can improve community resilience (Ziegler et al. 2023; Hamilton et al. 2021; Chirico et al. 2017; Guidetti et al. 2014), including areas monitored by CCFRP. Therefore, we predicted that site (MPA vs

REF) would be a statistically significant predictor for all three species, with the relative intensity of sightings increasing in the MPA relative to the REF (Figure 3). This effect was largest for Vermilion rockfish, followed by Olive rockfish and smallest for Copper rockfish (Figure 3). It is likely that this pattern is driven by a combination of fishing pressure and sample size. Vermilion rockfish have historically been a highly targeted species in the recreational fishery in California (Monk et al. 2021). Other work has found that the magnitude of MPA responses can scale with pre-existing fishing pressure (Ziegler et al. 2022); this may partly explain the increase in VER. This combined with the larger sample size (n = 1,000) compared to Copper rockfish (n = 448), could explain why Vermilion rockfish have the largest effect size of MPA protection. While we expected site to be an important predictor of fish presence and abundance, we also wanted to know if the relationship with habitat varied between the MPA and REF sites; this is why we included spatially varying coefficients.

There has been no work done to examine how MPAs or fishing pressure influence the habitat usage of fish species. However, understanding this relationship could improve MPA management and designation. We tested this relationship by including bathymetry, for Vermilion and Copper rockfish, and substrate, for Olive rockfish, as spatially varying coefficients. For Copper rockfish, it appears that there is a difference in the effect size of bathymetry between Point Buchon (Figure 4A) and Piedras Blancas (Figure 4B), with depth having a larger effect size in Point Buchon compared to Piedras Blancas. Within Point Buchon, it appears that the effect size is larger still in the MPA compared to the REF (Figure 4A), whereas in Piedras Blancas the spread between the MPA and REF seems relatively equal. While the effect size of bathymetry varies more for Vermilion rockfish than Copper rockfish, the variance appears to be more equally spread across Point Buchon and Piedras Blancas MPA and REF (Figure 4C-D). However, within the Point Buchon MPA, there is a large positive effect of depth on the relative intensity of sightings for Vermilion rockfish (Figure 4C). We included substrate as a spatially varying coefficient for Olive rockfish, which allowed for the effect of substrate on the intensity of Olive rockfish sightings to vary across the study area (Figure 8A-B). There is a slight difference in the effect substrate has on Olive rockfish between the Point Buchon and Piedras Blancas study areas, where substrate has a larger effect on Olive rockfish in Point Buchon relative to Piedras Blancas (Figure 8A-B). It is possible that this is a result of the large residuals of the Olive rockfish model. Since Olive rockfish are a water column species, there may be other factors that we were unable to account for that impact how Olive rockfish respond to substrate. While we are unable to definitively explain what causes the effect of bathymetry and substrate to vary across and within the study areas, these results demonstrate the value of further research examining how fishing pressure influences the habitat associations of different fish species.

Additionally, we generated maps of predicted intensity of sightings for both Vermilion and Copper rockfish (Figure 9A-D). For both species, the model predicts that relative intensity of sightings will be the largest in the Piedras Blancas MPA (Figure 9B & 9D). This may be due to the fact that Piedras Blancas MPA was where the majority of both Vermilion (n = 460) and Copper rockfish (n = 293) were caught. Additionally, Vermilion rockfish have a higher overall range of predicted intensity of sightings compared to Copper rockfish, which may be attributed to the larger sample size of Vermillion (n = 1000) relative to Copper rockfish (n = 448). The predicted intensity for

Olive rockfish was much larger than should be possible, and as a result we did not include these results. It is possible that as a water column species, there were factors that influence Olive rockfish distribution that were unaccounted for in the model, but we are uncertain what those factors might be.

Vermilion and Copper rockfish had smaller Matérn ranges (the distance at which two locations within the study area are no longer correlated) and spatial standard deviations compared to Olive rockfish (Figure 5). Smaller Matérn ranges suggest that the presence of Vermilion and Copper rockfishes are more patchy than Olive rockfish, which may relate to the species' mobility. Olive rockfish are a more mobile species with a larger home range, and therefore that they are more likely to move across a larger area–and more diverse habitats– than Vermilion and Copper rockfish (Love et al. 2002). Having smaller spatial standard deviations suggests that Vermilion and Copper rockfish are more homogenous in their distribution across the study area compared to Olive rockfish (the estimated magnitude of variability of the spatial field was approximately 7x higher for Olive rockfish). This could be because Vermilion and Copper rockfish are typically associated with the seafloor, while Olive rockfish occupy the water column and are a schooling species (Love et al. 2002). Since Olive rockfish are not as associated with the seafloor and are more mobile, this could lead to a less homogenous distribution across the study area compared to Vermilion and Copper rockfish.

It is important to recognize the limitations of these models to make generalizations about the larger rockfish population. The data for this study were collected using hook-and-line methods, an imperfect metric of a species presence and abundance. Absence of a caught fish does not definitively mean that a fish was not in the sampled location. By including multiple years of data, we believe we minimized the impact of false absences on our results, but additional years of data could improve model results. Further, though the locations assigned to each caught fish were interpolated from in situ waypoints, we believe that a buffer of five meters helps to account for any variability in the true location of the fish when it was caught. Though hook-and-line methods are imperfect, we believe this study shows that it is an important tool that could be used as an easier, more affordable way to gather valuable fish presence data, as opposed to remotely operated vehicle or SCUBA methods. Especially as habitat data and GPS accuracy continue to improve, hook-and-line catch data may be able to shed new light on how habitats influence distributions and how protection may modulate those relationships. Understanding the relationship between species distributions and habitat variables can help to inform management decisions. By understanding how specific species utilize habitat, managers can make spatially informed decisions to support the resilience of commercially and recreationally valuable species (Young & Carr 2015).

1.5. Conclusion

We sought to use hook-and-line, catch-and-release fishing as a new way of collecting data to generate species distribution models for three species of commonly caught rockfish on the central coast of California. We utilized PPMs within a spatial GLMM framework as a way to model this presence-only data. Additionally, we included the habitat variables as spatially varying coefficients to test if there was a difference in the way the rockfishes utilized habitat inside and outside of MPAs. We found bathymetry and site to be an important predictor for Vermilion, Copper, and Olive rockfish, and substrate to be important for Copper and Olive rockfish. Bathymetry was also a spatially varying coefficient for Vermilion and Copper rockfish, while substrate was spatially varying for Olive rockfish. Understanding the way habitat influences a species distribution and whether that affect varies spatially has many broader management and protection implications. Specifically, it can improve the ability of scientists to determine areas that are essential for fish conservation, as well as estimate habitat availability (Anderson & Yoklavich 2007).

WORKS CITED

Anderson, S. C., Ward, E. J., English, P. A., & Barnett, L. A. K. (n.d.). *sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields*. https://doi.org/10.1101/2022.03.24.485545

Anderson, T. J., Syms, C., Roberts, D. A., & Howard, D. F. (2009). Multi-scale fish-habitat associations and the use of habitat surrogates to predict the organisation and abundance of deep-water fish assemblages. *Journal of Experimental Marine Biology and Ecology*, *379*(1–2), 34–42. https://doi.org/10.1016/j.jembe.2009.07.033

Anderson, T. J., & Yoklavich, M. M. (2007). Multiscale habitat associations of deepwater demersal fishes off central California. *Fisheries Bulletin*, *105*(2), 168–179.

Barnett, L. A. K., Ward, E. J., & Anderson, S. C. (2021). Improving estimates of species distribution change by incorporating local trends. *Ecography*, *44*(3), 427–439. https://doi.org/10.1111/ecog.05176

Bozec, Y., Dolédec, S., & Kulbicki, M. (2005). An analysis of fish-habitat associations on disturbed coral reefs: chaetodontid fishes in New Caledonia. *Journal of Fish Biology*, *66*, 966–982. https://doi.org/10.1111/j.1095-8649.2005.00652.x

Carr, M. H. (1991). Habitat selection and recruitment of an assemblage of temperate zone reef fishes. In *Mar. Biol. Ecol* (Vol. 146).

Caselle, J. E., Love, M. S., Fusaro, C., & Schroeder, D. (2002). Trash or habitat? Fish assemblages on offshore oilfield seafloor debris in the Santa Barbara Channel, California. *ICES Journal of Marine Science*, *59*. https://doi.org/10.1006/jmsc.2002.1264

Chirico, A. A. D., McClanahan, T. R., & Eklöf, J. S. (2017). Community- and government-managed marine protected areas increase fish size, biomass and potential value. *PLoS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0182342

Chittaro, P. M. (2004). Fish-habitat associations across multiple spatial scales. *Coral Reefs*, *23*(2), 235–244. https://doi.org/10.1007/s00338-004-0376-z

Cope, J., Dick, E. J., Maccall, A., Monk, M., Soper, B., & Wetzel, C. (2015). Data-moderate stock assessments for brown, China, copper, sharpchin, stripetail, and yellowtail rockfishes and English and rex soles in 2013.

El-Gabbas, A., & Dormann, C. F. (2018). Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography*, *41*(7), 1161–1172. https://doi.org/10.1111/ecog.03149

Friedlander, A. M., Brown, E., & Monaco, M. E. (2007). Defining reef fish habitat utilization patterns in Hawaii: Comparisons between marine protected areas and areas open to fishing. *Marine Ecology Progress Series*, *351*, 221–233. https://doi.org/10.3354/meps07112

Gleason, M., Fox, E., Ashcraft, S., Vasques, J., Whiteman, E., Serpa, P., Saarman, E., Caldwell, M., Frimodig, A., Miller-Henson, M., Kirlin, J., Ota, B., Pope, E., Weber, M., & Wiseman, K. (2013). Designing a network of marine protected areas in California: Achievements, costs, lessons learned, and challenges ahead. *Ocean and Coastal Management*, *74*, 90–101. https://doi.org/10.1016/j.ocecoaman.2012.08.013

Guidetti, P., Baiata, P., Ballesteros, E., di Franco, A., Hereu, B., Macpherson, E., Micheli, F., Pais, A., Panzalis, P., Rosenberg, A. A., Zabala, M., & Sala, E. (2014). Large-scale assessment of mediterranean marine protected areas effects on fish assemblages. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0091841

Hanan, D. A., & Curry, B. E. (2012). Long-Term Movement Patterns and Habitat Use Of Nearshore Groundfish: Tag-Recapture in Central and Southern California Waters. *The Open Fish Science Journal*, *5*, 30–43.

Hannah, R. W., & Rankin, P. S. (2011). Site fidelity and movement of eight species of pacific rockfish at a high-relief rocky reef on the oregon coast. *North American Journal of Fisheries Management*, *31*(3), 483–494. https://doi.org/10.1080/02755947.2011.591239

Hamilton, S. L., Caselle, J. E., Malone, D. P., & Carr, M. H. (2010). Incorporating biogeography into evaluations of the Channel Islands marine reserve network. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18272–18277. https://doi.org/10.1073/pnas.0908091107

Hamilton, S. L., Starr, R. M., Wendt, D. E., Ruttenberg, B. I., Caselle, J., Semmens, B. X., Bellquist, L., Morgan, S., Mulligan, J., Tyburczy, J., Ziegler, S. L., Brooks, R. O., Waltz, G., Mason, E., Honeyman, C., Small, S., & Staton, J. (2021). *California Collaborative Fisheries Research Program (CCFRP)-Monitoring and Evaluation of California Marine Protected Areas*.

Ivens-Duran, M. (2014). A spatial analysis of changes in recreational fishing pressure on the central coast of California subsequent to MPA implementation.

Johnson, S. W., Murphy, M. L., & Csepp, D. J. (2003). Distribution, habitat, and behavior of rockfishes, Sebastes spp., in nearshore waters of southeastern Alaska: observations from a remotely operated vehicle. *Environmental Biology of Fishes* (Vol. 66).

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, *70*. https://doi.org/10.18637/jss.v070.i05

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. In *J. R. Statist. Soc. B* (Vol. 73).

Lindholm, J. B., Auster, P. J., Ruth, M., & Kaufman, L. (2001). Modeling the Effects of Fishing and Implications for the Design of Marine Protected Areas: Juvenile Fish Responses to Variations in Seafloor Habitat. *Conservation Biology*, *15*(2), 424–437. https://www.jstor.org/stable/2641840?seq=1&cid=pdf-

Love, M. S. (2011). *Certainly More Than You Want to Know About The Fishes of The Pacific Coast*. Really Big Press.

Love, M. S., Caselle, J. E., & Buskirk, W. V. (1998). A severe decline in the commercial passenger fishing vessel rockfish (Sebastes spp.) catch in the southern California bight, 1980-1996. In *CalCOFl Rep* (Vol. 39).

Love, M. S., Morris, P., McCrae, M., & Collins, R. (1990). *Life history aspects of 19 rockfish species (Scorpaenidae: Sebastes) from the Southern California Bight* (Issue February).

Love, M. S., Nishimoto, M. M., Snook, L., Bull, A. S., Laidig, T., Kui, L., Watters, D., & Yoklavich, M. (2022). A structured deepwater fish community in an isolated benthic feature off Southern California. *Bulletin of Marine Science*, *98*(3), 221–246. https://doi.org/10.5343/bms.2022.0005

Love, M. S., & Schroeder, D. M. (2007). A characterization of the fish assemblage of deep photic zone rock outcrops in the Anacapa Passage, Southern California, 1995 to 2004, with evidence of a regime shift. In *CalCOFI Rep* (Vol. 48).

Love, M. S., Schroeder, D. M., Lenarz, B., & Cochrane, G. R. (2006). Gimme Shelter: The Importance of Crevices to Some Fish Species Inhabiting a Deeper-Water Rocking Outcrop in Southern California. In *GIMME SHELTER CalCOFI Rep* (Vol. 47). www.pcouncil.org/reserves/reservesback.html

Love, M. S., Yoklavich, M., & Schroeder, D. M. (2009). Demersal fish assemblages in the Southern California Bight based on visual surveys in deep water. *Environmental Biology of Fishes*, 84(1), 55-68. https://doi.org/10.1007/s10641-008-9389-8

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

Love, M. S., & York, A. (2006). The relationships between fish assemblages and the amount of bottom horizontal beam exposed at California oil platforms: fish habitat preferences at man-made platforms and (by inference) at natural reefs. *Fishery Bulletin*, *104*(4), 542–549.

Lowe, C. G., Anthony, K. M., Jarvis, E. T., Bellquist, L. F., & Love, M. S. (2009). Site Fidelity and Movement Patterns of Groundfish Associated with Offshore Petroleum Platforms in the Santa Barbara Channel. *Marine and Coastal Fisheries*, *1*(1), 71–89. https://doi.org/10.1577/c08-047.1

Monk, M. H., Dick, E. J., Field, J. C., Saas, E. M., Rogers, T. L., Monk, M. H., Field, J. C., & Saas, E. M. (2021). *The status of Vermilion Rockfish (Sebastes miniatus) and Sunset Rockfish (Sebastes crocotulus) in U.S. waters off the coast of California north of Point Conception in 2021*.

Muller K. (2023). Hms: pretty time of day. R package version 1.1.3.

Odum, E. P. (1963). *Ecology: A Bridge Between Science and Society*. Houghton Mifflin Harcourt School.

Pittman, S. J., Christensen, J. D., Caldow, C., Menza, C., & Monaco, M. E. (2007). Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecological Modelling*, *204*(1–2), 9–21. https://doi.org/10.1016/j.ecolmodel.2006.12.017

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., & Warton, D. I. (2015). Point process models for presence-only analysis. In *Methods in Ecology and Evolution* (Vol. 6, Issue 4, pp. 366–379). British Ecological Society. https://doi.org/10.1111/2041-210X.12352

Rohit Raj Kolora, S., Owens, G. L., Manuel Vazquez, J., Stubbs, A., Chatla, K., Jainese, C., Seeto, K., McCrea, M., Sandel, M. W., Vianna, J. A., Maslenikov, K., Bachtrog, D., Orr, J. W., Love, M., & Sudmant, P. H. (2021). Origins and evolution of extreme life span in Pacific Ocean rockfishes. In *Science* (Vol. 374). https://www.science.org

Starr, R.M., D.E. Wendt, C.L. Barnes, C.I. Marks, D. Malone, G. Waltz, K.T. Schmidt, J. Chiu, A.L. Launer, N.C Hall, and N. Yochum (2015). Variation in responses of fishes across multiple reserves within a network of marine protected areas in temperate waters. *PLoS One*, 10(3), e0118502.

Stephens, J., Wendt, D., Wilson-Vandenberg, D., Carroll, J., Nakamura, R., Nakada, E., Rienecke, S., & Wilson, J. (2006). Rockfish Resources of the South Central California Coast: Analysis of the Resource from Partyboat Data, 1980-2005. In *CalCOFI Rep* (Vol. 47).

Temple Lang D. (2024). XML: tools for parsing and generating XML within R and S-plus. R package version 3.99-0.17.

Thorson, J. T., Barnes, C. L., Friedman, S. T., Morano, J. L., & Siple, M. C. (2023). Spatially varying coefficients can improve parsimony and descriptive power for species distribution models. *Ecography*, *2023*(5). https://doi.org/10.1111/ecog.06510

Wedding, L., & Yoklavich, M. M. (2015). Habitat-based predictive mapping of rockfish density and biomass off the central California coast. *Marine Ecology Progress Series*, *540*, 235–250. https://doi.org/10.3354/meps11442

Wendt, D. E., & Starr, R. M. (2009). Collaborative Research: An Effective Way to Collect Data for Stock Assessments and Evaluate Marine Protected Areas in California. *Marine and Coastal Fisheries*, *1*(1), 315–324. https://doi.org/10.1577/c08-054.1

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan., L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686

Williams, E. H., & Ralston, S. (2002). Distribution and co-occurrence of rockfishes (family: Sebastidae) over trawlable shelf and slope habitats of California and southern Oregon. *Fishery Bulletin*, *100*(4), 836–855.

Young, M., & Carr, M. H. (2015). Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef fishes. Diversity and Distributions, 21(12), 1428–1440. https://doi.org/10.1111/ddi.12378

Young, M. A., Iampietro, P. J., Kvitek, R. G., & Garza, C. D. (2010). Multivariate bathymetry-derived generalized linear model accurately predicts rockfish distribution on Cordell Bank, California, USA. *Marine Ecology Progress Series*, *415*, 247–261. https://doi.org/10.3354/meps08760

Zeileis, A., & Grothendieck, G. (2005). Zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*. *14*(6), 1-27. doi:10.18637/jss.v014.i06

Ziegler, S. L., Brooks, R. O., Hamilton, S. L., Ruttenberg, B. I., Chiu, J. A., Fields, R. T., Waltz, G. T., Shen, C., Wendt, D. E., & Starr, R. M. (2022). External fishing effort regulates positive effects of no-take marine protected areas. *Biological Conservation*, *269*. https://doi.org/10.1016/j.biocon.2022.109546

Ziegler, S. L., Johnson, J. M., Brooks, R. O., Johnston, E. M., Mohay, J. L., Ruttenberg, B. I., Starr, R. M., Waltz, G. T., Wendt, D. E., & Hamilton, S. L. (2023). Marine protected areas, marine heatwaves, and the resilience of nearshore fish communities. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-28507-1

Ziegler, S. L., Brooks, R. O., Bellquist, L. F., Caselle, J. E., Morgan, S. G., Mulligan, T. J., Ruttenberg, B. I., Semmens, B. X., Starr, R. M., Tyburczy, J., Wendt, D. E., Buchheister, A., Jarrin, J. R. M., Pasparakis, C., Jorgensen, S. J., Chiu, J. A., Colby, J., Coscino, C. L., Davis, L., Castro, F., Elstner, J. T., Honeyman, C., Jarvis Mason, E. T., Johnston, E. M., Small, S. L., Staton, J., Waltz, G. T., Basnett, B., Satterthwaite, E. V., Killeen, H., Dibble, C. D., & Hamilton, S. L. (2024). Collaborative fisheries research reveals reserve size and age determine efficacy across a network of marine protected areas. *Conservation Letters*, 17(2). https://doi.org/10.1111/conl.13000

Appendix

A: GPS Standard Operating Procedure

Introduction

The California Collaborative Fisheries Research Program (CCFRP) was created in 2007 to monitor marine protected areas (MPAs) off the coast of California. CCFRP regularly monitors 12 MPAs and nearby reference areas using hook-and-line, catch-and-release fishing for standardized periods of time (for more information on the methodology of CCFRP see (Wendt and Starr 2009). This program has been collecting valuable population data on nearshore groundfish for 17 years and has had many interesting findings on how MPAs influence rockfish populations (Ziegler et al. 2024; Ziegler et al. 2023; Starr et al. 2015; etc.). However, location data collected prior to 2024 focused primarily on ensuring fishing activities aligned with standardized sampling design and not necessarily on collecting fine-scale location data for each individual fish. CCFRP samples pre-determined sampling cell locations, by fishing in 15-minute increments called drifts, during which the vessel is drifting throughout the sampling cell and volunteer anglers are consistently fishing. In years prior to 2024, the latitude and longitude were recorded at the beginning and end of every 15-minute fishing drift. Several times throughout this drift a waypoint, the latitude and longitude that corresponded to a caught fish, was recorded. Therefore, some location data was captured during each drift, but most individual fish do not have an associated location. By interpolating the start, end and any intermediate waypoints, it is possible to assign an approximate catch location to any individual for which location data were not recorded. While this is a good method for obtaining an approximate catch location from data that were previously collected, we are proposing a new method to collect location information in the field with higher accuracy and precision. This new method involves tracking the vessel's movement throughout the day on a handheld GPS unit and recording the exact time a fish is captured as often as possible. The GPS records latitude and longitude with a 3 m error every 15 seconds. This GPS track can then be matched with each fish for which a time is recorded, producing fine scale location data for nearly every fish. Having access to fine scale catch locations such as this would increase the value of data gathered by CCFRP, lay groundwork for additional analysis, and elucidate groundfish habitat associations in the nearshore environment in California. Thus, the methods outlined in this document represent a novel approach to collecting and processing fine scale catch location data that can be implemented for the entire statewide CCFRP program.

Methods

This Standard Operating Procedure describes in detail how to implement this new approach for CCFRP data collection. The methods section is broken up into the following parts: (1) Set up, (2) Taking data in the field, (3) Data prior to 2024, (4) Data post 2024.

Set Up

This protocol assumes users have installed and are familiar with R and RStudio. Users need to install the necessary packages in RStudio (versions noted are those in use at the time of this writing, late 2024): tidyverse (vers 2.0.0, Wickham et al. 2019), zoo (vers 1.8-12, Zeileis & Grothendieck 2005), hms (vers 1.1.3, Muller 2023) and XML (vers 3.99-0.17, Temple Lang 2024).

```
library(tidyverse) #for data management
library(zoo) #for an.approx function
library(hms) #for the as_hms function
library(XML) #reading and extracting data from the .gpx files
library(here) #for reading in and exporting the files

#this will ensure that all the gps coordinate decimals points are shown
options(digits = 8)
```

Taking Data in the Field

Tracks: At the beginning of the day, turn the GPS unit on and navigate to the set-up menu. Navigate to the track manager and set the interval to every “15 seconds” and “recording on the map”. Throughout the day, record the time in 12-hour format with the hours, minutes and second in the comments section for as many caught fish as possible. Once the fishing day is over, navigate to the track manager and save the tracks for the day, leaving the file named with the date.

Waypoints: Once a fish is caught, select the “mark” button and then “done”, and record the waypoint number on the data sheet. Do this with as many caught fish as possible. Once the fishing day is done, navigate to the waypoint manager, select the waypoints taken that day and save them with the date name.

Note: Regardless of which method is used, it is very important that in the field fish are recorded on the datasheet in the order that is closest to the order in which they were caught. This will ensure that the catch locations assigned to each fish are as accurate as possible.

Data Prior to 2024

For data taken prior to 2024, there are two requirements that must be met, (1) the fish must be entered into the database in the same order that they are recorded on the field datasheet, and (2) there must be no infilled waypoints. Infilled waypoints are from years prior to 2024 when fish without waypoints were assigned the same location as the last fish with a recorded waypoint. These two requirements are important, because the order on the field datasheet is the closest thing to the order in which the fish were caught. The fish must be entered in this order to evenly divide them along the drift path. There cannot be infilled waypoints because this method is going to create a new waypoint for every fish that was recorded without a waypoint. If either of these requirements are not met, data cleaning is required.

Once those two requirements have been met, export the desired data from the Access database to an Excel file, with source formatting. Be sure to include the Trip ID, Area, Year Automatic, Drift ID, Start Time, End Time, Start_LatDD, Start_LonDD, End_LatDD, End_LonDD, Fish ID, Species Code, Lat Released, Lon Released and Comments columns in the exported file. Once this file is exported, save it as a csv file, rename the columns to match the following: tripid, area, year, driftid, st_time, end_time, st_lat, st_lon, end_lat, end_lon, fishid, species, waypoint_lat, waypoint_lon, and comments (Figure 1) and format the st_lat, st_lon, end_lat, end_lon, waypoint_lat and waypoint_lon columns to be numbers with five decimal places. This csv file can now be read into RStudio. The data must be arranged by drift id and fish ID to ensure that the fish are organized and read into RStudio in the order in which they were recorded in the field.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	tripid	area	year	driftid	st_time	end_time	st_lat	st_lon	end_lat	end_lon	fishid	species	fork	waypoint_lat	waypoint_lon	comments
2	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	65	VER	495	35.69472	-121.31567	8:59:53
3	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	66	BLU	178			
4	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	67	OLV	399			
5	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	68	BLU	258			
6	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	69	BLU	344	35.69473	-121.31586	9:01:15
7	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	71	OLV	399			
8	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	73	GPR	242			
9	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	74	BLU	284	35.69471	-121.31585	9:02:09
10	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	76	BLU	289			
11	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	78	BLU	345	35.69471	-121.31585	9:02:32
12	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	82	BLU	298	35.69471	-121.31586	9:02:54
13	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	83	OLV	450			
14	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	85	BLU	308	35.6947	-121.31587	9:03:38
15	BLT01_23	BL	2023	BLM07112	8:59:08 AM	9:14:16 AM	35.69469	-121.316	35.69448	-121.316	87	BLU	274			

Figure 10. An example of what the California Collaborative Fisheries Research Program (CCFRP) data should look like once it has been exported from Access and converted into a csv file. The columns species, waypoint_lat, waypoint_lon and comments are outlined in red to highlight what the data looks like before the waypoints have been interpolated. Data for this figure were collected in 2023 by the Cal Poly San Luis Obispo chapter of CCFRP.

```
CCFRP_DATAFRAME_NAME <- read.csv(here("INSERT_CCFRP_DATA_FILE_NAME_HERE.csv"))

CCFRP_DATAFRAME_NAME <- arrange(CCFRP_DATAFRAME_NAME, driftid,
             fishid)
```

Next, test whether there are any waypoints that match the end coordinates and remove them if this is true. This is to ensure that when the end coordinates are later assigned to the last fish caught on the drift, they do not occur earlier in the drift.

```
CCFRP_FINAL_DATAFRAME <- CCFRP_DATAFRAME_NAME %>%
 mutate(
  waypoint_lat = if_else(
   condition = waypoint_lat == end_lat & waypoint_lon == end_lon,
   true = NA,
   false = waypoint_lat),
  waypoint_lon = if_else(
```

```
    condition = waypoint_lat == end_lat & waypoint_lon == end_lon,
    true = NA,
    false = waypoint_lon))
```

Then, make a vector that contains all the unique Drift IDs.

```
uniquedrift <- sort(unique(CCFRP_FINAL_DATAFRAME$driftid))
```

Finally, run the data and that vector through a loop. This loop assigns the first and last fish caught on a drift the start and end coordinates, respectively. It also evenly divides the remaining fish between the start/end coordinates and any intermediate waypoints that were taken. Lastly, it returns a data frame where every caught fish has an approximate catch location. This data frame can be exported as a csv file or used as a data frame for further analysis in RStudio.

```
# Note: This loop can only be run once

if(exists('alldata')==T){rm(alldata)}

for(drift in unique_drift){

  tempdrift <- CCFRP_FINAL_DATAFRAME %>% #change this to match the final dat
aframe
    filter(driftid == drift) %>%
    mutate(fishlat = gps_lat,
           fishlon = gps_lon)

  l <- nrow(tempdrift)

  # adding the start lat and long or end lat and lon in cases where the
   # start and end fish have no waypoint

  tempdrift$fishlat[1] <- if_else(is.na(tempdrift$fishlat[1]),
                              tempdrift$st_lat[1],
                              tempdrift$fishlat[1])

  tempdrift$fishlon[1] <- if_else(is.na(tempdrift$fishlon[1]),
                              tempdrift$st_lon[1],
                              tempdrift$fishlon[1])

  tempdrift$fishlat[l] <- if_else(is.na(tempdrift$fishlat[l]),
                              tempdrift$end_lat[l],
                              tempdrift$fishlat[l])

  tempdrift$fishlon[l] <- if_else(is.na(tempdrift$fishlon[l]),
                              tempdrift$end_lon[l],
                              tempdrift$fishlon[l])
```

```
##evenly dividing waypoints across fish on a drift

gps_approx <- tempdrift %>%
  mutate(approx_lat = na.approx(fishlat),
         approx_lon = na.approx(fishlon))

if(exists('driftgps')==T){driftgps <- rbind(driftgps, gps_approx)}
if(exists('driftgps')==F){driftgps <- gps_approx}

FINAL_DATA_NAME_HERE <- driftgps
FINAL_DATA_NAME_HERE$fishlat <- NULL
FINAL_DATA_NAME_HERE$fishlon <- NULL}

# When exporting this into a csv file convert the numbers to scientific notation
```

Data Post 2024

GPS Time Stamp

Before beginning GPS time stamp data processing, download the tracks from the GPS to the computer. To do this, plug the GPS unit used in the field into the computer and copy the .gpx file(s) named with the date(s) that are to be downloaded to the computer. Then these gpx files will be read into RStudio individually and the coordinates and time will be extracted and formatted into the correct structure and placed into a data frame with the following columns: lat, lon, date and time (Figure 2).

	A	B	C	D	E	F
1	lat	lon	elevation	date_time	date	time
2	35.63675	-121.278	2.54	7/11/2023 7:07	7/11/2023	7:07:20
3	35.63746	-121.279	2.54	7/11/2023 7:07	7/11/2023	7:07:35
4	35.63807	-121.279	3.98	7/11/2023 7:07	7/11/2023	7:07:50
5	35.6387	-121.28	3.98	7/11/2023 7:08	7/11/2023	7:08:05
6	35.63935	-121.281	4.46	7/11/2023 7:08	7/11/2023	7:08:20
7	35.63998	-121.281	3.98	7/11/2023 7:08	7/11/2023	7:08:35
8	35.64061	-121.282	3.5	7/11/2023 7:08	7/11/2023	7:08:50
9	35.64125	-121.283	3.5	7/11/2023 7:09	7/11/2023	7:09:05
10	35.64187	-121.284	3.98	7/11/2023 7:09	7/11/2023	7:09:20
11	35.64253	-121.284	3.5	7/11/2023 7:09	7/11/2023	7:09:35
12	35.6432	-121.285	3.5	7/11/2023 7:09	7/11/2023	7:09:50
13	35.64373	-121.286	3.02	7/11/2023 7:10	7/11/2023	7:10:05
14	35.64431	-121.286	3.5	7/11/2023 7:10	7/11/2023	7:10:20
15	35.64487	-121.287	3.5	7/11/2023 7:10	7/11/2023	7:10:35

Figure 11. An example of what the GPS data should look like once the gpx files have been read into RStudio and extracted into a data frame. The columns species, lat, lon and time are outlined in red to highlight what the data looks like straight from the GPS. Data for this figure were collected in 2023 by the Cal Poly San Luis Obispo chapter of CCFRP.

```
extractgpx <- function(gpsfilename){

  gpx <- htmlTreeParse(file = gpsfilename,
                       useInternalNodes = TRUE)

  #getting the coordinates
  coords <- xpathSApply(doc = gpx, path = "//trkpt", fun = xmlAttrs)

  #getting the time
  gpstime <- xpathSApply(doc = gpx, path = "//trkpt/time", fun = xmlValue)

  #putting the time into date-time structure
  gpstime_r <- strptime(gpstime[1], "%Y-%m-%dT%H:%M:%SZ")

  #formatting into correct timezone
  gpstime_r <- as.POSIXct(gpstime, format = "%Y-%m-%dT%H:%M:%SZ", tz = 'UTC')

  attr(gpstime_r, 'tzone') <- 'America/Los_Angeles'

  #binding the data into a dataframe with new column names
  gps <- data.frame(
  lat = as.numeric(coords["lat", ]),
  lon = as.numeric(coords["lon", ]),
  date_time = gpstime_r) %>%
  mutate(date = date(date_time),
         time = str_extract(date_time, "[0-9]{2}:[0-9]{2}:[0-9]{2}"))

  return(gps)}


#make sure to create a new date object for each individual GPS file date you read in

INSERT_DATA_NAME_HERE1 <- extractgpx("INSERT GPS FILE NAME HERE.gpx
")

INSERT_DATA_NAME_HERE2 <- extractgpx("INSERT GPS FILE NAME HERE.gpx
")

INSERT_DATA_NAME_HERE3 <- extractgpx("INSERT GPS FILE NAME HERE.gpx
")
```

Before combining these data frames, filter to make sure that there is only one date in each of the files. This can be a problem, because sometimes the GPS stores the tracks for more than one date in a file.

```
#Repeat for every date you have - The date structure is YYYY-MM-DD
INSERT_DATA_NAME_HERE1 <- INSERT_DATA_NAME_HERE1 %>%
```

```
  filter(date == "INSERT DATE HERE")

INSERT_DATA_NAME_HERE2 <- INSERT_DATA_NAME_HERE2 %>%
  filter(date == "INSERT DATE HERE")

INSERT_DATA_NAME_HERE3 <- INSERT_DATA_NAME_HERE3 %>%
  filter(date == "INSERT DATE HERE")
```

Create a list of all the newly filtered dates; and pass it to the function bindgps. This function will bind all the GPS files, and round the time to the nearest 15 seconds. The CCFRP time will also be rounded later, and this will ensure that the time associated with the fish will match a time associated with the GPS. This function will also convert the time into 12-hour format to match the CCFRP time. It then creates a new column called date_rdtime, this column combines the new rounded time and the date. This will be the column that will join the CCFRP and GPS data.

```
bindgps <- function(listname){

  gps_tracks <- bind_rows(listname) %>%
    mutate(time = as_hms(time),
           date = as_date(date)) %>%
    mutate(round_time = round_hms(time, secs = 15)) %>%
    mutate(round_time = as.numeric(str_remove_all(round_time, ":")),
           round_time = if_else(
             condition = round_time >= 130000 & round_time < 140000,
             true = round_time - 120000,
             false = round_time),
           round_time = if_else(
             condition = round_time >= 140000,
             true = round_time - 120000,
             false = round_time),
           round_time = gsub("(\\d\\d)(\\d\\d)(\\d\\d)",
                             "\\1:\\2:\\3",
                             round_time),
           round_time = gsub("(\\d)(\\d\\d)(\\d\\d)",
                             "\\0\\1:\\2:\\3",
                             round_time)) %>%
    unite("date_rdtime", date, round_time, sep = "_")

  return(gps_tracks)}

#include all the dates you have
INSERT_LIST_NAME_HERE <- list(INSERT_DATA_NAME_HERE1, INSERT_DA
TA_NAME_HERE2,
                INSERT_DATA_NAME_HERE3)

INSERT_GPS_DATA_NAME_HERE <- bindgps(INSERT_LIST_NAME_HERE)
```

To prepare the CCFRP data, export the data from the Access database to an excel file, be sure to keep source formatting, with the following columns: Trip ID, Area, Year Automatic, Drift ID, Start Time, End Time, Start_LatDD, Start_LonDD, End_LatDD, End_LonDD, Fish ID, Species Code, Lat Released, Lon Released and Comments. Once this file is exported, save it as a csv file, rename the columns to match the following: tripid, area, year, driftid, st_time, end_time, st_lat, st_lon, end_lat, end_lon, fishid, species, waypoint_lat, waypoint_lon, and comments (Figure 1) and format the st_lat, st_lon, end_lat, end_lon, waypoint_lat and waypoint_lon columns to be numbers with five decimal places. Now run the data through the cleanfish function. This function will read in the csv file, create a date column, extract the time from the comments section and round it to the nearest 15 seconds to match the GPS data, it will then create a new column called date_rdtime, this column combines the new rounded time and the date. This will be the column that will join the CCFRP and GPS data.

```
cleanfish <- function(ccfrpfilename){

  ccfrpdat <- read.csv(ccfrpfilename)

  ccfrpdat <- ccfrpdat %>%
    mutate(date = str_extract(driftid, pattern = "\\d\\d\\d\\d\\d\\d"),
           date = mdy(date)) %>%
    mutate(twodighour = str_extract(comments,
                                    pattern ="\\d\\d\\:\\d\\d:\\d\\d"),
           onedighour = str_extract(comments,
                                    pattern = "\\d\\:\\d\\d:\\d\\d")) %>%
    mutate(time = if_else(
      condition = is.na(twodighour),
      true = onedighour,
      false = twodighour)) %>%
    select(-c(onedighour, twodighour)) %>%
    mutate(time = as_hms(time)) %>%
    mutate(round_time = round_hms(time, secs = 15)) %>%
    unite("date_rdtime", date, round_time, sep = "_")

  return(ccfrpdat)}

INSERT_CCFRP_DATA_NAME_HERE <- cleanfish("INSERT CCFRP FILE NAME
HERE.csv")
```

To join the CCFRP and GPS data by the date_rdtime column, feed the CCFRP data object and GPS object created above into the join function. This function will create one new data frame with all the data from the CCFRP and GPS data, as well as two new columns, gps_lat and gps_lon which will have the latitude and longitude recorded on the GPS track.

```
join <- function(ccfrpdata, gpsdata){
```

```
  final <- left_join(ccfrpdata, gpsdata,
                     by = "date_rdtime") %>%
    select(-c("date_time")) %>%
    separate_wider_delim(date_rdtime, delim = "_",
                         names = c("date", "round_time")) %>%
    rename(gps_lat = lat,
           gps_lon = lon)

  return(final)}

INSERT_FINAL_DATA_NAME_HERE <- join(INSERT_CCFRP_DATA_NAME_HE
RE,
                                    INSERT_GPS_DATA_NAME_HERE)
```

After obtaining GPS coordinates for all fish with time stamps, give the remaining fish approximate catch locations using the same methods that were outlined in the data prior to 2024 section. Arrange the data by drift id and fish id to ensure that the fish are organized and read into RStudio in the order in which they were recorded in the field. Next, test whether there are any waypoints that match the end coordinates and remove them if this is true. This is to ensure that when the end coordinates are later assigned to the last fish caught on the drift, they do not occur earlier in the drift.

```
# Comment out the mutate function if fish are not processed after the end of the drift
# It is included if multiple fish got the end coordinate but were not caught at the end

sort.prep <- function(joineddata){

  final <- arrange(joineddata,
                   driftid,
                   fishid) %>%
    mutate(gps_lat = if_else(condition = gps_lat == end_lat
                             & gps_lon == end_lon,
                             true = NA,
                             false = gps_lat),
           gps_lon = if_else(condition = gps_lat == end_lat
                             & gps_lon == end_lon,
                             true = NA,
                             false = gps_lon))

  return(final)}

INSERT_FINAL_DATA_NAME_HERE <- sort.prep(INSERT_FINAL_DATA_NAME
_HERE)
```

Then, make a vector that contains all the unique Drift IDs. Finally, run the data and that vector through a loop. This loop assigns the first and last fish caught on a drift the start and end coordinates, respectively. It also evenly divides the remaining fish between the start/end coordinates and any intermediate waypoints that were taken. Lastly, it returns a data frame where every caught fish has an approximate catch location. This data frame can be exported as a csv file or used as a data frame for further analysis in RStudio.

```
# Note: This loop can only be run once

#getting each unique driftid
unique_drift <- unique(INSERT_FINAL_DATA_NAME_HERE$driftid)

if(exists('alldata')==T){rm(alldata)}

for(drift in unique_drift){

 tempdrift <- INSERT_FINAL_DATA_NAME_HERE %>%
  filter(driftid == drift) %>%
  mutate(fishlat = gps_lat,
      fishlon = gps_lon)

 l <- nrow(tempdrift)

 # adding the start lat and long or end lat and lon in cases where the
 # start and end fish have no waypoint

 tempdrift$fishlat[1] <- if_else(is.na(tempdrift$fishlat[1]),
                  tempdrift$st_lat[1],
                  tempdrift$fishlat[1])

 tempdrift$fishlon[1] <- if_else(is.na(tempdrift$fishlon[1]),
                  tempdrift$st_lon[1],
                  tempdrift$fishlon[1])

 tempdrift$fishlat[l] <- if_else(is.na(tempdrift$fishlat[l]),
                  tempdrift$end_lat[l],
                  tempdrift$fishlat[l])

 tempdrift$fishlon[l] <- if_else(is.na(tempdrift$fishlon[l]),
                  tempdrift$end_lon[l],
                  tempdrift$fishlon[l])

 ##evenly dividing waypoints across fish on a drift

 gps_approx <- tempdrift %>%
  mutate(approx_lat = na.approx(fishlat),
      approx_lon = na.approx(fishlon))

 if(exists('driftgps')==T){driftgps <- rbind(driftgps, gps_approx)}
 if(exists('driftgps')==F){driftgps <- gps_approx}
```

```
  FINAL_DATA_NAME_HERE <- driftgps
  FINAL_DATA_NAME_HERE$fishlat <- NULL
  FINAL_DATA_NAME_HERE$fishlon <- NULL}

# When exporting the data to a csv convert the numbers to scientific notation
```

GPS Waypoints

Before beginning GPS waypoint data processing, download the waypoints from the GPS to the computer. Plug the GPS unit used in the field into the computer and copy the .gpx file(s) named with the date(s) that are to be downloaded to the computer. Then these gpx files will be read into RStudio individually and the coordinates and waypoint names, and date will be extracted and formatted into the correct structure and placed into a data frame with the following columns: lat, lon, date, waypoint and wpt_date. The wpt_date column is a combination of the waypoint name and the date on which it was taken. This is an important column as it will be the column by which the gps waypoint data is joined to the CCFRP data.

```
extractgpx.wp <- function(waypointfilename){

  gpx <- htmlTreeParse(file = waypointfilename,
                       useInternalNodes = TRUE)

  #getting the coordinates
  coords <- xpathSApply(doc = gpx, path = "//wpt", fun = xmlAttrs)

  #getting the date
  time <- xpathSApply(doc = gpx, path = "//wpt/time", fun = xmlValue)

  #getting the name
  name <- xpathSApply(doc = gpx, path = "//wpt/name", fun = xmlValue)

  #putting the time into date-time structure
  gpstime_r <- strptime(time[1], "%Y-%m-%dT%H:%M:%SZ")

  #formatting into correct timezone
  gpstime_r <- as.POSIXct(gpstime_r, format = "%Y-%m-%dT%H:%M:%SZ", tz = 'UT
C')

  attr(gpstime_r, 'tzone') <- 'America/Los_Angeles'

  #binding the data into a dataframe with new column names
  waypoint <- data.frame(
  lat = as.numeric(coords["lat",]),
  lon = as.numeric(coords["lon",]),
  date_time = gpstime_r,
```

```
  waypoint = name) %>%
  mutate(date = date(date_time),
         date = str_remove_all(date, "-"),
         wpt_date = str_glue("{date}_{waypoint}"))

  return(waypoint)}


#make sure to create a new date object for each individual GPS waypoint file date you re
ad in

INSERT_DATA_NAME_HERE1 <- extractgpx.wp("INSERT GPS FILE NAME HERE.
gpx")

INSERT_DATA_NAME_HERE2 <- extractgpx.wp("INSERT GPS FILE NAME HERE.
gpx")

INSERT_DATA_NAME_HERE3 <- extractgpx.wp("INSERT GPS FILE NAME HERE.
gpx")
```

To import and clean the CCFRP fish data, export the desired data from the Access database to an excel file, be sure to keep source formatting, with the following columns: Trip ID, Area, Year Automatic, Drift ID, Start Time, End Time, Start_LatDD, Start_LonDD, End_LatDD, End_LonDD, Fish ID, Species Code, Lat Released, Lon Released and GPS Waypoint. Once this file is exported, save it as a csv file, rename the columns to match the following: tripid, area, year, driftid, st_time, end_time, st_lat, st_lon, end_lat, end_lon, fishid, species, waypoint_lat, waypoint_lon, and gps_waypoint and format the st_lat, st_lon, end_lat, end_lon, waypoint_lat and waypoint_lon columns to be numbers with five decimal places. Run the data through the function cleanfish.wp. This function will read in the csv file, create a date column, and create a new column called wpt_date, this column combines the date and the waypoint name. This will be the column that will join the CCFRP and GPS waypoint data.

```
cleanfish.wp <- function(ccfrpfilename){

  ccfrpdat <- read.csv(ccfrpfilename)

  ccfrpdat <- ccfrpdat %>%
    mutate(date = str_extract(driftid, pattern = "\\d\\d\\d\\d\\d\\d"),
           date = mdy(date),
           date= str_remove_all(date, "-"),
           wpt_date = if_else(is.na(gps_waypoint),
                              NA,
                              str_glue("{date}_{gps_waypoint}"))) %>%
  select(-c(date))

  return(ccfrpdat)}
```

```
INSERT_CCFRP_DATA_NAME_HERE <- cleanfish.wp("INSERT CCFRP FILE NAM
E HERE.csv")
```

To join the CCFRP and GPS data by the wpt_date column, feed the CCFRP data object and GPS object created above into the join.wp function. This function will join the two data frames and create two new columns, gps_lat and gps_lon which will have the latitude and longitude recorded on the GPS track.

```
join.wp <- function(ccfrpdata, wpdata){

  final <- left_join(ccfrpdata, wpdata,
               by = "wpt_date") %>%
  select(-c("waypoint", "date_time", "date")) %>%
  rename(gps_lat = lat,
        gps_lon = lon)

  return(final)}

INSERT_FINAL_DATA_NAME_HERE <- join(INSERT_CCFRP_DATA_NAME_HE
RE,
                              INSERT_GPS_DATA_NAME_HERE)
```

After obtaining GPS coordinates for all fish with waypoints, give the remaining fish approximate catch locations using the same methods that were outlined in the data prior to 2024 section. Arrange the data by drift id and fish id to ensure that the fish are organized and read into RStudio in the order in which they were recorded in the field. Next, test whether there are any waypoints that match the end coordinates and remove them if this is true. This is to ensure that when the end coordinates are later assigned to the last fish caught on the drift, they do not occur earlier in the drift.

```
sort.prep <- function(joineddata){

  final <- arrange(joineddata,
               driftid,
               fishid) %>%
   mutate(gps_lat = if_else(condition = gps_lat == end_lat
                    & gps_lon == end_lon,
                    true = NA,
                    false = gps_lat),
         gps_lon = if_else(condition = gps_lat == end_lat
                    & gps_lon == end_lon,
                    true = NA,
                    false = gps_lon))

  return(final)}
```

```
INSERT_FINAL_DATA_NAME_HERE <- sort.prep(INSERT_FINAL_DATA_NAME
_HERE)
```

Then, make a vector that contains all the unique Drift IDs. Finally, run the data and that vector through a loop. This loop assigns the first and last fish caught on a drift the start and end coordinates, respectively. It also evenly divides the remaining fish between the start/end coordinates and any intermediate waypoints that were taken. Lastly, it returns a data frame where every caught fish has an approximate catch location. This data frame can be exported as a csv file or used as a data frame for further analysis in RStudio.

```
#Note: This loop can only be run once

#getting each unique driftid
unique_drift <- unique(INSERT_FINAL_DATA_NAME_HERE$driftid)

if(exists('alldata')==T){rm(alldata)}

for(drift in unique_drift){

  tempdrift <- INSERT_FINAL_DATA_NAME_HERE %>%
    filter(driftid == drift) %>%
    mutate(fishlat = gps_lat,
           fishlon = gps_lon)

  l <- nrow(tempdrift)

  # adding the start lat and long or end lat and lon in cases where the
  # start and end fish have no waypoint

  tempdrift$fishlat[1] <- if_else(is.na(tempdrift$fishlat[1]),
                                  tempdrift$st_lat[1],
                                  tempdrift$fishlat[1])

  tempdrift$fishlon[1] <- if_else(is.na(tempdrift$fishlon[1]),
                                  tempdrift$st_lon[1],
                                  tempdrift$fishlon[1])

  tempdrift$fishlat[l] <- if_else(is.na(tempdrift$fishlat[l]),
                                  tempdrift$end_lat[l],
                                  tempdrift$fishlat[l])

  tempdrift$fishlon[l] <- if_else(is.na(tempdrift$fishlon[l]),
                                  tempdrift$end_lon[l],
                                  tempdrift$fishlon[l])

  ##evenly dividing waypoints across fish on a drift
```

```
gps_approx <- tempdrift %>%
  mutate(approx_lat = na.approx(fishlat),
         approx_lon = na.approx(fishlon))

if(exists('driftgps')==T){driftgps <- rbind(driftgps, gps_approx)}
if(exists('driftgps')==F){driftgps <- gps_approx}

FINAL_DATA_NAME_HERE <- driftgps
FINAL_DATA_NAME_HERE$fishlat <- NULL
FINAL_DATA_NAME_HERE$fishlon <- NULL}

# When exporting the data to a csv convert the numbers to scientific notation
```

Results

This proposed method was tested by the San Luis Obispo chapter of CCFRP in the summer of 2023. Both the traditional field method where the latitude, longitude and time were written, and the new time stamp method where only the time was written down while the GPS recorded tracks were done to test the differences between the two methods. There were 675 waypoints taken using the traditional field method, while there were 1,716 with the new time stamp method. In addition, to being easier to do in the field, the new method is more accurate compared to the previous method; see Figure 3A-B.

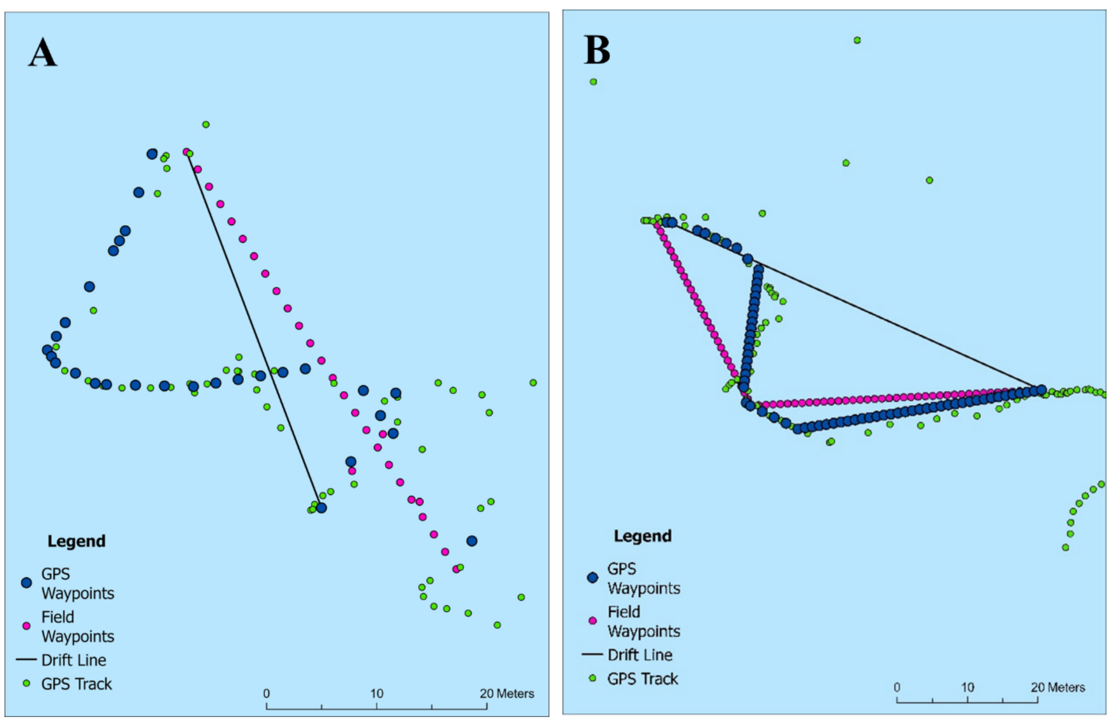

Figure 12. The difference in the data collected using (A) traditional field waypoint method and (B) the new GPS time stamp method. Data for these figures were collected in 2023 by the Cal Poly San Luis Obispo chapter of CCFRP. The blue points represent the fish catch locations obtained using the GPS time stamp method, the pink

points represent the fish catch locations obtained using the traditional field waypoint method, the black line represents the drift line from the start coordinates of the drift to the end coordinates and the green points represent the GPS tracks that were taken every 15 seconds.

Discussion

Having tried both the traditional field method and the new GPS time stamp method, it is apparent that the GPS time stamp method is not only easier to do in the field, but it is also more accurate with regard to fish catch location data. Thus, going forward, CCFRP will continue to use either the new GPS time stamp method, or the GPS waypoint method, if that is preferred, to continue collecting more accurate fish location data. This document includes all the code and instructions necessary to carry out these new methods, as well as how to convert older CCFRP data into a format which can also be used for spatial analyses. This document will add to the already incredibly useful and informative data that CCFRP continues to collect on rockfish species along the entire coast of California.