Methods to utilize known habitat to filter data for indices of abundance from a recreational fishery survey in California

Melissa Hedges Monk[a,1,*], Rebecca R. Miller[b], Grant Waltz[1], Dean Wendt[c]

[a]*Southwest Fisheries Science Center, 110 McAllister Way, Santa Cruz, 95060,*

[b]*University of California Santa Cruz, Street Address, Santa Cruz, 95060,*

[c]*California Polytechnic State University, Street Address, San Luis Obispo, 93407,*

Abstract

This is the abstract.

Keywords: fisheries dependent data, habitat association, groundfish, index of abundance

---

1. Introduction

Melissa will refocus the introduction - I keep changing my mind.

Integrated fisheries stock assessment models utilize a variety of data sources to develop the most complete picture of the stock and current status in relation to management thresholds.

Catch per unit effort is one

More often(TYPICALLY??), an index of abundance is used as a relative measure of the population and requires a time series to inform the stock assessment model. An index of

*Corresponding author

Email addresses: melissa.monk@noaa.gov (Melissa Hedges Monk), rebecca.miller@noaa.gov (Rebecca R. Miller), cat@example.com (Grant Waltz), cat@example.com (Dean Wendt)

[1]This is the first author footnote.

relative abundance assumes that changes in the index are proportional to changes of abundance in the population (Harley et al., 2001). Fishery-independent data collected from standardized survey designs provide a more unbiased estimation of the trend in a fisheries population. However, fishery-independent surveys are costly, labor intensive and often require a long time series to be considered informative in fisheries stock assessments. In an ideal situation, both fishery-dependent and fishery-independent surveys would be used to inform the stock assessment model. However, there are cases when only fishery-dependent data are available and the caveats of each data stream must be considered (cite assessments). Fishery-dependent data are collected directly from the the fishery and are less costly than (than what????) the whose operations are not constrained by sampling designs, but dependent on the behaviors of the captain and vessel and, in the case of recreational trips, customer preference.

Fishery-dependent surveys sample the fishing fleets and are subject to potential sampling biases. The sampling is dependent on the fishing boat's behavior, which is to maximize catch. Sampling of the fishing fleet is often opportunistic based on the availability of samplers and the availability of trips to sample. Sampling the fisheries can also be constrained to the current regulations, which may prohibit the retention of a species or fishing at certain depths, i.e., California Department of Fish and Wildlife has varying spatial and temporal depth and season closures implemented through six management regions. There is also a network of Marine Protected Areas (MPAs) designated from 2007-2012 that prohibit recreational fishing, and are therefore areas no longer sampled by the recreational fishing fleet.

Catch per unit effort (CPUE) is a common metric available from fishery-dependent surveys (Maunder and Punt, 2004).

A common characteristic of ecological data is a high proportion of zero observations across samples and the question as to whether the sampling occurred within the species' habitat and the species was not observed or if the sampling occurred outside of the species' habitat (structural zeroes). **you might mention someplace in introduction that nearshore rockfish are strongly associated with rocky habitat– perhaps here. Also, consider this sentence/paragraph as the opening paragraph??** Fisheries survey data are often subset to exclude structural zeroes using the Stephens-MacCall method (2004), which models the probability of observing the target species given the other the presence/absence of other species. However, the onboard observer survey collected location-specific information on each fish encountered. To subset the onboard observer survey data and exclude structural zeroes,we used the positive catch locations (I DON"T TOTALLY UNDERSTAND THIS SENTENCE- only postiive catches are included?) this sentence seems like it is in the wrong place? more of a methods?) as a proxy for suitable habitat.

To explore the changes in data filtering related to structural zeroes, we utilized high resolution fishery sampling and and bathymetry data, we evaluated data from a recreational party boat onboard observer survey, which collects location- and species-specific CPUE information from the commercial passenger fishing vessel (CPFV; also know as party boat) fleet (Monk et al., 2014). The data were collected at the level of a fishing drift and fine-scale habitat data are available for a large fraction of California state waters. Paired with recently available high-resolution bathymetry data provided an opportunity to overlay each individual fishing drift onto known habitat type (hard vs. soft substrate), and has been the method utilized for stock assessments since 2015. To explore how data selection methods and the resulting indices would change if the data were only available at a courser resolution, we used the same data set to develop standardized indices of relative abundance based on three different data filtering methods. We applied these methods across six nearshore rockfish species with different life histories,habitat preferences and commonness in the data.

The three data treatment methods included filtering the drift-level data based on known location, i.e., the status quo **(what is status quo? StevensMcCall?)**, treating the drift-level data as if the location of the drifts were not available, and lastly, an aggregating of the catches at the drift-level data to a trip. In addition, for the model filtered based on known rocky reef habitat, we weighted the index by area of habitat within pre-defined regions. For the two cases where we removed the location information, we filtered the data using the Stephens-MacCall method.

MOVE TO INTRODUCTION: It is not often the case where high-resolution habitat and fishing location information are both available, and for many fishery-dependent surveys an analyst will have to determine which subset of the data to use based on available information. The onboard observer data provide an opportunity to explore what information we gain from explicit knowledge of fishing locations.

The Stephens-MacCall (2004) filtering approach was used to predict the probability of encountering the target species, based on the species composition of the catch in a given trip. The method uses presence/absence data within a logistic regression to identify the probability of encountering a target species given the presence or absence of other predictor species. This method is commonly used to filter data that were collected dockside after a vessel returned to port or when location data are not provided.

Prior to 2013, these data had not been used to develop an index of abundance for West Coast fisheries stock assessments. That was partially due to the availability of data and at that point in time not many full stock assessments had been conducted for the nearshore groundfish species. These data were first used without the habitat data in a suite of 2013 data moderate stock assessments (**?**), where data were filtered using positive species observations and an alpha hull method commonly used to estimate species home range sizes and distribution models for terrestrial species (**??**)

Include a paragaph in the introduction that briefly summarizes the purpose of the S-M method, i.e. using species composition of the catch to identify effective fishing effort for the target species, and cite it there. That way, you can mention it in the methods and the reader will already be familiar with it.

cut from another place: The onboard observer survey data provide a high-resolution of catch, effort and the ability to map the fishing drifts to fine-scale habitat data. This paper explores methodological differences in data treatment to see what we gain by having the high-resolution habitat data and using that as a mechanism to filter out trips that are not targeting the species of interest

This paper explores methodological differences in data treatment to determine changes in *trends in indices and the associated error among three alternative assumptions and data filtering strategies. All of the methods described below started with the same subset of drifts from the onboard observer survey data, restricted to state waters and the years 2004-2016. In the case of application to stock assessments, all potential data are explored, which may be why trends in indices differ in this paper than what has previously been published in stock assessments. Since the most recent stock assessments in 2021, the data have undergone a major quality assurance effort by the authors.*

2. Methods

We developed indicies of abundance for six species or species pairs of rockfish (*Sebastes spp.*) that are of management interest on the U.S. West Coast: black rockfish (*S. melanops*), the blue and deacon rockfish (*S. mystinus*/*S. diaconus*), brown rockfish (*S. auriculatus*), China rockfish (*S. nebulosus*), gopher rockfish (*S. carnatus*), and the vermilion and sunset rockfish (*S. miniatus*/*S. crocotulus*). The two cryptic species pairs (blue/deacon and sunset/vermilion rockfish) are genetically identifiable, but not separable within the onboard observer survey time series. These six species all have different latitudinal distributions, exploitation histories, and habitat and depth preferences(**?**).

2.1. Survey Data and Habitat-based Filtering

The California Department of Fish and Wildlife (CDFW) began a fishery-dependent onboard observer survey of the Commercial Passenger Fishing Vessel (CPFV or party/charter boat) fleet in 1999. In 2004, the survey became part of the CDFW's California Recreational

Fisheries Survey (CFRS, *add year and cite website*) that includes additional surveys to quantify catch and effort by the recreational fleet. In response to a request from the fishing industry, the California Polytechnic State University Institute of Marine Science, San Luis Obispo (Cal Poly) began a supplemental onboard observer survey in 2001 of the CPFV fleet based in Port Avila and Port San Luis along the Central Coast [#fig-map]. Both the CDFW and the Cal Poly onboard observer surveys continue through present day; however, due to both spatial and temporal recreational regulation changes we limited the data for this research to the years 2004 to 2016. Between 1999 and 2003, the recreational regulations evolved from no restriction on the number of lines or hooks an angler could deploy to a one line and two-hook maximum, as well as implementation of depth restrictions. Subsequent management allowed a relaxation of depth restrictions beginning in 2017, potentially shifting fishing effort relative to the 2004-2016 period (**?**).

While only a small portion of the total CPFV trips taken are sampled as part of the onboard observer survey, the onboard observer survey collects a large amount of data during each trip. During each trip the sampler records information for each fishing drift, defined as a period starting when the captain announced "lines down" to when the captain instructs anglers to reel their lines up. Just prior to the start of each fishing drift, the sampler selected a subset of anglers to observe, at a maximum of 15 anglers per drift. The sampler records all fish encountered (retained and discarded) by the subset of anglers as a group, i.e., catch cannot be attributed to an individual angler. Samplers also record the start and end times of a drift, location of the fishing drift (start latitude/longitude and for most drifts, end latitude/longitude), and minimum and maximum bottom depth. Fish encountered by the group of observed anglers are recorded as either retained or discarded. This provides information on the catch (count of each species) and effort (time and number of anglers fished) during each fishing drift. While both surveys include records of discarded fish, we only used the retained catch in these analyses. Discarded fish can often represent a different size structure than retained fish, either due to size limits or angler preference, or represent fish encountered during a temporal or spatial closure.

The SWFSC developed a relational database for the CDFW onboard survey from 1999-2010(2014) that has been updated annually. The Cal Poly data are also provided to the SWFSC annually. All data were checked for potential errors at the drift-level by SWFSC staff.

The CPFV data included only areas north of Point Conception ($34°27'N$) due to gaps in habitat coverage further south. To further remove drifts that may not accurately define a successful fishing drift or represent data errors, the upper and lower 1% of the recorded time fished and recorded observed anglers were removed. Given that the fishery was closed deeper than 40 fathoms for the entire time period from 2004-2016, we filtered the data to retain 99% of all drifts based on average drift depth. We calculated average depth from the recorded minimum and maximum depths when available or the imputed minimum and maximum depth from the bathymetry layer described in the next paragraph. A depth cutoff slightly deeper than the maximum allowed is reasonable given the variability in habitat fished and all retained drifts occurred within California state waters (up to 3 nm from shore).

High resolution seafloor mapping data allowed us to map each drift from the onboard observer surveys with predicted habitat (referred throughout the paper as the drift-level, habitat-informed data). We utilized the bathymetry and backscatter data collected by the California Seafloor Mapping Program (CSMP) (**?**). The CSMP mapped California state waters at a 2 m resolution north of of Point Conception to the California-Oregon border. A total of 137 CSMP substrate blocks that ranged in size from 16 km^2 to more than 400 km^2 were mosaicked together by authors. Rough and smooth substrates were identified by CSMP using two rugosity indices, surface:planar area, and vector ruggedness measure (VRM) of the bathymetric digital elevation model [#fig-map2]. The CSMP set a varying VRM threshold for each of the substrate blocks, removed any artifacts, and is considered a conservative estimate of rough habitat.

The 137 CSMP substrate raster blocks were then mosaicked together by authors, and converted the pixels designated as rough habitat (rocky habitat proxy) from a raster format to polygons, and calculated a 5 m buffer around the rough habitat polygon to allow for any small errors in positional accuracy using ArcMap 10.7 (ESRI citation). The area of each reef polygon was calculated, and those reefs greater than or equal to 100 m^2 were included. Contiguous polygons identified as rocky substrate were defined as a singular rocky reef, regardless of size. The area of rocky habitat for this paper was calculated to exclude portions of the reef that extended outside of California state waters (further than 3 nm from shore). The mapped area does not include very shallow areas close to shore, which extend approximately 200-500 m from the shoreline. Fishing by the CPFV fleet is limited in these waters due to shallow depths and kelp beds. We assigned fishing drifts to reefs based on the recorded start location of a drift, given that the end locations of drifts were not always recorded. The distance from the recorded drift start location to the nearest rocky habitat was calculated in meters. For each target species, we calculated the cumulative distribution of distance to rocky reef for drifts that retained the target species and used a distance cutoff of 90% for each species. To illustrate the similarities and differences among the six species, we plotted the percent of fishing drifts within an aggregated region that where the species was present and retained. To show the differences in the general commonness or rarity of the species we calculated the average CPUE, before standardization, for each species and aggregated area. We also downloaded the effort estimates for the CPFV trips from RecFIN to compare the the the area of rocky habitat with the effort in each region as well as the distribution of observed trips.

2.2. Stephens-MacCall Data Filtering

We applied the Stephens-MacCall method to both the drift-level data and the trip-level data (2004). For the drift-level data we removed all location and depth identifiers for a drift and kept the county of landing as a spatial identifier. To construct a data set that mimicked trip-level data, we took the drift-level data, aggregated the observed retained catch within a trip, and kept the county of landing as a spatial identifier. We then compared results using two levels of aggregation (catch rates by drift and trip) to illustrate the impact of having less spatially-explicit data on both data filtering and the resulting indices of abundance.

Prior to any filtering a total of 19,425 drifts that aggregated to 2,270 trips were available for the analyses. The number of initial samples used for the Stephens-MacCall filtering method were higher than the habitat-informed data described in the previous section because retained drifts with missing locations (latitude/longitude).

Before applying the Stephens-MacCall method, we identified a suite of potentially informative predictor species for each of the six target species. Species that never co-occurred with the target species and those present in fewer than 1% of all drifts and 3% of all trips were removed to reduce the number of species to those that were informative. A lower threshold of 1% was selected for the drift-level data due to the change in magnitude of the number of samples when using drifts vs trips.

The remaining species all co-occurred with the target species in at least one trip and were retained for the Stephens-MacCall logistic regression. Coefficients from the Stephens-MacCall analysis (a binomial generalized linear model) were positive for species that are more likely to co-occur with the target species, and negative for species that were less likely to be caught with target species. The intercept represented the probability of observing only the target species in a sample. We also calculated the 95% confidence interval for each coefficient.

Stephens and MacCall proposed filtering (excluding) samples from index standardization based on a criterion of balancing the number of false positives and false negatives from the predicted probability of encounter. False positives (FP) are trips that are predicted to encounter the target species based on the species composition of the catch, but did not. False negatives (FN) are trips that were not predicted to encounter the target species, given the catch composition, but caught at least one target species. Stephens and MacCall recommended a threshold where the false negatives and false positives are equally balanced, however, this threshold does not have any biological relevance and for this particular data set where trained samplers identify all fish. We assumed that if the target species was encountered, the vessel fished in appropriate habitat.

Of interest for the index of abundance was the elimination of trips that had a low probability of catching the target species given the other species caught on the trip. Therefore, we retained all of the trips that caught the target species and those trips that did not catch the target species, but had a probability higher than the threshold balancing the false negatives and false positives. This practice has commonly been used in recent stock assessments of rockfish on the West Coast.

2.3. Indices of Abundance

Four standardized indices of abundance were generated for each of the six species, one each for the data filtering method (drift-level habitat-informed, drift-level Stephens-MacCall, trip-level Stephens-MacCall) and an area-weighted index from the habitat-informed drift-level data. All indices were modeled using Bayesian generalized linear models (GLMs) and the delta GLM method (Lo et al., 1992; **?**). The delta GLM method is commonly used to standardize catch-per-unit effort for stock assessments [citations]. The delta method models the the data with two separate GLMs; one for the probability of encountering the species of interest from a binomial likelihood and a logit link function and the second models the positive encounters with either gamma or lognormal error structure. The error structure of the positive model was selected via the Akaike Information Criterion (AIC) from models with the full suite of considered explanatory variables.

The response variable for the positive models was angler-retained catch per unit effort. For the indices modeled at the level of a drift, effort was calculated as the number of angler hours fished on a drift. The trip-level effort was calculated as angler days, using the average number of observed anglers across all drifts on a trip.

To keep comparisons across data filtering methods similar, depth was not considered as an explanatory variable in the habitat-informed index. Depth is often a significant explanatory variable for rockfish species, with many rockfish species and populations separated by depth (**?**). Year was always included in as an explanatory variable in model selection, even if it was not significant, because the goal of the index of abundance was to extract the year effect. Other explanatory variables considered for the habitat-informed index were aggregated regions rocky reefs (categorical variable*)* and wave (a 3-month aggregated period of time, e.g., January-March). The area-weighted index also included a year/rocky reef interaction term, even if it was not statistically significant, to allow us to weight the index by the area of rocky reef. The regions of rocky reef were aggregated differently for each species to ensure adequate sample sizes to explore the year/rocky reef interaction.

Explanatory variables for the two indices using the data filtered using Stephens-MacCall method (blind to habitat information at the drift- and trip-level) included only year, wave and aggregated counties of landing. California has 14 coastal counties north of Point Conception, 11 of which were represented in these data. We aggregated the northern counties of Del Norte, Humboldt and Mendocino into one region, Sonoma and Marin counties just north of San Francisco into another region and Alameda and San Francisco counties into a third region. The remaining counties of San Mateo, Santa Cruz, Monterey and San Luis Obispo remained unaggregated.

Model selection for the binomial and positive observation models was based on AIC using the lme4 package in R, and unless very different predictors were selected, the same predictors were used in each of the two Bayesian models. The Bayesian models were run with 5,000 iterations and weakly informative priors. Posterior predictive model checks were examined for both the binomial and positive observation models, including the predicted percent positive compared to the maximum likelihood estimates. We constructed the final year index by multiplying the back-transformed posterior draws from the binomial model with the exponentiation of positive model draws, and taking the mean and standard deviation for each year.

The area-weighted habitat-informed index was developed by extracting the posterior draws of from each year and area combination of the binomial and positive posterior predictions, and then summing across the product of the back-transformed posteriors weighted by the fraction of total area within each reef. To compare the indices across the three data filtering methods and the area-weighted index, each index was scaled to its mean value.

3. Results

3.1. Survey Data and Habitat-based Filtering

The data sets were filtered for errors within the relational database before analyses and the data used here reflects changes from the QA/QC process that may not be reflected in data available directly from the CDFW. Approximately 21% of all the CDFW observed CPFV trips from 2004-2016 occurred north of Point Conception and it is important to note that north of Bodega Bay, California the majority of charter boats are smaller 6-pack vessel that may not have the capacity to carry a sampler onboard. The addition of the Cal Poly onboard observer survey to the CDFW survey increased the sample sizes of observed trips in San Luis Obispo county by an average of 155% from 2004-2016.

From 2004-2016 the drift-level data contained a total of 19,425 fishing drifts, and after removing drifts with missing effort information (time fished and/or observed anglers), 19,180 drifts remained. The filter for fishing drifts and observed anglers resulted in fishing drifts lasting between three and 96 minutes and three to 15 observed anglers, and reduced the data to 18,591 fishing drifts. The remaining data filter for depth resulted in a cutoff of 46.6 fathoms, and retained 18,405 drifts based on average drift depth.

We defined 108 areas of rocky habitat at the finest scale within California state waters from Point Conception to the California/Oregon border. The 2 m resolution of the substrate shows the patchiness and heterogeneity of the rocky substrate (insets A and B of [#fig:map]). While the location-specific data from the fishing fleet is governed by confidentiality, a high proportion of the fishing drifts were associated with rocky habitat. This was verified by the distributions of the distance from rocky habitat for each of the six species. The distance from rocky habitat cutoff for blue, China and gopher rockfish was six meters, eight meters for vermilion rockfish, 14 meters for black rockfish and 16 meters for brown rockfish. This final data filter resulted in xxx drifts for blue, China and gopher rockfish, zxxxx for vermilion rockfish, xxx for black rockfish and xx for brown rockfish.

After exploratory analyses and considering the the availability of data, the areas rocky habitat were grouped into five regions to ensure adequate sample sizes for developing indices of abundance (Figure 2). While covering a small area (5% of the rocky habitat), the number of observed fishing drifts within state waters around the Farallon Islands off the coast of San Francisco was high enough to warrant keeping it as a separate area of rocky habitat. The region defined from the California/Oregon to San Francisco encompasses 49% of the total rocky habitat in state waters by area, but only 12% of the observed drifts fished in this area. Each of the four remaining regions of rocky habitat defined from San Francisco to Point Conception contained an average of 12% of the available habitat. The CDFW estimated fishing effort by district, which does not exactly align with our areas of grouped reef habitat. Only considering the fishing effort north of Point Conception, CDFW estimated an average of 9% of the CPFV from the California/Oregon border through Mendocino County, 38% from Sonoma through San Mateo County, and 53% from Santa Cruz to Point Conception.

The differences in latitudinal distribution of the six species is apparent from the maps of percent of positive observations (Figure 3). Black rockfish are distributed north fo San Francisco, a more northerly distribution reflected in the aggregation of data from Santa Cruz and south, whereas brown rockfish is distributed across coastal California. Percent postive catch generally showed higher catches south of San Francisco for vermillion, gopher, brown, and blue rockfish. Black rockish showed higher postive catches in the north, while the percent of drifts retaining China rockfish were all around low coastwide. The average CPUE was highest for blue rockish between San Francisco south to Big Sur (Figure 4). Black rockfish average CPUE higher in the north, while gopher rockfish CPUE was generally consistent throught the coast, albiet slightly higher south of Big Sur. China rockfish CPUE catch was typically low coastwide, with slightly higher catch rates in the Farallon Island reefs.

The final aggregation of the reefs and total area within each region are found in Table () and reflect the disribution and patterns in the visual representation of commonness in the data. The fraction of drifts retained for the indices of abundance was high for all six species (80% or greater), indicating that many of drifts within these data occurred near areas of rocky habitat.

3.2. Stephens-MacCall Data Filtering

A total of 19,425 drifts that aggregated to 2,252 trips were used for the trip-level Stephens-MacCall filtering. In general, the co-occurring species used for the Stephens-MacCall method were similar for the drift-level and the trip-level data. We present the coefficients and 95% confidence intervals for the species coefficients for black rockfish and brown rockfish in Figure (fdfsfdj). The plots for the remaining four species are available in the supplemental materials. The confidence intervals were larger for the trip-level data and the co-occuring species at the drift-level provide a refined look at species that have positive coefficients. For black rockfish, a noticible difference is the intercept. At the trip-level the intercept (probability the only black rockfish is encountered) is uninformative and at the drift-level the intercept is strongly negative. A higher fraction of the co-occuring species provide uniformative information (the 95% confidence interval crosses zero) for the trip-level data than the drift-level.

The percent of the samples retained for each data method differed by species, but followed the general trend that the lowest percent of samples were retained from the Stephens-MacCall filtering at the drift level, ranging from 12% of samples retained for China rockfish and 54% for blue rockfish. A much higher percent of samples were retained both from the other two methods, with an average of 83% of drifts retained when habitat was included as a filter. Data filtering for the indices with data aggregated to the trip-level and using the status quo of retaining all positive observations resulted in a high proportion of positive samples (0.70 - 0.86) for all species.

3.3. Indices of Abundance

All but three of the 24 indices of relative abundance were modeled with a lognormal distribution. The trip-level indices for black, blue and gopher rockfish were modeled using a gamma distribution. In general, the larger increases and decreases in the indices were similar among the four indices developed for each species (Figure Figure 6). The generalized approach used in this paper to create indices with comparable methods resulted in different results for each species. The area-weighted indices are reflective of the total available habitat and use all of the available high resolution habitat and fishing drift data. The effects of this can be seen in the plots where the area-weighted indices depart from the habitat-informed drift-level indices. For example, the effect of the area-weighting is apparent for black rockfish in 2005, 2007, 2009 and 2013. For China rockfish the habitat-informed indices present a more variable index, whereas both the Stephens-MacCall filtered datasets are more similar. For vermilion rockfish, while the trends are similar among all four indices, the effect of area-weighting dampens the increase modelled from the habitat-informed drift level data from 2004-2006.

China rockfish is the only species for which the trip-level index had the lowest average coefficient of variation, which increased with the the habitat-informed filtering (Table **?**). For all other species, the habitat-informed filtering resulted indices with a lower average CV than the trip-level filtering. This is most apparent for brown and gopher rockfish where the estimated error shrinks drastically for all of the drift level indices versus the trip-level index (Figure xx) T

The average CVs between the drift-level area-weighted index and the drift-level habitat informed indices were similar, as expected, since they both used the same data with the only difference being the year:area interaction in the models. However, the average CV between drift-level habitat-informed filtering and Stephens-MacCall filtering for the drift-level data differed by species. The average CV for brown rockfish from the Stephens-MacCall filtering was large (0.679) compared to the habitat informed filtering (0.142).

4. Discussion

Data were limited to the California coast north of Point Conception ($34°27'N$). The composition of the fish communities in southern California differ, and the recreational fisheries are fundamentally different, with a higher percentage of trips targeting mixed species and pelagic and highly migratory species, as well as more limited access to rocky habitat nearshore. Point Conception is a biogeographic break (citation) and a number of stock assessments In addition, complete habitat data are not available for areas in southern California. The data were also temporally restricted to the years 2001-2016. Earlier and more recent data were excluded to preserve a data set with the most consistent gear and depth regulations.

Habitat layers The characteristics and classification of the rocky habitat into more specific substrate types, e.g., boulder vs pinnacle, is available for a small fraction of the mapped area. Therefore, all areas of rocky substrate are currently created equal. A number of video surveys have shown habitat associations differ by species and ,

Oftentimes a captain will position the vessel adjacent to rocky habitat so that the current allows the vessel to drift over the rocky habitat.

The Stephens-MacCall model was developed to approximate habitat for recreational fisheries data with unknown fishing locations. The onboard observer surveys coupled with the high resolution rocky reef habitat maps remove the uncertainty in both fishing locations and the availalbe habitat. While the Stephens-MacCall filter is useful in identifying co-occurring or non-occurring species it assumes all effort was exerted in pursuit of a single target species.

The targeting of more than one species or species complex ("mixed trips") can result in co-occurrence of species in the catch that do not truly co-occur in terms of habitat associations informative for an index of abundance. This was clearly shown in the differences between the trip-level Stephens-MacCall filtering that relies on the information gathered from an entire trips to the drift-level Stephens-MacCall filtering that reflects the species encountered at a single location.

Both blue and black rockfish have high affinity to rocky habitat, but occur higher off the bottom and are both schooling species. It is not uncommon to have a a drift dominated by blue rockfish in central California, or black rockfish further north. However, the Stephen-MacCall approach does not account for this by modeling presence/absence.

The choice of a threshold value to use as a data filter from the Stephen-MacCall method should be reviewed to determine how sensitive an index of abundance is to that method. The

people have been addressing SM questions and how to deal with space in stock assessmetn for awhile

The majority of groundfish species targeted by the CPFV fleet north of Point Conception during the time period of this study all have high associations to rocky habitat. In this case, the Stephens-MacCall method can be considered a proxy for habitat when the species of interest has known associations. This can be expanded in areas where trips are known to target species of interest, but no habitat data are available the proportion of trips encountering the target species could be used as a proxy for habitat. This does not hold for areas where multiple species complexes are targeted on same trip, e.g, a multi-day trip may target large pelagic species and once trip limits are reached, the trip may focus on a secondary target, which is the case for the California CPFV fleet fishing south of Point Conception.

The suite of six species that we modelled in this paper is a concrete example of why habitat is important and also varies among the species. The high proportion of retained drifts across species when using habitat as a data filter indicates that hate majority of drifts occurred over, or very close to, rocky habitat.

There are a number of key assumptions made when using the onboard observer data in a stock assessment. A key assumption of the onboard observer surveys is that fishing behavior remains the same when samplers are not onboard the vessel. If a captain only fishes particular locations or targets a different suite of species when a sampler is onboard the vessel, additional bias is introduced in the data.

spatio-temporal modelling.

Versions of the indices filtered based on habitat were approved by the Pacific Fisheries Management Council's Science and Statistical Committee for use in the 2013 stock assessments and have been used all of the stock assessment process since. Comparisons should not be drawn between the indices presented here and the stock assessment documents as the indices in this paper were simplified to develop direct comparisons among methods. When filtering and modelling data for a stock assessment, additional filtering steps would be taken, such as excluding areas where species are rare, e.e., south of Santa Cruz for black rockfish. However, this is also a function of the lower sampling rates along the coast north of San

Francisco.

Addtional factors not considered in the simplified models presented here include the fact that the catch from the recreational CPFV fishery is dependent on a number of factors including weather, distance from port, the clientele preferences, angler experience and captain's knowledge. These models also do not account for distance to the nearest port, which has been shown to significantly impact the access to fish as well as historical fishing pressure....In addition, in 2004 the CDFW implemented spatial and temporal closures to the recreational nearshore groundfish fishery.

The fishery-dependent indices of abundance undergo higher levels of scrutiny during stock assessment reviews due to the nature of the data being driven by fisher behavior. The one fishery-independent survey for nearshore groundfish in California north of California tends to have similar trends to the fishery-dependent indices for the shallower nearshore species like gopher and China rockfish.

The influence of an index of abundance is sometime the can have a large influence on end year estimation of stock status (find examples).

accepted for management (China, gopher/black-and-yellow, vermilion/sunset, blue/deacon, black, lingcod - cite assessments).

Recent studies have identified the need to investigate the assumptions and uncertainty in relative indices of abundance from visual surveys (Bacheler and Shertzer, 2015; Campbell et al., 2015) and simulation studies (Siegfried et al., 2016).

Prioritize data for stock assessments (Magnusson and Hilborn, 2007).

Stock synthesis weighting of indices based on CVs - is the CV tighter for the fishery-independent survey to give it have an edge over the onboard observer survey?

Composition data from recreational surveys had the largest impact on simulation results, but individual survey components did not have individual effects on benchmarks (Siegfried et al., 2016).

5. Tables

Species	Drift-level habitat-informed	Drift-level SM-filtered	Trip-level
Black rockfish	0.886	0.252	0.408
Blue rockfish	0.830	0.538	0.871
Brown rockfish	0.855	0.243	0.490
China rockfish	0.808	0.121	0.515
Gopher rockfish	0.787	0.401	0.755
Vermilion rockfish	0.799	0.382	0.821

Table 1: Area of rocky habitat in state waters aggregated to levels modelled for each species. The shaded blocks for each species indicate which areas ere aggregated to ensure appropriate samples sizes to explore an area-weighted index.

<table>
<tr><th>Rocky Reef Desginations</th><th>Blue rockfish & Vermilion rockfish</th><th>Black rockfish</th><th>Brown rockfish</th><th>China rockfish</th><th>Gopher rockfish</th></tr>
<tr><td>California border to San Francisco</td><td>439.546</td><td>439.546</td><td>439.546</td><td rowspan="2">547.970</td><td rowspan="4">735.825</td></tr>
<tr><td>San Francisco to Santa Cruz</td><td>108.424</td><td>108.424</td><td rowspan="2">498.967</td></tr>
<tr><td>Farallon Islands</td><td>50.252</td><td rowspan="4">390.543</td><td>50.252</td></tr>
<tr><td>Moss Landing to Big Sur</td><td>137.603</td><td rowspan="2">228.027</td><td>137.603</td></tr>
<tr><td>Big Sur to Morro Bay</td><td>90.424</td><td rowspan="2">202.688</td><td>90.424</td></tr>
<tr><td>Morro Bay to Point Conception</td><td>112.264</td><td>112.264</td><td>112.264</td></tr>
</table>

Table 2: The average Coefficient of Variation (CV) for each index of abundance, where SM-filtered is the Stephens-MacCall filtering.

Species	Drift level Area-weighted	Drift level habitat-informed	Drift level SM-filtered
Black rockfish	0.443	0.449	0.364
Blue rockfish	0.134	0.142	0.099
Brown rockfish	0.242	0.240	0.679
China rockfish	0.320	0.301	0.233
Gopher rockfish	0.179	0.183	0.138
Vermilion rockfish	0.152	0.178	0.133

6. Figures

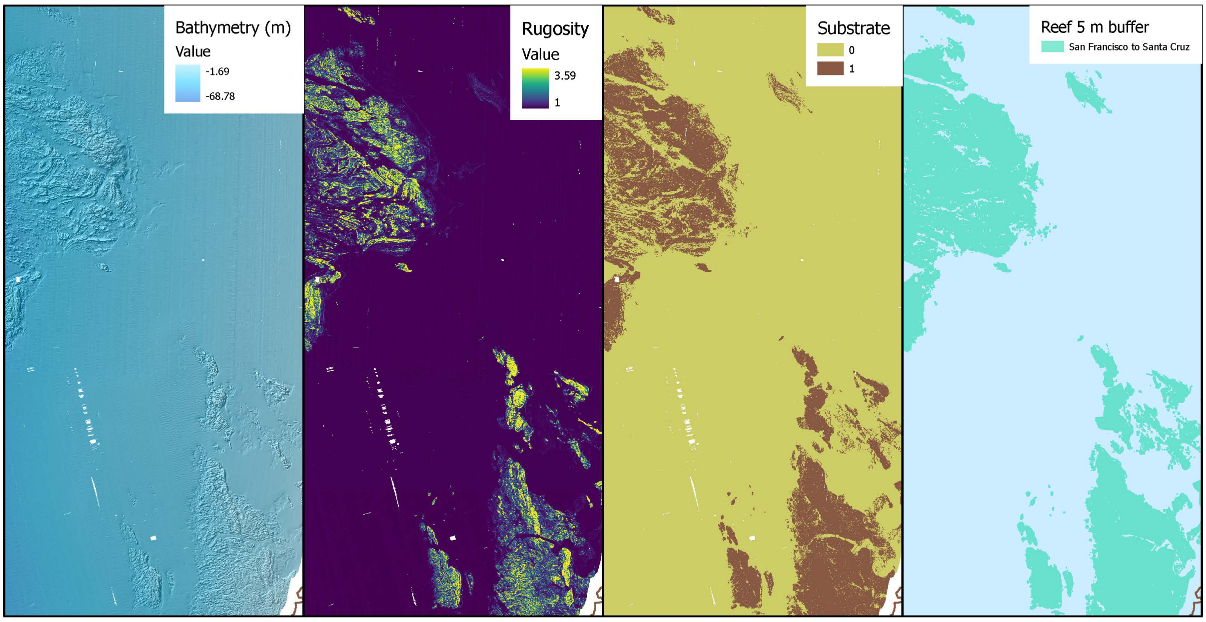


Figure 1: A example of the high resolution bathymetric data and components used to create the rocky reef habitat layer for groundfish (far right panel).

References

Bacheler, N.M., Shertzer, K.W., 2015. Estimating relative abundance and species richness from video surveys of reef fishes. Fishery Bulletin 113.

Campbell, M.D., Pollack, A.G., Gledhill, C.T., Switzer, T.S., DeVries, D.A., 2015. Comparison of relative abundance indices calculated from two methods of generating video count data. Fisheries Research 170, 125–133.

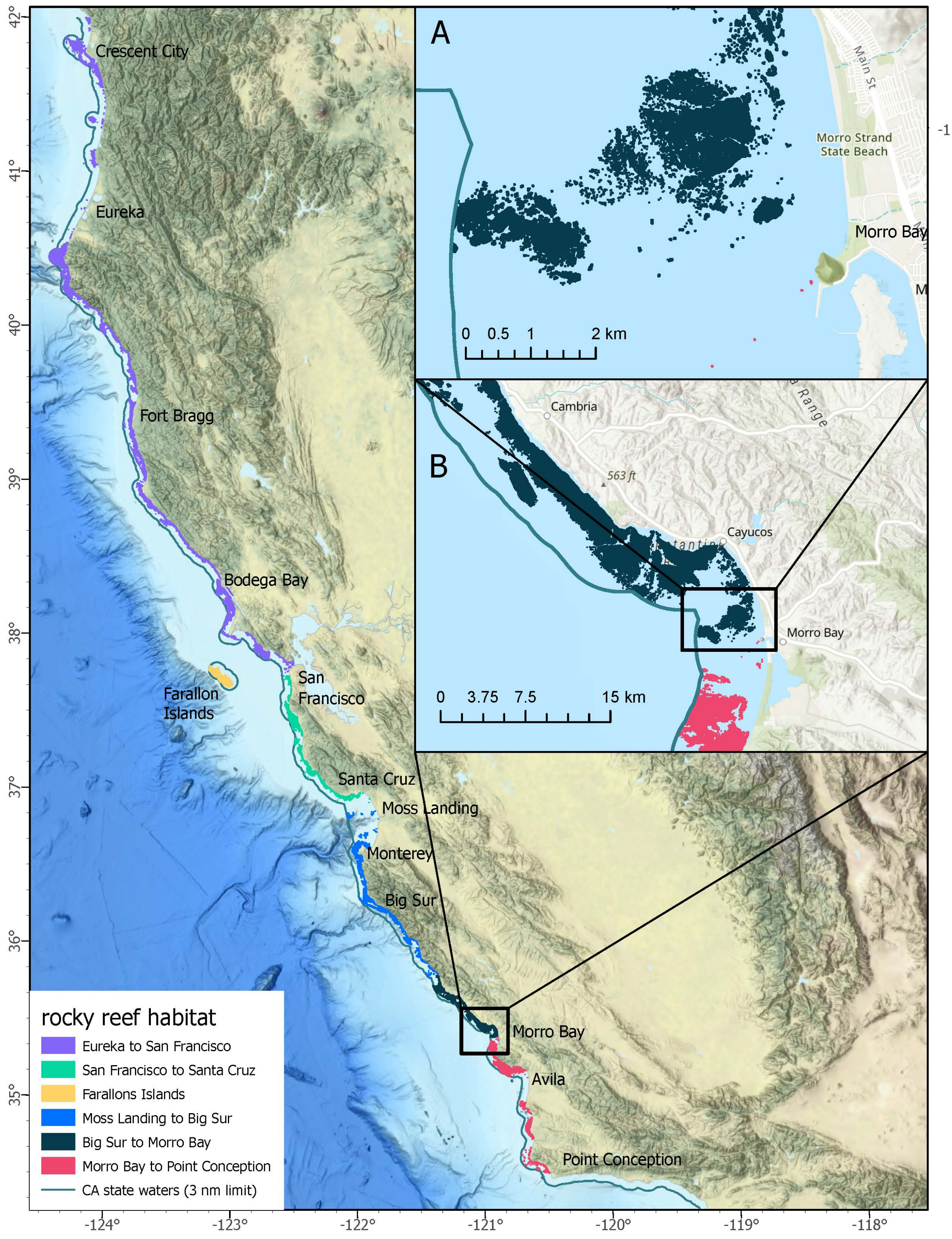


Figure 2: A maps of California state waters north of Point Conception colored by the aggregated areas of rocky reef habitat, including inset A depicting the rocky reef habitat in relation to 3 nm state water boundary state waters and inset B showing the high resolution rocky habitat in the area.

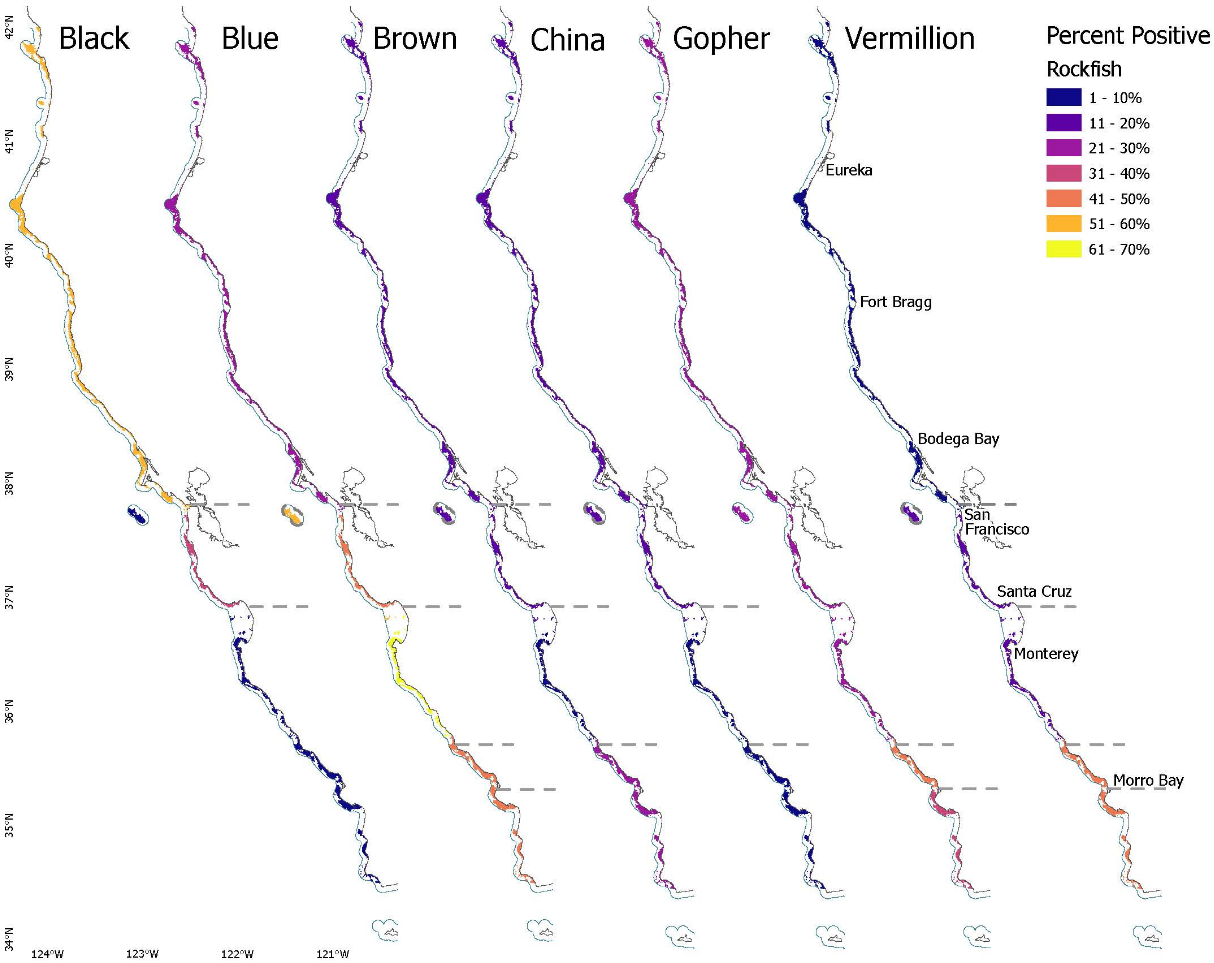


Figure 3: The percent of drifts that retained the target species, within grouped areas of rocky habitat over all years of the time series. The grey dashed lines represent the aggregated rocky habitat used to develop an index of abundance.

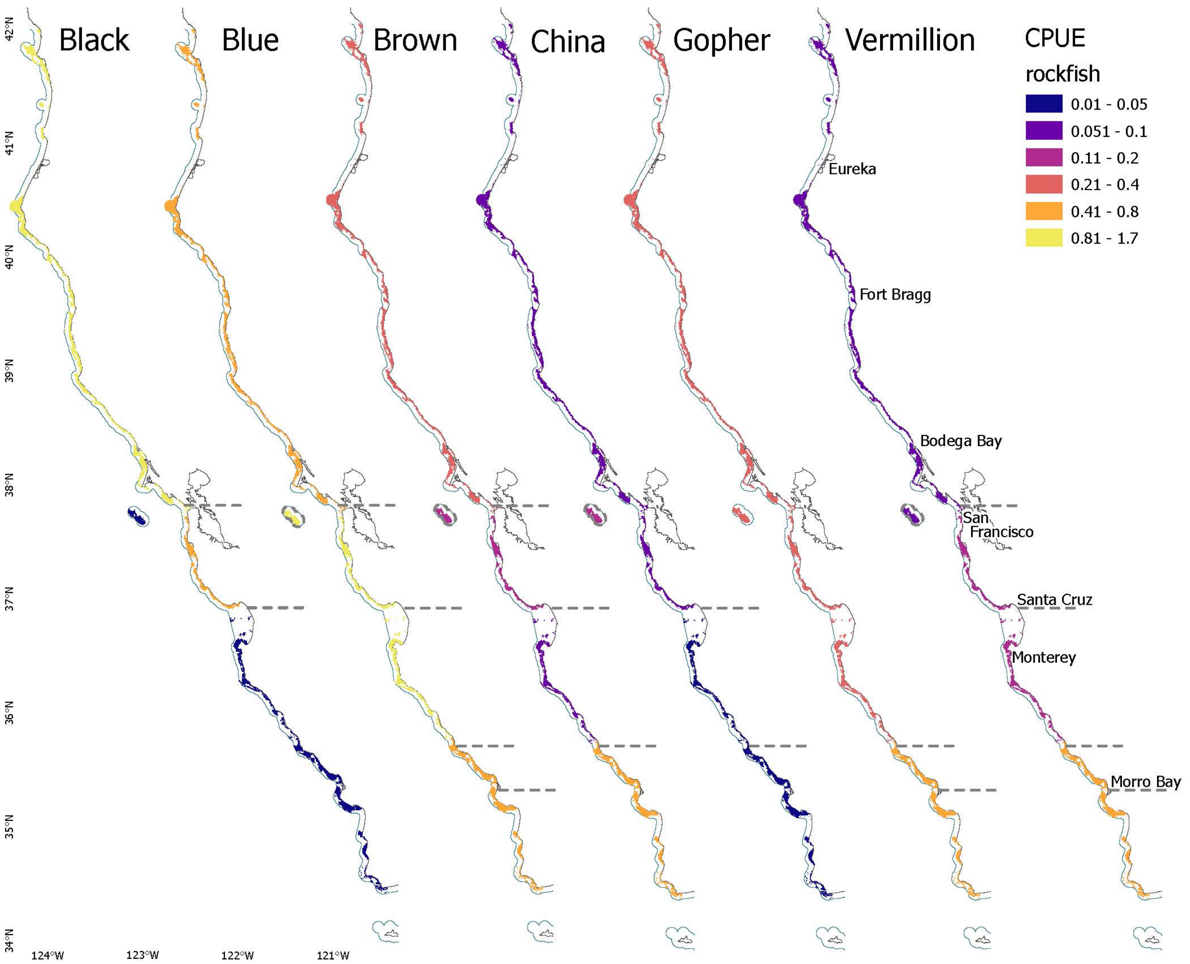
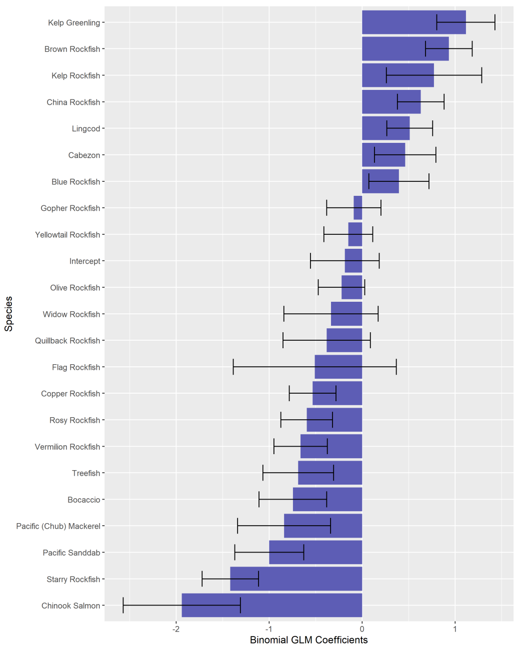
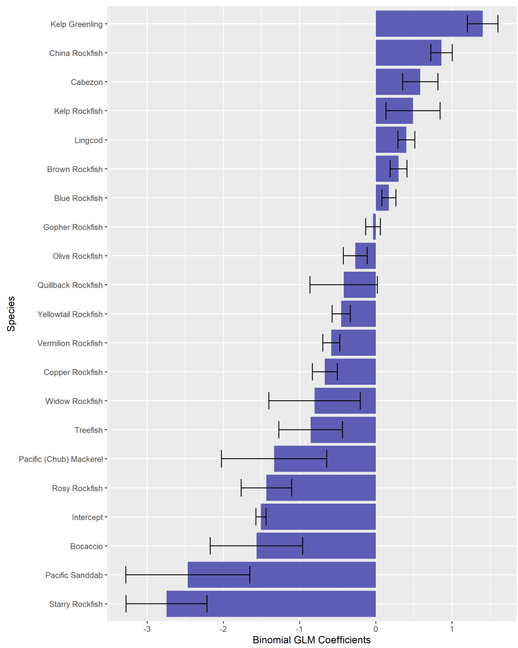


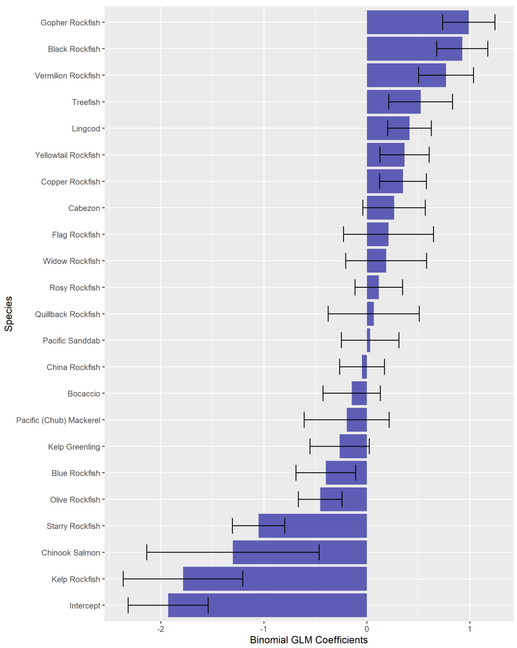
Figure 4: The average CPUE across all years of the time series for each of the six species. The grey dashed lines represent the aggregated rocky habitat used to develop an index of abundance.



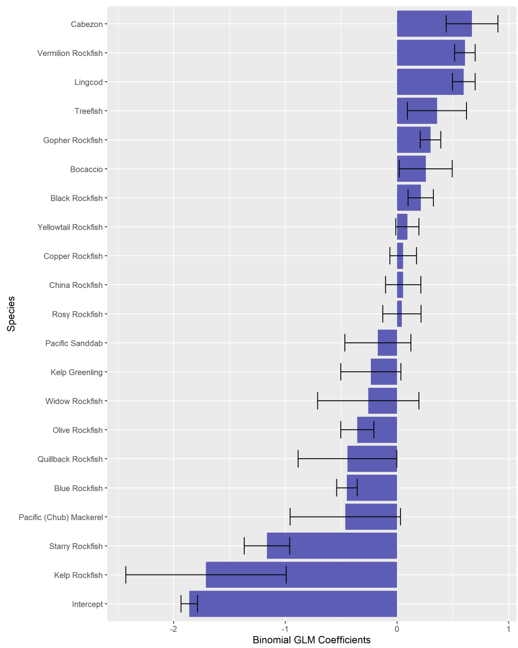
(a) Black rockfish trip-level



(b) Black rockfish drift-level



(c) Brown rockfish trip-level



(d) Brown rockfish drift-level

Figure 5: Examples of the species coefficients and 95% confidence intervals for the Stephens-MacCall filtering for black rockfish and brown rockfish for the trip-level and drift-level data.

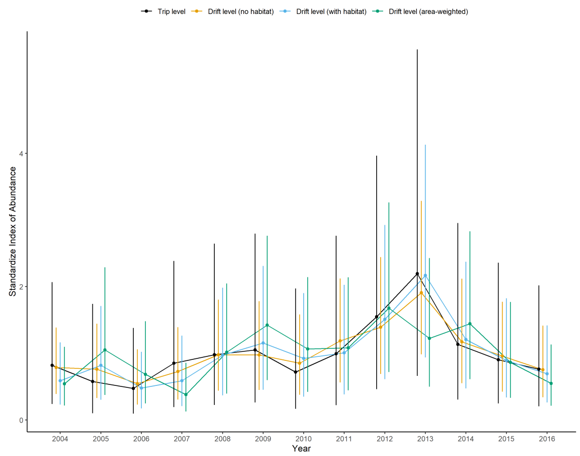


Figure 6: Black rockfish

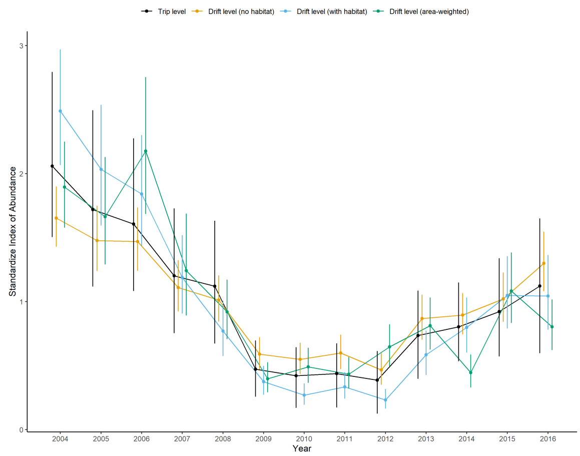


Figure 7: Blue rockfish

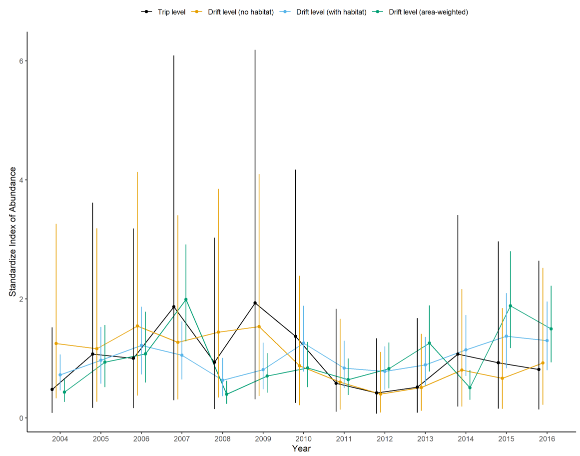


Figure 8: Brown rockfish

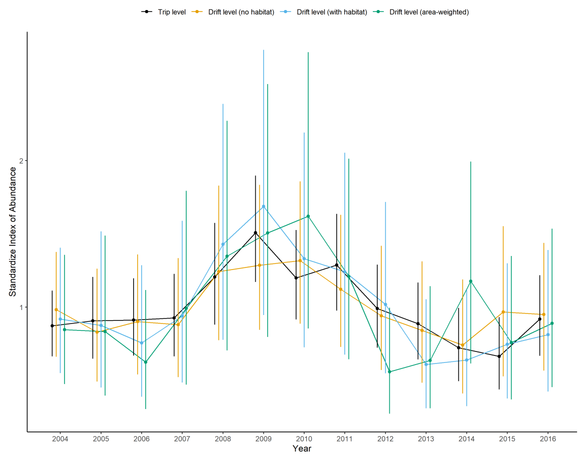


Figure 9: China rockfish

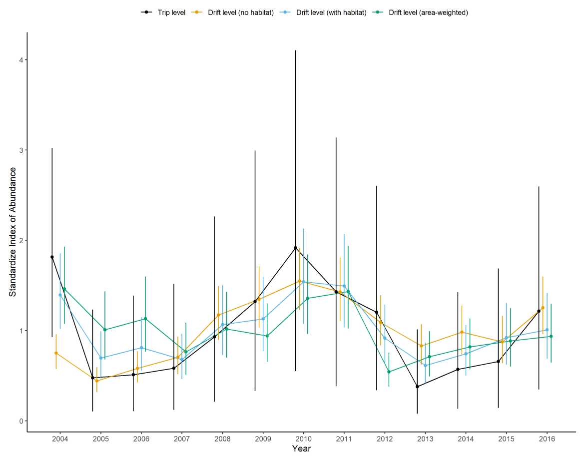


Figure 10: Gopher rockfish

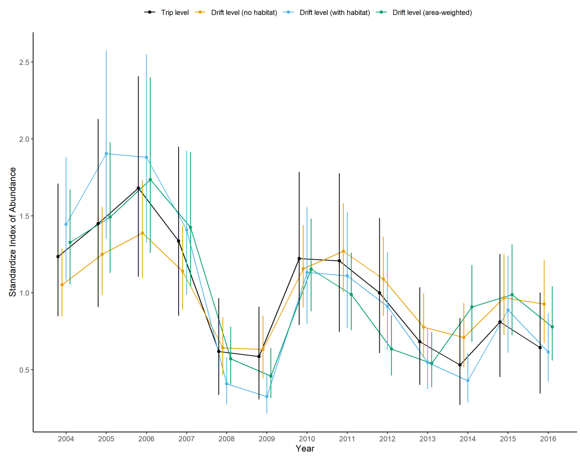


Figure 11: Vermilion rockfish

Scaled indices of abundance for four data filtering methods explored.

Harley, S.J., Myers, R.A., Dunn, A., 2001. Is catch-per-unit-effort proportional to abundance? Canadian Journal of Fisheries and Aquatic Sciences 58, 1760–1772. doi:10.1139/f01-112.

Lo, N.C., Jacobson, L.D., Squire, J.L., 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49, 2515–2526.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish and Fisheries 8, 337–358.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: A review of recent approaches. Fisheries Research 70, 141–159. doi:10.1016/j.fishres.2004.08.002.

Monk, M.H., Dick, E.J., Pearson, D., 2014. Documentation of a relational database for the California recreational fisheries survey onboard observer sampling program, 1999-2011. NOAA-TM-NMFS-SWFSC-529 .

Siegfried, K.I., Williams, E.H., Shertzer, K.W., Coggins, L.G., 2016. Improving stock assessments through data prioritization. Canadian Journal of Fisheries and Aquatic Sciences 73, 1703–1711.

Stephens, A., MacCall, A., 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research 70, 299–310.