Title: Methods for using remotely operated vehicle survey data in assessment of nearshore groundfish stocks along the California coast.

Authors: Dr. John Budrick, Ms. Laura Ryley and Mr. Mike Prall, California Department of Fish and Wildlife

Introduction: Nearshore groundfish stock assessments have identified the lack of fishery-independent data sources as a research and data need (Agenda Item E.2, Attachment 1, September 2017). In addition, methods currently utilized in stock assessments do not explicitly account for abundance inside of no-take Marine Protected Areas (MPAs). Remotely operated vehicles (ROVs) provide a non-lethal sampling method in areas where harvest is prohibited. They also allow collection of data on overfished species and nearshore species which constrain take of healthy stocks. Because ROVs employ only non-lethal data collection methods, they avoid need for research set-asides or other allocative considerations that may arise between fisheries and research sectors.

The California Department of Fish Wildlife (CDFW) in collaboration with Marine Applied Research and Exploration (MARE) conducted ROV surveys to measure differences in fish density (fish/m^2, hereafter referred to as density) and size of fish inside MPAs and at reference locations open to fishing. These data can be applied in stock assessments as indices of relative abundance and expanded using habitat area estimates to provide estimates of abundance.

In this study, the depth distribution of sampling and life history of encountered species were evaluated to select candidate species for which the ROV survey could provide indices of abundance or estimates of abundance. Generalized linear models (GLM) with alternative distributions were evaluated as an appropriate means of deriving indices of abundance from the estimates of density provided by the ROV survey taking into account overdispersion and other characteristics of the data set. We also identified variables correlated with density estimates that should be accounted for in normalizing indices of abundance, including depth, latitude, proportion of hard substrate along the transect and whether take is allowed in the area sampled. The results help inform future development of indices of abundance for use in integrated stock assessments once multiple years of sampling have been conducted to provide a suitable time series.

The California Seafloor Mapping Program (CSMP) is a collaborative effort that has performed high resolution bathymetry mapping allowing the categorization of seafloor for the vast majority of California's state waters, which encompasses most of the habitat of nearshore groundfish species. In combination, seafloor mapping, density estimates from the ROV survey, and proxy estimates of average weights from the California Recreational Fishery Survey (CRFS) were used to estimate the abundance of gopher rockfish between Pigeon Point (San Mateo County) and Point Conception (Santa Barbara County), California. Both design-based and model-based estimates of abundance were conducted to provide comparable estimates. The design-based estimates considered variation in the density with depth, proportion of hard seafloor and latitude identified by the GLM through stratification of estimates. Model-based estimates considered the same variables as well as terrain attributes describing seafloor relief from the CSMP data and used a Generalized Additive Model (GAM) to derive coefficients defining the relationship of variables to the observed density of gopher rockfish. The resulting model describing density was expanded to the entire study region using CSMP data using the Marine Geospatial Ecology Tool (MGET) described in Roberts et. al (2010) with methods analogous to those applied in Young and Carr (2015).

Absolute estimates of abundance that account for the amount of suitable habitat can be used as indices in integrated stock assessments provided that future monitoring produces a suitable time series. Alternatively, a single estimate of abundance from these methods could be used to inform the scale of stock assessments, which is subject to considerable uncertainty in the absence of absolute estimates of abundance, to "peg" the indices of abundance from long standing surveys or derived from fishery data providing information on trend but lacking scale. In addition, fish density maps or the distribution of abundance can be used to inform allocation of annual catch limits across management areas. Lastly, the estimates of abundance in combination with proxies for fishing mortality at maximum sustainable yield (F_{MSY}) for groundfish stocks can be used to estimate stand-alone overfishing limits forming the basis for annual catch limits as category 2 or 3 stock assessments.

The CDFW provides the following analysis of the use of the density estimates and expanded estimates of abundance generated using data from this ROV survey method in nearshore stock assessments for methodology review. A desk review was conducted in 2019 by the methodology review panel members to provide input on potential refinements and was received in October. Responses to suggestions or questions from reviewers are provided in grey highlight below the section of the methodology to which they pertain.

Overview of Methodology

The following sections provide an overview of the technical attributes of ROV equipment, the data it collects and the sampling design. Thereafter, we describe the CSMP, how terrain attributes that describe seafloor relief are derived from CSMP data and are paired based on location to the segments of the ROV transects used as the density sampling unit for this methodology. Subsequently, we discuss the species to which the proposed methods can be applied given the interplay of their life history and the sampling design. Following, methods used to develop models for deriving indices of abundance from fish observations and habitat variables available from the ROV and CSMP are described and evaluated including appropriate GLM distributions and variable selection criteria. The results of the GLM are used to inform poststratification of data in design-based methods of expanding density to provide absolute estimates of abundance with estimates of habitat area from CSMP data. We use MGET integrating R and ArcGIS to develop a GAM of gopher rockfish density to provide a model-based estimate of abundance with the variables from the ROV survey and terrain variables from the CSMP data expanded using CSMP data in raster format. The results are then compared to design-based estimates of abundance based on estimates of density in depth and latitude strata expanded by area identified as habitat using the CSMP. We go on to discuss the use of the results in stock assessments and conclude with discussion of potential refinements to improve the methods in the future. The following table of contents provides additional information on the organization of the document and its contents.

Table of Contents:

ROV Data Collection:

Field Methods Overview: The statewide ROV survey conducted from January 2014 to December 2016, visited 148 sites providing observational data for density estimates from within MPAs and reference locations with similar habitat (Appendix 1). Survey transect lines within sites were positioned based on the location of rocky habitat and distributed across the entire depth range of rocky reef where possible. At each site, four to ten transect lines started at a random point were surveyed to achieve four km of survey transects within rocky habitat at each site. The ROV positioning system calculated the longitude, latitude and depth every two seconds allowing observations to be georeferenced. Ranging sonars were used to estimate transect width allowing calculation of the observed swath width along the course of each transect at one second intervals. The ROV was outfitted with stereo cameras and paired lasers allowing estimates of lengths of encountered fish. Additional details of aspects of the field data collection are provided below.

ROV System and Configuration

ROV Underwater Maneuvering

The ROV used in these surveys was a Deep Ocean Engineering Vector M4, named ROV Beagle, owned and operated by MARE. The ROV was equipped with a three-axis autopilot including a rate gyroscopically damped compass and sonar altimeter. Together, these aids allowed the pilot to maintain a constant heading (± 1 degree) and constant altitude (± 0.3 meter) above the seafloor with minimal pilot corrections. In addition, an adjustable forward thrust control was used to help the pilot maintain a consistent forward velocity between 0.25 and 0.5 m/sec.

ROV Positioning and Transect Swath Measurement

An ORE Offshore Trackpoint III® ultra-short baseline acoustic positioning system with ORE Offshore Motion Reference Unit (MRU) pitch and roll sensor was used to reference the ROV position relative to the ship's Wide Area Augmentation System Global Positioning System (WAAS GPS). The ship's heading was determined using a KVH magnetic compass. The Trackpoint III® positioning system calculated the geographic position of the ROV relative to the ship at approximately two-second intervals. The ship-relative position was corrected to real world position and recorded in meters as X and Y coordinates using the World Geodetic System (WGS 1984) Universal Transverse Mercator (UTM) coordinate system using HYPACK® 6.2 hydrographic survey and navigation software. Measurements of ROV heading, depth, altitude, water temperature, camera tilt and ranging sonar distance both forward and downward to the substrate, were averaged over a one-second period and recorded along with the position data.

Two Tritech® 500 kHz ranging sonars, which measure distance across a range of 0.1–10 m using a 6° conical transducer, were used as the primary method for measuring transect width for both the forward and downward facing video. The ranging sonars were fixed below and parallel to the camera between two forward-facing red lasers spaced 100 mm apart. Each transducer was pointed at the center of view in each camera and was used to calculate the distance to middle of screen, which was subsequently converted to width using the known optical viewing angle of each cameras field of view. Readings from these sonars were averaged five times per second and recorded into the sensor data file at a one-second interval. Measurements of transect width using a ranging sonar are accurate to ± 0.1 m (Karpov et al. 2006). Transect width is combined with distance traveled by the ROV to calculate area swept by the video cameras thereby providing transect swath for determining density and area covered of observations.

The ROV system positioning from both the ODFW and CDFW programs could also be improved through the addition of a DVL aided IMU system to provide positioning and attitude measurements independent of the USBL acoustic tracking system. (Dr. Trembanis)
Improvement of acoustically derived positioning is and has been of interest to our team as positioning error can be significant especially when small transect segments are needed to correlate with finer scale bathymetry derived seafloor mapping data. There is significant added cost of DVL equipment and added processing complexity, thus careful consideration of the achievable increase in accuracy and how this propagates into the error estimates of our analysis is necessary to evaluate the cost benefit. We have currently chosen the scale of our analysis to be much larger than the known accuracy inherent in our positioning data to minimize the propagation of this source of error. For finer scale analysis, investment in increased positional accuracy may be warranted.

ROV Imagery Capture Configuration
The ROV was equipped with four standard resolution (640 by 480) color cameras: two locally recorded stereo cameras for highly accurate measurements of size and two primary data collection cameras; one facing forward approximately 30° below the horizon and the other pointing directly downwards. The two-camera system provided a continuous, slightly overlapping view from above the horizon to directly below the ROV. Video for both cameras was captured on SONY® DSR 45 digital video tape recorders and Pioneer DVR510 digital video disc recorders. In addition to capturing biological and habitat observations, the forward video was overlaid with an on-screen display of text characters representing real time sensor data (time, depth, temperature, range, altitude, forward camera angle and heading). The ROV was also equipped with a high definition (HD) video and 5.6 mega pixel digital still camera. The HD and still cameras were mounted forward facing and were locally recorded on a hard drive housed in the ROV.   At the end of each survey day, this imagery was downloaded and saved to a portable hard drive.

ROV Imagery and Data Timestamping
A continuous time feed was necessary to relate all data and imagery collected during ROV dives. Time was extracted once per second from the ships GPS data stream and was used to provide a basis for relating ROV position, sensor data and video observations. A Horita® GPS3 and WG-50 were used to generate on screen overlay display of GPS time, as well as output Society of Motion Picture and Television Engineers (SMPTE) linear timecode (LTC) for capture on video recorder audio tracks synchronized to each video frame at an interval of 1/30th of a second. This stored SMPTE timecode can then be accessed from the audio track during subsequent video scoring of any observation. This method was improved by customizing HYPACK® navigational software to link all data collected in the field to the GPS time. ROV tracked position and sensor data were recorded directly by HYPACK® as a time-linked text file. A redundant one-second timecode file of sensor data was also collected in the field using a custom-built on-screen display and operating system software with timecode extracted from the navigation computer system's internal clock which was synchronized to GPS time.

ROV Data Management
All data collected by the ROV, along with subsequent observations extracted during post-processing of the video, was linked in a Microsoft Access® database using GPS time. Database management software, developed by MARE, was used to expand all data records to one second of Greenwich Mean Time (GMT) time. During video post-processing, a Horita® Time Code Wedge (model number TCW50) was used in conjunction with a customized computer keyboard to extract and record SMPTE timecode stored on the video audio track, to one second resolution, of observations into a customized Microsoft Access® database entry form.

Reports mention quality control processes and R scripts for data reduction and entry. These workflows and code bases should be made available for the follow-on review and shared with the broader AIASI community. (Dr. Trembanis)

CDFW is interested in making data available to the AIASI community. Data shared with NOAA Fisheries will be subject to a data sharing agreement (DSA) with CDFW. A memorandum of understanding is under development by CDFW for this DSA. Data requests from other agencies or organizations will be reviewed by CDFW on a case by case basis. Requests can be directed to the authors of this report. Discussion of quality control and data reduction of all the ROV data sources and intermediate manipulations are beyond the scope of this review. Relevant elements of these processes can be explored further as they relate to questions of accuracy and error propagation or other questions during the review process.

Consideration should be made for integration and transfer of local databases into FGDC compliant datasets. Of particular value moving forward is to arrange data and image structures to make them readily available to use and be used by image toolkits. A great example of this is the NOAA Fisheries Strategic Initiative on Automated Image Analysis program. (Trembanis)

- ***A major recommendation for both programs is to review and closely align the data collection, archiving, and image analysis to the tools and protocols outlined by the AIASI initiative.***
- ***A benefit to the larger community image analysis efforts would come from having access to the derived imagery from the ROV surveys to use for further development and testing particularly of automated machine learning algorithms for fish detection and sizing.***
- ***Relational database structures are recommended and the use of secure and backed up servers that can be made available to other users in the field. (Trembanis)***

The CDFW is interested in making our extensive annotated datasets available to the AIASI to inform machine learning and other similar efforts. Data shared with NOAA Fisheries will be subject to a data sharing agreement (DSA) with CDFW. A memorandum of understanding is under development by CDFW for this DSA. Data requests from other agencies or organizations will be reviewed by CDFW on a case by case basis. Requests can be directed to the authors of this report. All datasets are maintained in relational Microsoft Access format that could be readily translated to other data formats as needed, however funding and staff resources limit the ability of CDFW to make all datasets FGDC compliant and compatible with the AIASI. Reprocessing and reduction of data for data sharing will be considered along with data requests.

Support Vessels and ROV Sampling Operations

During a fall 2015 cruise performed between Point Arena, Mendocino County and Ano Nuevo, Santa Cruz County, ROV operations were conducted off the F/V Donna Kathleen, a 19 m fishing vessel owned and operated by Tim Maricich. During a central coast cruise performed in 2016, ROV operations were conducted off the R/V Miss Linda, a 23 m research vessel owned and operated by Captain Robert Pedro. Surveys were conducted between the hours of 08:00 and 17:00 PST to avoid the low light conditions of dawn and dusk that might affect fish abundance, measurements and underwater visibility. The ROV was flown off the F/V Donna Kathleen's starboard side and the R/V Miss Linda's port side using a "live boat" technique that employed a 317.5 kg clump weight. Using this method, all but 45 m of the ROV umbilical was isolated from hydrodynamic drag by coupling it with the clump weight cable and suspending the clump weight several meters off the seafloor. The 45 m tether allowed the ROV pilot sufficient maneuverability to maintain a constant speed (0.5 to 0.75 meters per second) and a straight course along the planned survey line. In addition, the ROV pilot and ship's helm used real-time video displays of ship and ROV location, to navigate along the 500 m line. The ship's helm used the displays to follow and maintain the position within 35 m of the ROV. At each site, the ROV was flown as close as possible along the pre-planned survey lines. In most cases, the ROV pilot maintained forward direction within ± 10 m of the planned line. The ROV pilot used the ranging sonar readings to sustain a consistent transect width by maintaining the distance from the camera to the substrate (at the screen horizontal mid-point) between 1.5 and 3 m.

Only using samples when water clarity is good is fine if water clarity is not a determinant of habitat preference and/or avoidance/attraction. (Dr. Berger)

Water clarity is a factor in determining whether sampling can be conducted on a given day. Threshold levels have been established to ensure that field of view or visible distance into the foreground does not affect detection probability. It is difficult to account for behavioral differences within the range of clarity that are suitable for sampling, as a result behavior affecting the probability of detection is assumed to be unaffected. It is uncommon to experience sufficient frequency of sampling or variation in clarity at a given location to allow for examination of changes in habitat preference with water clarity at a site level. Broader scale analysis may explain regional patterns of habitat preference associated with water clarity.

Imagery Post-Processing and Scoring Methods

ROV Positional Data and Transect Processing

The path of the ROV across the seafloor is determined for every one second during dives using acoustic tracking referenced to the ship's GPS systems. Acoustic tracking systems generate numerous erroneous positional fixes due to underwater acoustic noise and vessel movement that is not adequately compensated for by the tracking system pitch and roll sensors. For this reason, positional data was examined manually to remove outliers and averaged to better approximate the actual path of the ROV. Positional information was filtered for outliers and smoothed using a 21-position running mean created by averaging of ten (X,Y) values before and after every position (Karpov et al. 2006). Planar length of positions tracked was calculated for each second and combined with width to calculate area surveyed per second. These one second lengths were then added together to determine area swept for a given length of transect. We referred to each of these one second intervals, with estimated area swept, as microframes which are then assembled into segments of variable length or area according to the particular quantitative analysis being performed or particular metric being determined. The microframe is the smallest unit of transect that can be resolved for either quantification or associative analyses.

Gaps in the positional data that occurred due to deviations from quantitative protocols, such as pulls (ROV pulled back by ship induced tension on the umbilical), stops (ROV stops to let the ship catch up) or loss of target altitude caused by traveling over the back side of high relief structures (visual loss of 4 m target distance for more than 6 seconds which typically occurs on the downward slope of high relief habitat) were removed from the data to be used to generate quantitative transects along each survey line. The remaining usable portions of each survey line were then divided into two different transect types; fish density transects, and invertebrate density transects. Details on each transect type are described later in the post-processing methods.

It will be beneficial, for both programs, to review how the field of view of the camera is scaled and how the diminishing perspective viewing guidelines are constructed. (Dr. Williams)

The visual field of view for enumeration of fish was made by tracking fixed objects on the seafloor in the video while the ROV was moving over a flat level seafloor at a stable height off the bottom. This method was repeated for every survey deployment and when any adjustments or changes in the camera apparatus were made. A stationary object (rock, kelp, invertebrate etc.) was marked using a pen on a transparent overlay over the screen in successive video frames establishing the angle of the diminishing perspective in the upper corners of the video screen view. By tracking objects in several areas of the viewing area, the widest horizontal plane was established roughly in the middle of the video monitor. The upper corners of the viewing area, as established by object tracking, were excluded from the area used to enumerate fish. Verification of transect width estimation was made by driving the ROV across known width sections of one-inch diameter PVC pipe laid on the seafloor with calibration marks in 10 cm increments. Using these calibration tests, a constant scaling factor was established and multiplied by the ranging sonar values to obtain an estimate of the width of the viewing area for every one second of video along the transect.

Substrate Determination and Scoring

A protocol to characterize observed substrate along survey transect lines was developed to allow computation of area coverage of individual substrate types or combinations across transects at variable scales that may be desirable for analyses at different scales or purposes.

The video record was reviewed, and substrate types observed were classified independently as rock, boulder, cobble, gravel, sand or mud for every microframe of transect data. Rock was defined as any igneous, metamorphic or sedimentary substrate; boulder as rounded rock material that is between 0.25 and 3.0 m in diameter and clearly detached from the base substrate; cobble as broken or rounded rock material that is between 6 and 25 cm in diameter and clearly detached; gravel as any granular material with a diameter between 0.5 and 6 cm; sand as any granular material less than 0.5 cm (may include organic debris such as shell or bone, gravel or pebble); and mud as fine material whose granularity is not discernible via the ROV imagery.

To determine substrate percent cover during review of the video, a transparency film overlay with diminishing perspective guidelines approximating a parallel swath was placed over the video monitor screen. Each of the substrate types are identified by the processor during independent viewings of video and were recorded as discrete segments of the transect by noting where it was present with a beginning and ending timecode. Thus, the segments of substrate types may overlap each other along the survey line, creating areas of mixed substrate combinations (e.g. rock/sand, sand/cobble) along the transect. A substrate segment was considered continuous until a break of two meters or greater occurred along the survey line or the substrate dropped below 20% of the total combined substrates for a distance of at least three meters.

After the scoring process, the substrates were combined to create three independent habitat categories: hard (rock and/or boulder), mixed (rock and/or boulder with any combination of cobble, gravel, sand and/or mud), or soft (any combination of cobble, gravel, sand, and/or mud). Due to the independent scoring of substrate types, other habitat categories can be created by combining substrates. Additionally, ecotones between substrate types can be categorized and applied to discrete segments of transect as desired. The substrate and habitat determinations were recorded in the database for every microframe of transect.

Machine Vision techniques could be incorporated to extract information from digital image using algorithms- either through traditional image manipulation techniques or artificial intelligence "Deep Learning" neural network strategies. The hope is that machine vision will be used to more efficiently collect accurate data on the detection, quantification, and measurement of organisms and the classification of species and seafloor substrata. Automation of the identification of animals and

habitats is part of the solution, and can be a useful tool, depending on the level of taxonomic and/or physical identification required. (Dr. Trembanis)

Given the extensive geographic coverage of these datasets, increased efficiency from data processing informed by machine learning could be worthwhile. As the reviewer notes "While the basic quality of the image is a function of platform and image processing, it should be noted that machine vision can be challenged by the complexity of the habitat and the diversity of organisms". Imagery from our extensive survey coverage within California varies greatly in habitat complexity and visual quality (lighting, water clarity, backscatter etc.). Machine vision techniques may have value for certain species and applications however, the current surveys have been designed for broad characterization and monitoring of ecological conditions and habitats which requires human observers to identify of hundreds of fish and invertebrate species. As acknowledged by the reviewer "A major challenge with machine vision detection systems is the need for large annotated image datasets for training and testing of the algorithms and the intensive work needed by trained human annotators to build such datasets". It is unlikely that development of machine vision for a few target fish species will reduce this time investment significantly. The development of such methods would require long-term commitments for future funding to ensure the recoupment on investment of development costs compared to the cost savings gained by staff reduction.

Fish Scoring and Enumeration

Fish viewed within the forward video were classified to the lowest taxonomic level possible. Individuals that could not be classified to the species level were grouped into higher taxonomic levels or a complex of visually similar species. Video processors used the downward video and still photos taken of a particular fish to aid in identification where necessary. A transparent screen overlay with lines representing a diminishing perspective was used during fish review to approximate the three dimensional transect extending away from viewing screen. The overlay served as a guide for determining if a fish was in or out of the ROV transect. Fish enumeration was restricted to a maximum distance of four meters to avoid missing fish being obscured by objects in the foreground or their shadows at greater distances adversely affecting the ability to accurately identify fish. Using the sonar range value depicted on the screen as a gauge, the processor determined if a fish was within four meters as it entered the viewing area. Fish that entered the viewing area were only counted if more than half the fish crossed the overlay guidelines.

To accurately correlate the location of the fish with habitat, timecode entry was made when the fish crossed the mid-screen line. For fish that were within four meters but swam away before they crossed the mid-screen line, timecode entry was made when the location where the finfish had been observed reached the mid-screen point. All data entries were recorded in a Microsoft Access® database linked with the time.

Fish size (total length) was estimated by the video observer with the use of two parallel lasers placed 10 cm apart aimed to hit the seafloor in the center of the video viewing screen of the forward-facing camera. Fish sizes were estimated to the nearest cm and when possible tagged for future stereo sizing. Criteria for stereo sizing included fish orientation (almost perpendicular) and distance (within two meters) to the cameras.

Techniques based on calibrated paired-camera imagery are needed to acquire accurate fish size data. These are being implemented but data are not yet evaluated. Evaluation should include the cost of data processing (annotation) against the quantum of measurements required for different stock assessment applications. (Dr. Williams)

In 2014 we began collecting paired camera (stereo) video for size estimation using SeaGIS Event Measure scaling software. Funding limitations have prevented extensive processing of this video for more than a few species in select locations. Sizes for Gopher rockfish and other species of interest over the entirety of the subject area in this analysis have not been completed at the time of this review. Efforts are underway to add stereo sizing to our current processing workflow. Crucial to this development is creating time efficient methods for processors and examining encounter rates of sizeable fish to evaluate achievable sample sizes by species and locations.

Some work was done on this as part of baseline MPA monitoring performed on the northern California coast from MPA and reference sites between Fort Bragg (Mendocino County) and Crescent City (Del Norte County). In this analysis sizes of select rockfish species obtained by processors using 10 cm parallel scaling lasers were compared to stereographic estimates (Kline et al. 2014). This work examines the relationship between laser and stereo based size estimates. In-depth analysis of stereo sizeable encounter rates by species was not part of this analysis, however for copper and vermillion rockfish approximately 50% of fish observed were sized successfully using the stereo method. Further examination of our current datasets has indicated likely rates of between 30-50% for demersal rockfish species.

Although stereo sizing provides much greater accuracy and precision than laser-based estimates the encounter rate of target species during ROV transects is likely to be a limiting factor in achieving sample sizes needed to adequately inform stock assessment applications in some cases. Cost estimates of incorporating stereo based sizing with surveys in 2019 and 2020 is under investigation. Initial cursory examinations suggest that processing costs would increase by 1.5 to 2 times the cost of fish enumeration depending on the number of target species and necessary sample sizes making incorporation into current funding levels feasible and with further refinement of protocols additional efficiency can be gained.

Spatial Distribution of Sampling

In the surveys used for this analysis, sampling was designed to monitor and detect changes in species density due to protection implemented by California's marine protected area (MPA) network. CDFW and MARE developed an index site sampling design where fixed sites were chosen inside and outside of MPAs with similar habitats to monitor change over time that may be attributed to the prohibition of fishing within MPAs. Sites consisted of a defined rectangular survey region that covered the depth profile of rocky reef at each location. The rectangular regions were always 500 m wide but varied in length depending on the local extent of reef. Benthic survey lines were conducted across the width of the survey region (i.e. 500 m long survey lines) utilizing a random systematic design. A random starting point was chosen in the shallow end of the survey area that allows the required number of equally spaced transects to be deployed across the rectangular area. Typically, the aim was to acquire 4 km of linear transect across rocky habitat at each site to ensure adequate statistical power to detect differences in density between protected and reference sites for MPA monitoring (Karpov 2010). Across all areas surveyed, this resulted in allocation of between 4 to 10 survey lines at a site depending on the local rocky reef characteristics with sparse patchy habitats and areas with wide depth ranges requiring more survey lines to achieve the goal of 4 km of rocky habitat. An example of the spatial positioning sampling grids and orientation of transects with bathymetry is provided in Figure 1.

Figures depicting the distribution of sample locations across the state and a table summarizing the number of survey lines sampled and number of fish observed in each survey area are provided in Appendix 1. The focus of our analysis is on the area north of Point Conception. Charts showing the distribution of rocky reef habitat and degree of coverage from sampling locations along the coast from California/Oregon border to Point Arena, from Point Arena to Pigeon Point and Pigeon Point to Point Conception are provided in Figures 2, 3 and 4, respectively.

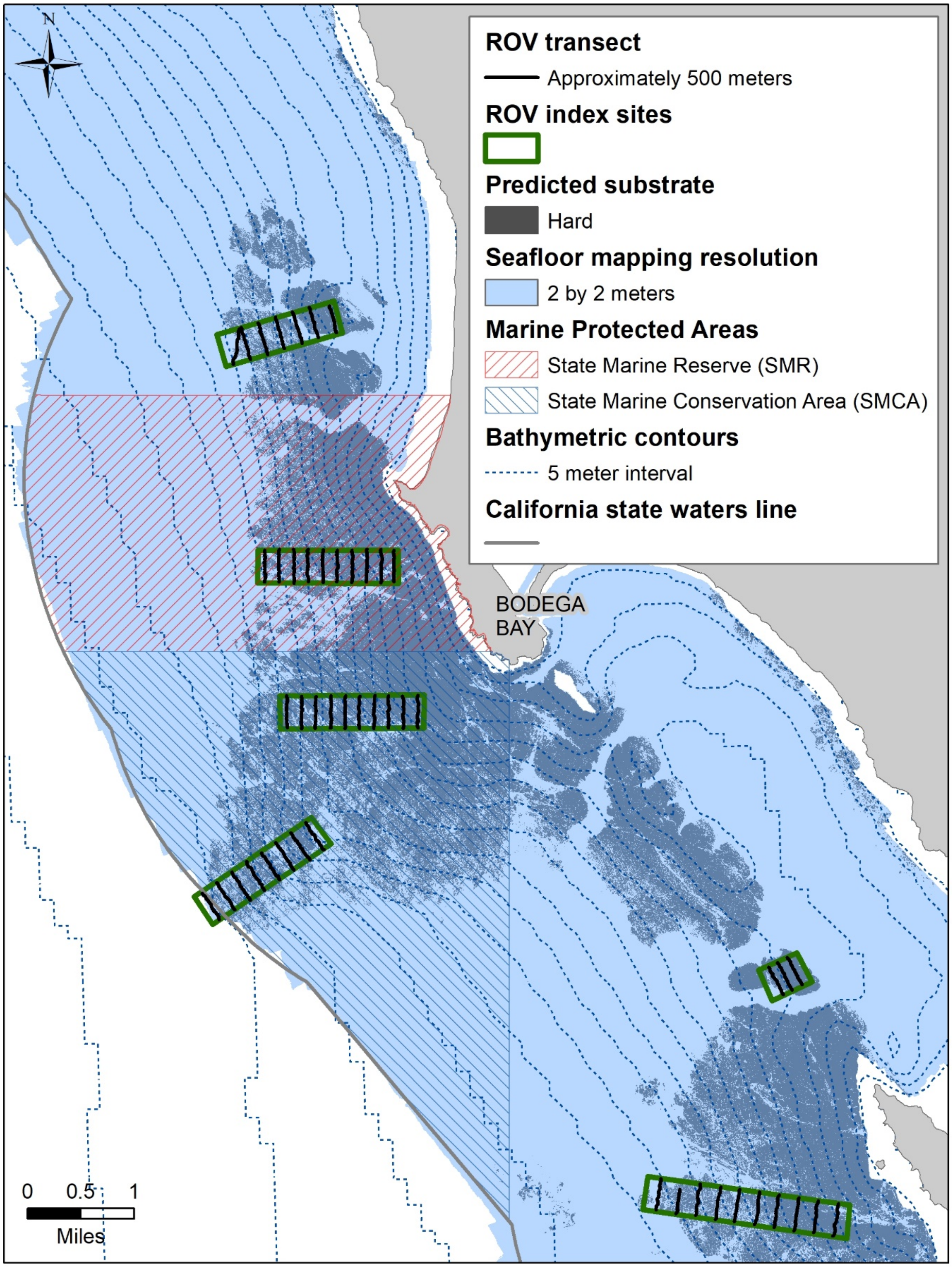

Figure 1. Depiction of the sampling design showing the boxes that identify sampling locations over hard substrate and the 500 m transect lines oriented to align with bathymetry contours and other features pertinent to the study.

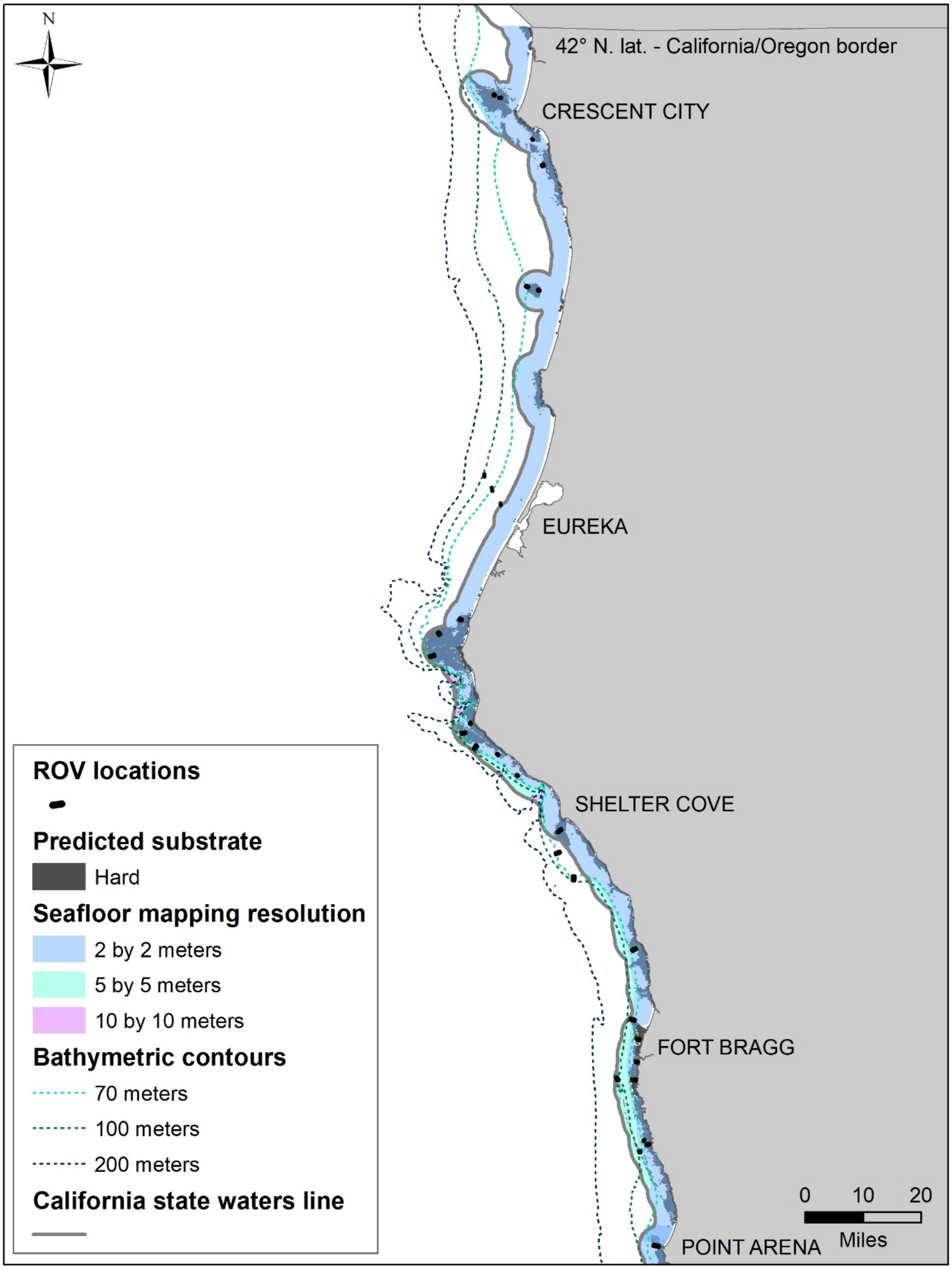

Figure 2. The Northern California coast from the California/Oregon border to Point Arena showing the distribution of hard substrate and degree of ROV coverage from sampling locations along the coast.

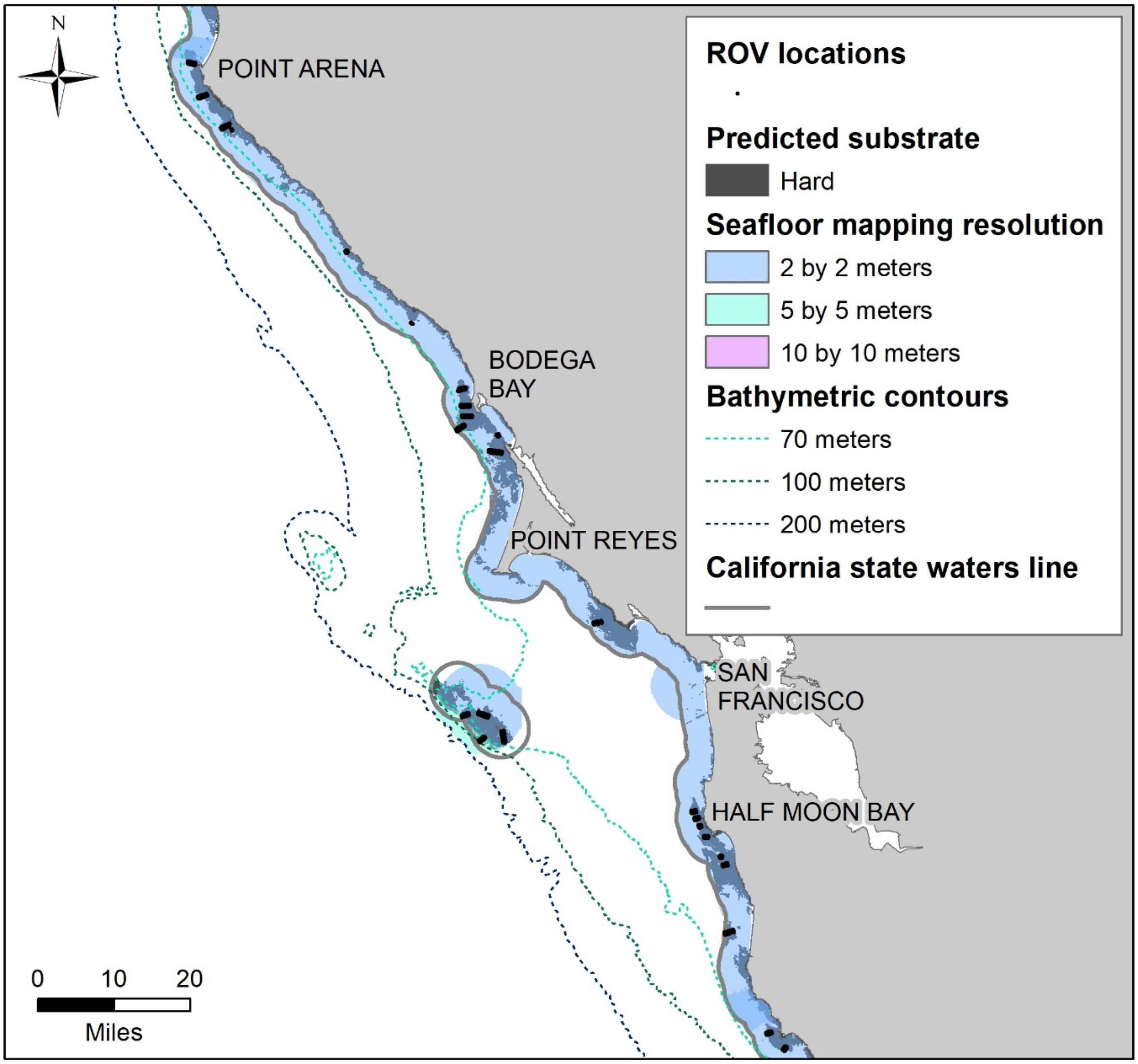

Figure 3. The North-Central California coast from Point Arena to Pigeon Point showing the distribution of hard substrate and degree of ROV coverage from sampling locations along the coast.

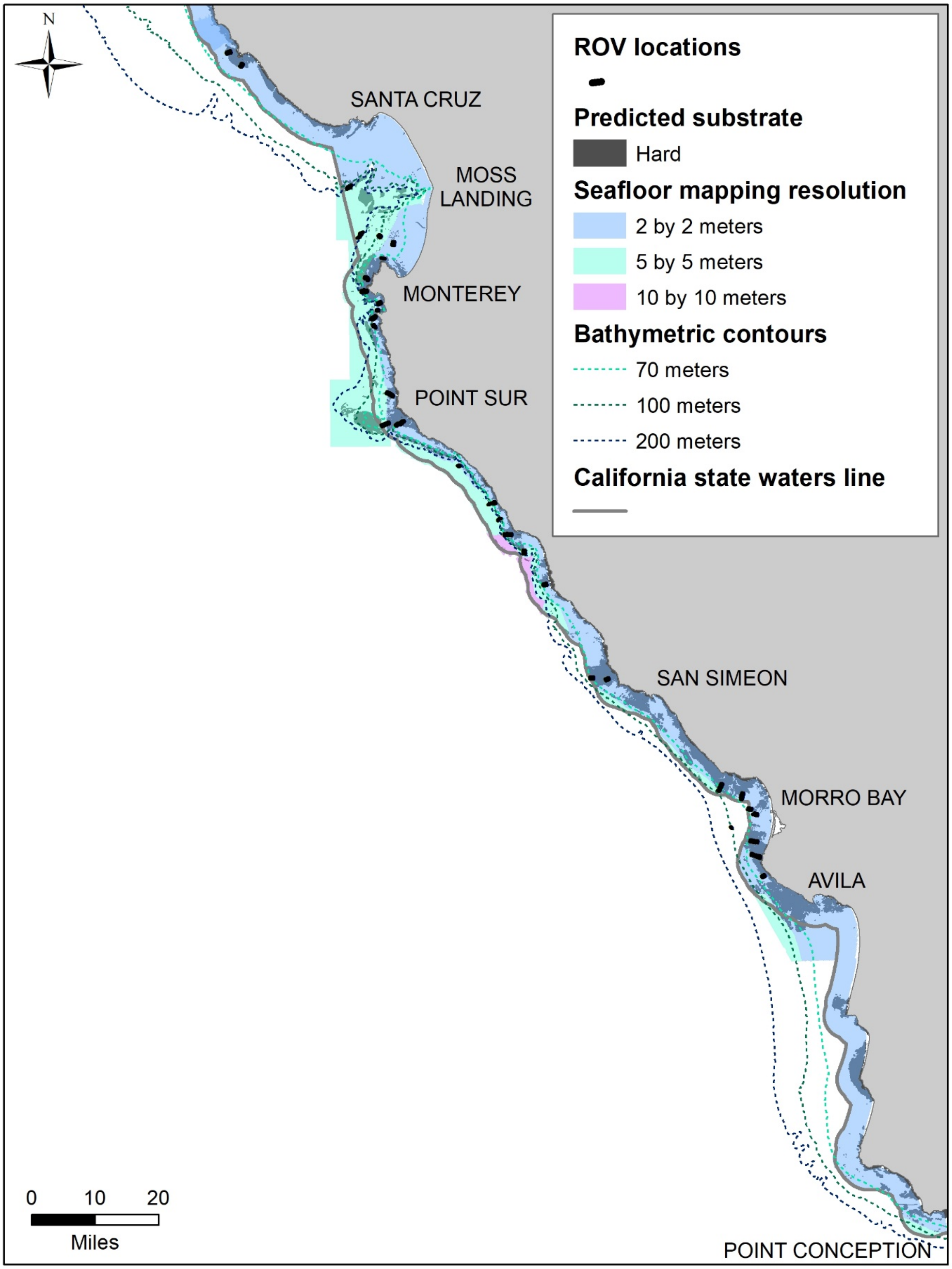

Figure 4. The South-Central California coast from Pigeon Point to Point Conception showing the distribution of hard substrate and degree of ROV coverage from sampling locations along the coast.

Considerations Concerning Poststratification of Survey Transects and Selected Sample Units
ROV transects are typically 500 m in length and composed of individual data points known as microframes collected at a one second interval. For the purposes of this analysis, transects were split into smaller segments with the goal of pairing fish observations with distinct habitat types.

The continuous collection of data by the ROV along each transect offered flexibility in defining subunits or segments for further analysis. The "legacy" data set, originally derived for the purpose of evaluating density inside and outside of MPAs, was composed of 25 m^2 constant area segments that could be combined into 100 m^2 subunits. This resulted in segments of varying length with discontinuities due to stop pulls and backsides. Given the intent is to evaluate not only density, but to analyze the correlation of density with various habitat variables and other factors, alternative sampling units were considered that better suited the purposes described below. Another consideration was the degree of spatial error of ROV observations relative to the CSMP data, which presented the need to address the spatial scale at which terrain attributes could be derived. Provided below is a qualitative evaluation of the factors leading to selection of the preferred sampling unit for analysis followed by a description of the sampling units considered.

Fixed Length Sampling Unit Used in Analysis
To provide sampling units sufficiently uniform in spatial coverage to allow representative habitat characteristics to be derived from seafloor mapping, a fixed 20 m fixed distance sampling unit was employed. Factors considered in the selection of the sample unit included the units of expansion, articulation with the statistical framework, potential for spatial autocorrelation and spatial error in the covariates, which were evaluated qualitatively in Table 1. The 20 m sampling unit was believed to provide a balance of the need for the sampling unit to be large enough to encompass the habitat of relatively sedentary rockfish species and the scale of spatial error, while being small enough to contain unique habitat characteristics so that associations between habitat and species observations could be examined.

ROV transects were divided into 20 m segments to increase the likelihood of the segment encapsulating just one habitat type as well as to align with the resolution of the seafloor mapping raster data and the GIS tools used to derive terrain attributes. See the Seafloor Mapping Data Aggregation and Neighborhood Size section below for a description of seafloor mapping data resolution and neighborhood size used to derive terrain attributes.

Unusable microframes, which are data points that resulted from stops, pulls, back sides, or any other event that renders the observation unusable, were included when grouping data into 20 m segments. This ensured that the segment did not represent greater than 20 m of distance travelled by the ROV, which increases the likelihood of each segment representing one habitat type. Exclusion of unusable microframes would break the segment into sub-segments that have differing gap sizes and the varying gap size may mean that the 20 m segment would actually represent more than 20 m of distance travelled by the ROV.

Consideration was also needed for transects that were not exactly 500 m in full length and therefore did not divide into perfect 20 m segments. At times this resulted in a segment at the end of the transect that was shorter than 20 m. Segments made up of less than 60% of usable data or that were less than 12 m in length were excluded from the analysis.

ROV height-off-bottom (HOB) is related to measurement error, and because it is directly related to transect width, is also influential on the species-specific probability of detection. Programs should review needs for a consistently restricted range of transect width. Evaluate existing data to determine if

a bias result from ROV HOB being greater over the most rugose reef and/or reef with high and abrupt relief (including 'backsides'). (Dr. Williams)

We examined the correlation between HOB and density which showed density to be significantly correlated for most species of interest, which are all associated with rocky reef habitat (Table A). The HOB was also positively correlated to terrain variables which represent greater habitat rugosity requiring the ROV to be flown at a greater height off the bottom resulting in a wider FOV and transect width. It is likely that the increase in density at greater height off bottom is due to the preference of the subject species for more complex habitat. Video imagery from transects with a lower height off bottom is usually better illuminated and allows the observer greater visual acuity to detect cryptic fish. Thus, apart from behavioral interactions with the ROV, it is likely that detection probability is greater at lower HOB (narrower transect width) despite the higher densities at greater HOB observed as a result of correlations resulting from operating logistics requiring greater height off of bottom associated with the rocky reef habitat that the species of interest are associated with.

The range of transect width was constrained to a minimum of approximately 0.5 m by the physical arrangement of the camera on the ROV. At the target HOB of 0.3 m the transect width is approximately 1.5 m. The maximum width was restricted by maintaining the target HOB during transects and post survey by filtering out data where the forward ranging sonar exceeded 4 m (3.4 m transect width) for more than 6 consecutive seconds. Under usual operational conditions and low to moderately rugose habitat the average transect width ranged between 1.5 and 3 m. In high rugosity the average width for transect segments ranged between 2 to 3.4 m. This protocol was developed to balance the need to cover area faster (greater HOB and speed) with the need to stay closer to the seafloor at a slower speed to capture better quality imagery to aid species identification. Further restricting this range will create a trade-off between these objectives. Restricting the upper range will result in increased removal of transect data in high relief habitats thereby further fragmenting transect segments. The use of sampling protocols, filters and cutoffs affecting the range of transect widths used in analysis help minimize potential bias from differential probability of detection.

We also tested for correlations of density with measures of the degree of "backsides" (where the ROV lost sight of the bottom after going over high relief habitat) along a transect, including the frequency, distance and area of backsides for species of interest. There were few correlations for quillback, brown and vermilion rockfish for some of the variables (Table B). These correlations should be expected given the association of this phenomenon with rocky reef habitat these species are associated with. The lack of correlation for some species and inconsistent results for others across regression methods (GLM vs GAM) or measures of the degree backside (Table B) did not support a strong effect of the omission of the backsides from segments on the probability of detection as measured by the density of species of interest.

Table A. Results of GAM and GLM of density with height off bottom with a negative binomial distribution (*** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, <0.1 is nearly significant).

Species	GAM	GLM
Gopher Rockfish	-	-
Copper Rockfish	***	***
Quillback Rockfish	***	***
Brown Rockfish	***	***
China Rockfish	***	***
Vermilion Rockfish	***	***
Kelp Greenling	***	***

Table B. Results of GAM and GLM of density with the number, distance and area of backsides with a negative binomial distribution (*** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, <0.1 is nearly significant).

	GAM			GLM		
Species	Count of Backsides	Distance of Backsides	Area of Backsides	Count of Backsides	Distance of Backsides	Area of Backsides
Gopher Rockfish						
Copper Rockfish						
Quillback Rockfish	**			**		
Brown Rockfish	***	*		*		
China Rockfish					.	.
Vermilion Rockfish			*			
Kelp Greenling						

Sampling Units Considered but Rejected

Legacy Fixed Area

The original ROV dataset that was received at the start of this analysis included a field that grouped ROV data points into 25 m^2 segments. This fixed area segment designation was used in initial exploratory analyses but was rejected due to inconsistency of segment length. For this analysis, the center point location of each fixed area segment was the location used to pair each segment's fish observations with a cell from each terrain attribute's raster. Inconsistent segment length results in varying levels of accuracy when establishing the relationship between fish observation and habitat (i.e. a short segment is more likely to be accurately described by the habitat value at its center point than a long segment).

Fixed area segments vary in length for several reasons. The equal area segment designation is based on the summation of area values to 25 m^2 from consecutive microframes of usable data. Microframe area is calculated by multiplying swath width by distance traveled. The swath width varies depending on how far off the bottom the ROV is and will increase as the distance between the ROV and the substrate increases. In addition, the speed of the ROV influences the distance it traveled in one second. Though the intent is to maintain consistent ROV speed, it isn't always possible and can result in inconsistent segment length. Any unusable microframes were excluded prior to calculation of cumulative area. When unusable microframes are excluded it breaks the segment into sub-segments that have differing gap sizes that also result in varying segment length.

Microframe

Use of microframes that were collected by the ROV at one second intervals were considered but rejected. Microframes were of interest because analysis of the data at this scale would produce the largest possible sample size. In addition, each fish observation would be paired with terrain attribute values from the high-resolution raster cell that intersected each point which in theory would result in observation data that are paired with the best available habitat data.

The use of microframes as our sampling unit was rejected because it resulted in an abundance of data points with no observations of fish which may result in an excess of zero values that would skew the data distribution necessitating the use of zero inflated methods. In addition, the spatial error associated with the positions defining any given microframe is estimated to be between three and six meters. Therefore, the pairings of observations and habitat data at this scale would have been error prone.

Full Transect

Use of full 500 m ROV transects was considered after the fixed area segment and microframe approaches had been ruled out. This was an easy option to consider since the data were already grouped into transects and because the transect level would have a very low number of units with zero observations. The full transect level approach was rejected because habitat can vary over the full length of a transect which makes associations between fish observations and habitat impossible to evaluate effectively and a more reasonable segment length was preferred.

Table 1. Considerations regarding potential sampling units for ROV analysis

Method	Description	Units of expansion	Articulation with statistical framework	Spatial Autocorrelation	Spatial Error of Covariates	Short Segment Exclusion	Pulls and Backsides	Zero Values	Articulation with Seafloor Mapping
Basic Consideration		Consistent with expansion w/ seafloor data?	Allows unbiased estimation of density and abundance. Consistent with depth and bottom type categorization.	Results near each other correlated due to unaccounted for behavior or habitat etc.- bootstrap resolved?	Terrain attributes or inferred values from seafloor mapping accurately represent transect characteristics.	Not result in fragments of various sizes with differing detection probability/loss of data from unused info	Unbiased by exclusion of pulls and backslides due to concatenated transect	Observed area not so small that positive too uncommon	Resulting variables can be defined in the seafloor for expansion i.e. continuous vs. discrete
Area Based Method	25m square with varying distance and width	Area is fixed	Width and time significant, concern as variables.	TBD Bootstrap resolved.	Ascribing to centroid limited resolution	Significant loss due to concatenation	Common due to fixed area	Smaller = more	Feasible fixed area or variable factor selected
Distance Based Method	Fixed Distance Transect	Area is a variable	Width will vary and requires a variable	TBD Bootstrap resolved	Ascribing to centroid limited resolution	Loss depends on length of fixed unit	If fixed length used, longer = more	Shorter = more	Feasible area as variable factor selected
Microframe	Uses 1 second observations 2 m by 2 m	Area is fixed	Percent rock not well informed	TBD Small frames more subject to spatial autocorrelation bootstrap resolved.	Low since unit is very small, but potential for false correlation if sample size low.	Low as small units prevents need for concatenation of segments	Non-issue due to small segment size	Far more zero value units, distribution	Feasible fixed area
Transect	Uses distance from transect line	Length and detection function constant over small width in our case	Too large of an area to associate the presence of fish with the habitat variables.	TBD	Continuous units of observation make evaluation of covariates difficult.	Less of an issue since units are continuous.	Non-issue as they are continuous units in strata i.e. depth range	Longer continuous unit less frequent.	Feasible but as transect lines across habitat rather than discrete area expansion

Were backsides of high-relief areas included in the length of transect calculation or were those areas removed or accounted for in other ways. (Dr. Berger)

ROV observation points representing backsides of high-relief areas are flagged as unusable in California's ROV dataset when the forward ranging sonar exceeded 4 meters for more than 6 consecutive seconds. The 20 m segments used for California's analyses were generated by grouping consecutive ROV observation points (both usable and unusable) until the resulting segment was 20 meters in length. After the data points were grouped into 20 m segments the unusable data points were removed from the 20 m segment meaning that these data points did not contribute to the surveyed area or to the counts of fish observed. If removal of unusable points resulted in a segment length was less than 12 m (60% of potential segment length) then the segment was excluded from subsequent analyses. For comparison, Oregon's methods indicate that less than 10 m squared was the threshold used for exclusion of short segments. California's 12 m threshold translates to approximately 25 m squared.

The California dataset is composed of 10,248 20 m segments. Of those, 867 segments were removed as a result of their usable distance being less than 12 m. Table C shows the degree to which the presence of backsides contributed to the removal of 584 segments. Conversely, 2,052 segments with backsides were included in subsequent analyses because the usable distance was greater than 12 m after the removal of the unusable backsides. The lack of correlation for density with the number, length or area of backsides for most species and inconsistent results for others across methods or measures of the degree backside effect did not support a strong effect of the omission of the area on the probability of detection as measured by the density of species of interest (Table B).

Table C. Reasons for removing short segments from subsequent analyses

# Segments Removed	Reason for Removal
278	100% related to presence of backsides
217	>=75% and <100% related to presence of backsides
52	>=50 and <75% related to presence of backsides
17	>=25 and <50% related to presence of backsides
20	<25% related to presence of backsides
223	Segment was less than 12 m before removal of unusable points (i.e. segment at start or end of transect)
60	Unrelated to backsides (i.e. stops, pulls, etc.)

Examine the effect of small (short) segments generated by 'gaps' created when bad data are excluded or when transects are segmented by substratum polygons defining geological categories since small segments create the potential for very high fish densities as the total view area approaches zero. (Dr. Williams)

California's methods included removal of segments less than 12 m in usable length from subsequent analyses to mitigate against the potential for very high fish densities as the total view area approaches zero. See response to Dr. Berger above for a description of the 12 m threshold and for a comparison to the threshold used in Oregon's methods. Segmentation by substratum polygons defining geological categories does not apply to California's methods as this was part of Oregon's methods only.

Figure 2 from Oregon's report shows "examples of 20 m segment fish density data exhibiting trends at low segment sizes". Figure A below shows the relationship for species of interest in California from our study. One potential cause of smaller usable area in each segment is that backsides have been removed from segments with pinnacles, making them shorter resulting in a smaller area. In addition, as discussed previously segments that are closer to the bottom will have a lower usable area as a result of the reduced field of view and resulting swath width observed. While these factors affect the usable area observed in each segment, the phenomenon reflected in the Figures is the result of the geometric relationship of density with increasing area for a given observed number of individuals per segment.

Though the density will tend toward infinity, the cutoff for the minimum observed distance mitigates this. While efforts are made to maintain a constant height above bottom, logistic constraints result in variability. This could be addressed in the future by constraining the swath width reviewed in processing to the minimum distance observed, though this would sacrifice observed area. The average density with segment length shows that the smaller segments do not systematically result in higher densities, in part due to the larger number of segments with no fish observed at smaller areas observed. The Usable Area was significant when included in the preferred GLM and GAM for gopher rockfish with variables reflecting latitude, depth, bottom relief and bottom type and resulted in a marginally improved R-squared value (0.064 vs 0.60) in k-fold validation. Inclusion of the usable area as a variable could address the geometric artifact induced by variable segment size. Fixed area segment design could be employed instead of constant distance though this may have implications for the derivation of terrain attributes to derive variables to evaluate correlations driving densities.

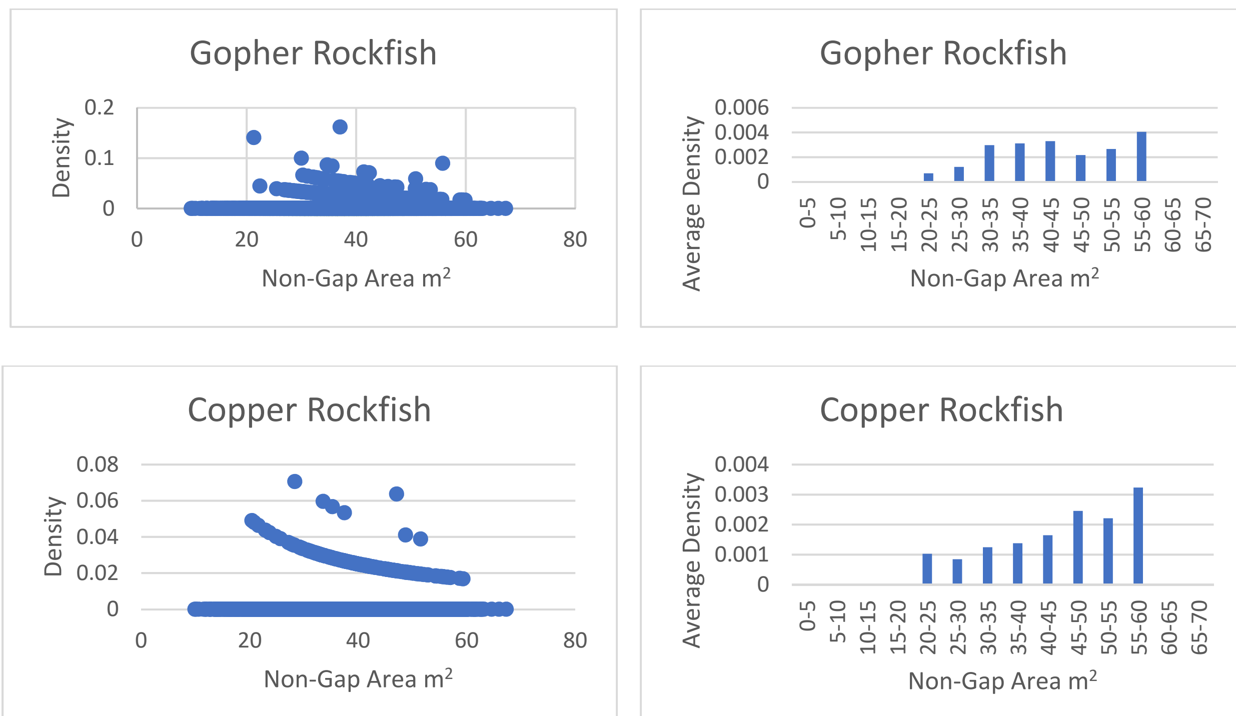

Figure A. Fish density versus usable segment area and average density with segment area.

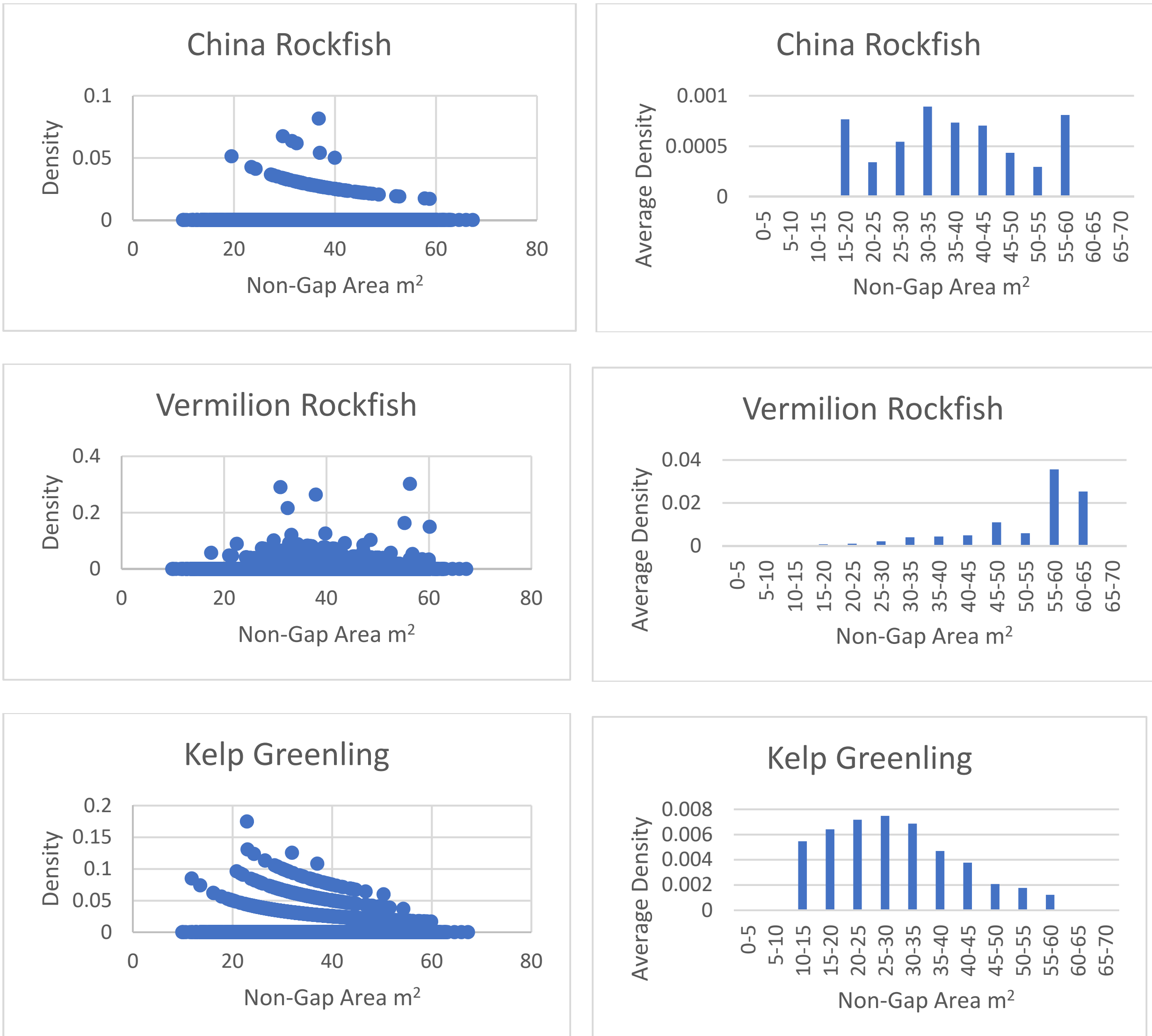

Figure A(cont.). Fish density versus usable segment area and average density with segment area.

Collection and Creation of Habitat Data to be used for Expansions

Use of California Seafloor Mapping to Derive of Terrain Attributes and Categorize Seafloor for Expansion

The CSMP provided raster-based estimates of depth and presence of hard bottom, not derived through traditional geologic interpretations but algorithmically defined using seafloor roughness as a proxy for determining areas likely to consist of rocky reef with significant relief in state waters. While seafloor is known for the path of the ROV from video observations, the degree of relief is not characterized. Terrain attributes can be derived from seafloor mapping depth data and paired based on location to the centroid of each 20 m segment of the ROV transects allowing for analysis of correlations of the fish observations with relief. In addition, the seafloor mapping provided the basis for expansion of density estimates in both design-based and model-based estimates of abundance. For the design-based expansions, the area of hard bottom habitat in each depth and latitude could be derived allowing estimation of the abundance of a given species. Model-based estimates of abundance relied on the seafloor area within a given latitude as well as derived terrain attributes, proportion hard bottom and depth from the seafloor mapping as the basis for expansion using the derived relationships between density and depth, latitude and proportion hard bottom and terrain attributes from GLMs. The Marine Geospatial Ecology Tool produced by Duke University used the raster-based grid values for factors and the model relating factors to fish abundance to produce an estimate of abundance analogous to the methods employed in Young and Carr (2015). The seafloor mapping and derived terrain attributes are described below, and further details of the modeling and expansion are provided in the analytical methods section.

California Seafloor Mapping Data

High-resolution seafloor mapping data were gathered in the majority of California state waters between the years 2007 and 2010 through a comprehensive mapping program managed by the California State Coastal Conservancy, Ocean Protection Council, CDFW, and the NOAA National Marine Sanctuary Program.

Multibeam bathymetric sonar was utilized to collect high resolution depth data presented as digital elevation models. Data collection met or exceeded International Hydrographic Organization Order 1 standards and was carried out at the maximum resolution obtainable using the best available equipment at the time. Resolution of bathymetric products is dependent on depth; two meters for data from the zero to 85 m depth range, five m for the 80 to 250 m depth range, and 10 m for the 230 to 1500 m depth range.

Seafloor Mapping Data Aggregation and Neighborhood Size

The spatial error of an ROV observation point is estimated to be between three and six meters. As a result, resolution of the seafloor mapping raster data used to derive terrain attributes was considered and selected at the same time the 20 m fixed segment length was selected.

Terrain attributes are derived from seafloor depth raster files using a variety of GIS tools. A commonality among these GIS tools is that each cell in the raster is assigned a value based on calculations performed on a neighborhood of cells, surrounding each cell. Some GIS tools allow the user to choose the neighborhood size while others are hard coded to use a three by three cell neighborhood.

Choosing a cell size for terrain attributes derived from seafloor mapping data was a balance of aggregating the cells from their original two by two meter resolution to a size that would create a sufficiently large neighborhood for the GIS tools that are hard coded with a neighborhood size of three by three cells while not aggregating so much that the depth values and terrain characteristics derived from depth data are oversimplified.

Terrain variables were created based on a five by five cell neighborhood of six by six-meter cells when GIS tools allowed for specification of a neighborhood size other than three by three cells. This resulted in a neighborhood size of 30 by 30 m. Since 20 m plus or minus a spatial error of three to six meters is 23 to 26 m this means the terrain that influenced the fish observations made by the ROV would be captured by the neighborhood of cells even while considering the possible offset due to spatial error. In addition, 30 by 30 m is a reasonable amount of area to consider when generating terrain variables for rockfish that exhibit high site fidelity and allows the habitat therein to be adequately characterized at an appropriate scale. A depiction of a 20 m segment of ROV transect superimposed on the 30 by 30 m neighborhood of six by six-meter CSMP raster cells used to evaluate terrain attributes is provided in Figure 5 as a visual aid.

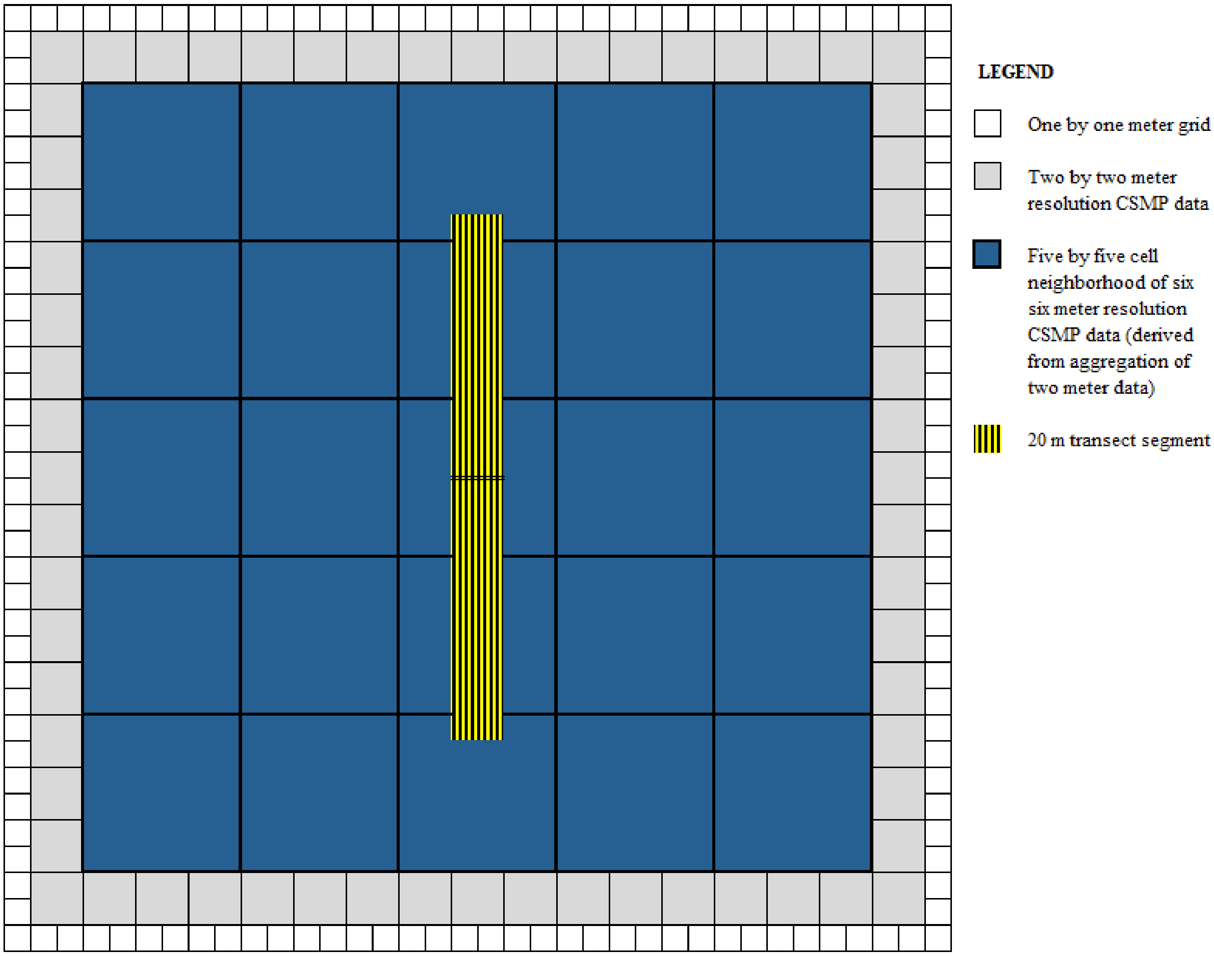

Figure 5. Depiction of our sampling unit (20 m segment of ROV transect) centered above the 30 by 30 m, five by five cell neighborhood of CSMP data used to derive terrain attributes when GIS tools allowed use of a five by five cell neighborhood.

When GIS tools required a three by three cell neighborhood size, this resulted in an 18 m by 18 m neighborhood size. Approximately two meters (and possibly more as a result of spatial error inherent in the ROV data) of the 20 m segment is not characterized by the 18 by 18 m neighborhood used to derive the terrain value. While not ideal, this approach was chosen because the alternative was to aggregate the raster data to eight or ten meters resulting in increased smoothing of the depth data which may also impact results. A depiction of a 20 m segment of ROV transect superimposed on the 18 by 18 m neighborhood of six by six-meter CSMP raster cells used to evaluate terrain attributes when a three by three cell neighborhood was required is provided in Figure 6 as a visual aid.

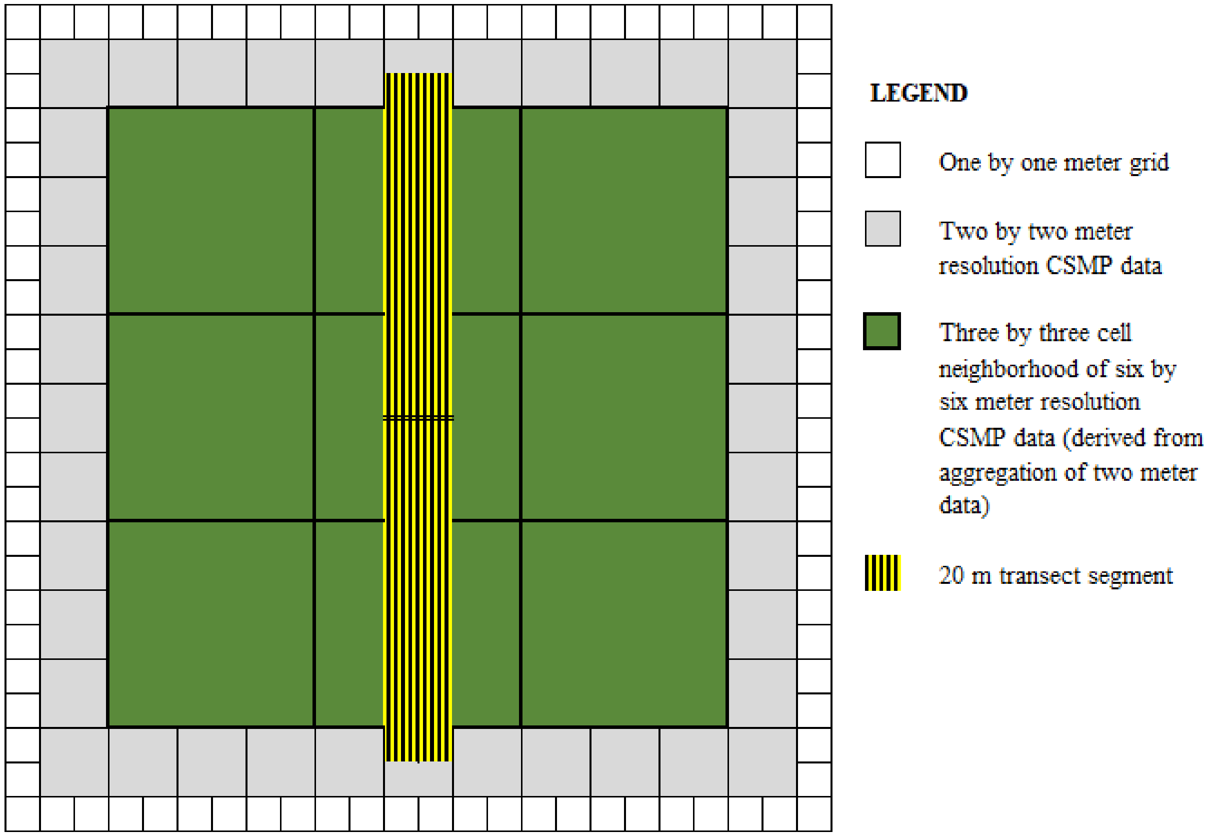

Figure 6. Depiction of our sampling unit (20 m segment of ROV transect) centered above the 18 by 18 m, three by three cell neighborhood of CSMP data used to derive terrain attributes when GIS tools required use of a three by three cell neighborhood.

Terrain Attributes

Many rockfish species are known to inhabit rugose high to moderate relief rocky reef habitat. As a result, terrain attributes that describe or are associated with relief may be found to have a predictive relationship with rockfish observations. Centroid points were generated from 20 m ROV segments. These centroids were intersected with raster files of terrain attributes using the ArcGIS Extract Multi Values to Points tool to facilitate analysis of the relationship between terrain attributes and rockfish observations. A variety of terrain attributes were considered but only a subset were used in this analysis.

Sources of Error in Associations between ROV Observations and Terrain Attributes

Unfortunately, associations between California ROV observations and terrain attributes are subject to possible spatial error due to the 3 to 6 m of spatial error of individual ROV microframes as well as the fact that the observations associated with the 20 m segment may have been made to either end of the segment as opposed to at the same location as the centroid. The best-case scenario is that all fish observations actually took place at the exact location of the centroid so that there is no spatial error or deviation. In this case the fish observations would be perfectly aligned with the cell at the center of the neighborhood from which the terrain attribute was calculated. The worst-case scenario would be if a segment is made up of locations with 6 m spatial error and all the observations occurred at one end of the segment. If this hypothetical segment were paired with a terrain variable derived from a five by five cell neighborhood of six by six meter cells, the neighborhood of cells around the actual location of fish observation may intersect as few as 15 of the 25 raster cells used to derive the terrain attribute paired with the observation. If this segment were paired with a terrain variable derived from a three by three cell neighborhood of six by six meter cells, the neighborhood of cells around the actual location of fish observation may intersect as few as three of the nine raster cells used to derive the terrain attribute paired with the observation.

Terrain Attributes Used in Analysis

Depth range, standard deviation of depth, slope, standard deviation of slope, and surface area to planar area (an estimate of rugosity) were included when making associations between terrain variables and ROV observations. These terrain variables are derived from the cells in the neighborhood surrounding the center cell, give equal weight to all the cells within the neighborhood, and therefore describe the habitat of the entire neighborhood. This is important for this analysis because the location of ROV observations is subject to three to six meters of spatial error which means the ROV observations may actually be up to six meters to either side of the terrain attribute cell that it intersects.

Depth range is the difference between the maximum and minimum cell depth in the cells that make up the neighborhood used to derive the value and was derived from raster data of depth using the ArcGIS Spatial Analyst Focal Statistics tool with statistics type set to range.

Standard deviation of depth is the standard deviation of the depth values in the neighborhood that is around and includes the center cell and was derived from raster data of depth using the ArcGIS Spatial Analyst Focal Statistics tools with statistics type set to STD.

Slope identifies the gradient or steepness for each cell of a raster by calculating the maximum rate of change in value from the center cell to its neighbors and was derived from raster data of depth with the ArcGIS Spatial Analyst Slope tool.

Standard deviation of slope is the standard deviation of the slope values in the neighborhood that is around and includes the center cell and was derived from raster data of slope using the ArcGIS Spatial Analyst Focal Statistics tool with statistics type set to STD.

Surface area to planar area represents the ratio of surface area to planar area in a three by three cell neighborhood surrounding each cell and was derived from raster data of depth using the Benthic Terrain Modeler for ArcGIS Surface Area to Planar Area tool.

Attributes Considered but Not Used in Analysis

Bathymetric position index, curvature, and relative difference to mean were considered but not included. Bathymetric position index, curvature, and relative difference to mean also describe rugosity and/or a location's position relative to a neighboring location (i.e. is a location on a ridge or in a depression) but were not included since they describe how a the position of a single cell compares to the position of cells that surround it as opposed to describing the variation amongst all the cells in a neighborhood, which is more robust to spatial error in characterizing relief in the habitat. Use of terrain attributes that describe a cell would be inappropriate for this project due to the three to six meters of potential spatial error associated with each ROV observation point. In addition, terrain attributes are paired with ROV segments through the segment's centroid location. Therefore, it is preferable that the value of the cell that intersects the centroid be representative of the entire 20 m segment length as opposed to describing just the centroid of the segment.

Bathymetric position index values are the product of the comparison of a specific cell's depth to the mean depth of the cells in the surrounding neighborhood and are derived using the Benthic Terrain Modeler tool for ArcGIS. A positive BPI value would indicate that the cell at the center of the neighborhood is at a greater elevation than the cells in its neighborhood. But the ROV observation may have been made over the cell three to six meters to either side of the center cell so the BPI value applied would be inaccurate for the observations.

Curvature is also specific to the unique location of each cell because it compares the cell at the center of a neighborhood to the eight surrounding neighbor cells to specifically describe if the individual cell is concave, convex, or flat. It was derived from depth raster data using the ArcGIS Spatial Analyst Curvature tool. It was also excluded due to the spatial error of ROV observations.

Relative difference in depth to mean is determined by dividing the difference between the neighborhood's mean depth and the center cell's depth value by the neighborhood's depth range using a python script. This terrain attribute also puts extra emphasis on the center cell when comparing it to the neighborhood that surrounds it and will also be excluded because of the possible spatial error.

Evaluation of Variable Detection Probability

Concerns regarding the implications of variable detection probability for density estimates between sites or transects raised by the SSC (Agenda Item E.3, SSC Report, September 2017) have been addressed in part through criteria for adequate sampling conditions, the sampling methodology itself and post processing methods. Video collected data was only used for density calculations when visibility was sufficient to view the entire video field of view at least 2 m in front of the ROV. During the course of a transect, the angle of the ROV camera relative to the substrate was adjusted by the pilot to maintain an oblique field of view with the horizon slightly below the top of the viewing area thereby insuring that fish behaving evasively in front of the ROV could be detected.

The behavior of observed fish and the distribution of distance of observations from centerline of the field of view can be examined to inform whether the behavioral response of a given species may have implications for detectability, though this would require a rescoring of the recordings for this specific purpose, which is time and cost prohibitive. In review of the recordings from our study, the overall behavior of encountered fish provided some indication of notable wariness or attraction to the ROV of a given species. The vast majority of demersal rockfish were found to be relatively unresponsive (MARE personal communication). Cabezon, treefish and California scorpionfish were relatively cryptic potentially affecting detectability. Schooling rockfish species such as blue, black or yellowtail rockfish were unavailable to the ROV in mid-water making the ROV based methods poorly suited to estimating their abundance without supplemental acoustic data and potential changes to the sampling methodology.

Fish that were scored as unidentified rockfish (UI Rockfish) made up a significant portion of observed rockfishes. A fish was scored as UI Rockfish when there was enough visual information such as size, body shape, color and fin orientation but not enough to confidently assign to the species level. The majority of the UI rockfish were unidentifiable because they were too far away from the ROV to be adequately illuminated and were most often smaller schooling individuals positioned off the seafloor. For this reason, we are confident that most of the unidentified rockfish are likely schooling and smaller species (blue/deacon rockfish, black rockfish, halfbanded rockfish, squarespot rockfish, widow rockfish, and yellowtail/olive rockfish). In areas with high observed abundance of these smaller schooling species UI rockfish counts also tended to be high, implicating them as a source of UI rockfish.

While the methods and selection of appropriate subject species address some of the potential issues relative to variable detection probability, the response of fish beyond the view of the ROV is unknown. Other studies conducted by the Alaska Department of Fish and Wildlife (Green et al. 2013) and National Marine Fisheries Service (Laidig and Yoklavich 2013) provide some insight on the degree species respond to ROVs in way that may affect detectability. Our analysis in the following section provides a review on a species by species basis taking into account the distribution and behavior of each species as well as results of other research conducted to date.

Further analysis of the research informing the presence and degree of variable detection probability and the appropriateness of application of ROV based survey methods to nearshore species were evaluated using the average density of fish at various transect widths across a number of species as depicted in bar plots in Figures 7. Limited data was available for widths less than 1 meter since the intent of the ROV survey design was to keep the ROV off the bottom. In addition, consistent observations of greater than four meters width over more than six seconds were eliminated due to the potential for shadows and objects in the foreground to obscure fish. For gopher rockfish, densities were relatively consistent across the primary range of transect widths (between 1.5 and 3 m) providing an indication that detection probability was consistent across segments. Either thresholds can be set to exclude segments outside this range of widths from analysis. Alternatively, transect width could be included as a variable to account for the effect on density estimates, though it is correlated with the effort in the denominator. Thus, we chose to relegate density estimates used in our estimation of abundance to the aforementioned range over which detection probability was assumed to be consistent.

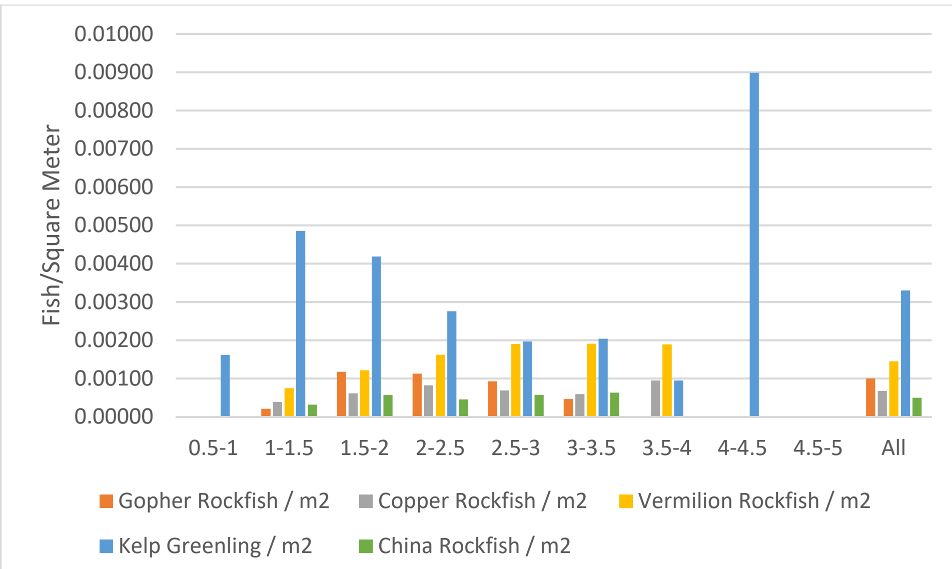

Figure 7. Fish per meter squared with segment width in meters for gopher rockfish, copper rockfish, vermilion rockfish, kelp greenling and China rockfish.

Literature on the topic of ROV selectivity (the reaction of fish to the platforms) should be reviewed, and the sampling design refined, and experiments conducted, if necessary, to reflect knowledge and knowledge gaps. Evaluate the Stoner et al. (2008) and Sward et al. (2019) reviews. (Dr. Williams)

Observations of fish reaction to our ROV appear similar to reported findings of other west coast studies and reviews (Laidig et al. 2013, Stoner et al. 2008, Sward 2019). In our surveys, we targeted a shallower assemblage of rockfish species than Laidig (2013) but the general findings of schooling vs. bottom associated species were similar to our observations. We have not quantified these reactions, but our observations indicate that the more sedentary species (gopher, quillback, china, and brown) rockfishes show little reaction to the ROV until a close distance (< 2m) and the schooling species, bocaccio, canary, blue, black, yellowtail and olive rockfish started to react at greater distances. Vermilion, yelloweye and copper rockfish showed an intermediate reaction response with no obvious avoidance reactions at distances that may lead to underestimation. Similar to observations reported from Oregon ROV surveys in Stoner et al. (2008), juvenile yelloweye rockfish seem to exhibit more frequent avoidance reactions than larger individuals. Non-rockfish species such as California sheephead, rock wrasse, kelp bass, and barred sand bass exhibit greater avoidance reaction and at greater distances that likely results in some missed observations.

In general, we have observed that most rockfish species we encounter have minimal response to the ROV until approached very closely, by which time they have been identified and counted. Strong reactions occur if the ROV bumps the substrate or makes other sudden low frequency mechanical noise. While stationary or moving very slowly, rockfish will attract and aggregate around the ROV. Steady smooth operation of the ROV results in less fish reaction while jerky movements or rapid aggressive thruster inputs and other noisy equipment causes avoidance reactions. To minimize these reactions, we chose to implement a constant speed for transects and use an autopilot thruster control to smooth the flight of the ROV and reduce pilot tendency to drive erratically or slow down to view fish or speed up during boring stretches.

Because our observations have agreed with other studies and our focus has been broader ecological analyses of MPA performance, we have not performed detailed analysis of species-specific reactions and limited this analysis to non-schooling and minimally reacting species. With ongoing funding for stock assessment applications, quantitative assessment of species-specific vehicle interactions could be made.

Circumstances and Species to which ROV Methodologies are Applicable

Whether indices of abundance or estimates of abundance are applicable to a given species is based on the depth distribution of the ROV survey, the depth distribution of available seafloor mapping data, the species depth distribution, and behavior of a species. The geographic scope of the analysis was relegated to the area north of Point Conception to the Oregon/California border due to the differing environmental conditions in the Southern California Bight for which comparisons of associations between regions should be conducted in the future. Thus, species for which the range extends south of Point Conception may be subject to limited areas of inference preventing full census or representation of trends in abundance to the south. High resolution seafloor mapping data collected by the CSMP is only available in state waters at present, thus species whose depth distribution includes a large proportion outside of the mapped area are not good candidates for estimation of abundance.

The predominant depth range sampled in this survey is less than 150 m, which encompasses the depth distribution of nearshore rockfish species allowing both indices of abundance and estimates of abundance to be estimated. A significant proportion of the depth distribution of shelf rockfish species is encompassed allowing estimation of indices of abundance. Schooling species that are semi-pelagic (i.e. black rockfish, blue rockfish and canary rockfish) present difficulties as a result of variable detection probability given that the ROV focuses on the sea floor. Demersal species that are not cryptic and do not exhibit avoidance thus evading detection are good candidates. The species also needs to be sufficiently unique to be easily identified without potential for misidentification and the frequency of presence in ambiguous classifications or unidentified rockfish groupings should be minimal. Consideration of these factors lead to the selection of the groundfish species in Table 2 to be considered for development of indices of abundance or abundance estimates.

Of these species, we focus on gopher rockfish, copper rockfish, vermilion rockfish, kelp greenling and China rockfish given their depth distribution, the depth distribution of sampling, and species behavior. Gopher rockfish is being assessed in 2019 making it the primary focus of our analysis. Copper rockfish and vermilion rockfish will be assessed in 2021 making them priority species for analysis. Kelp greenling was frequently encountered, and a full stock assessment has not been possible by other means making it a high priority. China rockfish was included since it had previously been assessed and estimates of abundance could be compared to results of the assessment.

Table 2. Groundfish species with potential for an index of abundance or estimates of abundance from ROV data to be considered.

Species	Index of Abundance	Abundance Estimate
Brown Rockfish	X	X
Quillback Rockfish	X	X
China Rockfish	X	X
Kelp Greenling	X	X
Copper Rockfish	X	X
Gopher Rockfish	X	X
Blue Deacon Rockfish Complex	X	
Black Rockfish	X	
Vermilion Rockfish	X	X
Canary Rockfish	X	
Yelloweye Rockfish	X	
Lingcod	X	

Suggestion to fill in gaps in multi-beam data that limit where data can be expanded by exploring available coastwide or statewide habitat data. (Dr. William and Dr. Shelton)

High-resolution seafloor mapping data with resolutions of 2, 5, or 10 m have been collected for most California state waters, but this dataset contains two critical gaps which are the white zone and the offshore of the California state waters line.

The first gap is a 50 to 500 m wide band of unmapped seafloor shoreward of the 2 m resolution layer from shore to 10 to 15 m depth along the California coast referred to as the white zone. Data collection in this band was prevented by navigation hazards and technical limitations that prevented ship-board mapping, and turbid waters or obstructions that prevented successful remote sensing (Saarman et al. 2015). See Figure 1 for a visual example of the gap that exists shoreward of the 2 by 2 m seafloor mapping data along the California coast.

Contributors from The University of California Santa Cruz, California Ocean Science Trust, and California Department of Fish and Wildlife used CSMP and National Oceanic and the National Oceanic and Atmospheric Administration Environmental Sensitivity Index (ESI) shoreline habitat categorizations to generate predictive maps of substrate characteristics in this "white zone" through interpolation. Spatial interpolation methods were used to create a raster that is 30 m in resolution and that predicted the proportion of rock substrate in each pixel. Interpolations were not intended to precisely predict the locations of specific reef features, but rather to provide a general estimate of the amount of rock versus soft bottom that is likely present in each area. Dataset metadata indicates that predictive accuracy was highest for narrow bands of white zone and the accuracy of substrate composition estimates decreased as the width of the band of white zone increased. Metadata also indicates that these data are unlikely to be useful for any purpose where precise location of substrate features at scales less than 100's of meters is required.

Depth data that match the white zone's 30 m resolution grid are not available. The best depth data available for the white zone is likely the U.S. Coastal Relief Model that was created by NOAA's National Geophyscial Data Center. This region wide dataset has a resolution of 3 arc seconds which translates to a pixel size of about 90 m. Table D below shows that depth data for 407.6 kilometers squared in the 0 to 10 m depth range must be sourced from the higher resolution NOAA Coastal Relief Model depth data compared to only 147.5 kilometers squared from the 2 m resolution CSMP data, limiting the resolution of depth data available.

The second gap is the waters offshore of California state waters. Figures B through D below show that the footprint of CSMP mapping data does not extend to the maximum of several nearshore rockfish species' common depth ranges. Figures B through D include depth contours that represent the maximum of the common depth range for five nearshore rockfish species. The 120 m depth contour is also included because table 4 below shows that 120 m is the deepest depth at which the CDFW ROV dataset indicates substantial observations. Figures B and C show that there is a substantial gap between where CSMP data coverage ends and the maximum depth of nearshore rockfish observations along the northern and north-central California coast. Figure D shows that CSMP data are available for a large proportion of the depth range of nearshore rockfishes in the central California coast even though there are still some gaps.

The CSMP dataset does not include the federal waters offshore of the California state waters line because CSMP efforts were focused exclusively on state waters. This is due to fact that the data collection was largely driven by the need for data to inform the design of California marine protected areas network and the high cost of data collection and ground-truthing (approximately 35 million dollars). As a result, there are gaps in high resolution depth and habitat data that exist in federal waters in depths where deeper nearshore rockfish species are known to occur.

The only region wide source of habitat classification data that was found offshore of the California state waters line was a GIS dataset in vector as opposed to raster format that classifies federal waters as hard, mixed, or soft. This dataset was created as part of the recent groundfish Essential Fish Habitat (EFH) review. Dataset authors are Oregon State University, Active Tectonics & Seafloor Mapping Lab and NOAA Fisheries Northwest Fisheries Science Center. Here are some key points derived from visual inspection of this dataset or from the dataset's metadata:

- Some shapes in this dataset look like hand drawn shapes which infers that a regionwide raster version of this dataset does not exist.
    - Metadata for a postage stamp area sized dataset that fed into the EFH layer was located and supports this theory. The metadata says that side scan sonar for data were printed out in "PosterShop" and mylar sheets were placed over the printed layouts. Then expert marine geologists interpreted the areas using pencil to draw polygons characterizing features based on their knowledge of the geology of the area.
- The metadata and/or paper linked from the metadata (Goldfinger et al. 2014) indicate:
    - The data are not intended to replace local site mapping nor are they meant to suggest that all areas are equally well known.
    - Some types of questions and analyses might require a level of uniformity or detail that regional products cannot provide due to their mixed resolution, and heterogenous quality.
    - Additional surveys and mapping will likely be needed to either verify habitat type in data poor areas or provide greater detail about habitat patchiness at local scales.
    - Next step identified as part of this work is collection of regional high-resolution data for the continental shelf and upper slope which infers that high-resolution seafloor mapping data does not exist for these regions.

Like the white zone, the best region wide depth raster data for federal waters is likely the U.S. Coastal Relief Model described above that has an approximate resolution of 90 m. There are several postage stamp size areas of higher resolution data available for federal waters offshore of the California state waters line but nothing that could come close to representing the entire study area.

Table D below shows that depth data for 8,904.0 kilometers squared in the 0 to 120 m depth range (both in the white zone and offshore of the CSMP data) must be sourced from the higher resolution NOAA Coastal Relief Model depth data, while only 5,943.3 kilometers squared is available from the 2 m resolution CSMP data.

Also of note is that 4,706.7 kilometers squared in the 20 to 90 m depth range must be defined by the 90 m NOAA CRM dataset as a result of the fact that these areas were not surveyed as part of the higher resolution CSMP seafloor mapping effort (Table D). This seems counterintuitive since there should be a higher probability of shallower depths being in state waters and therefore captured by CSMP. This could be partially explained by the fact that there may be shallow areas outside of state waters. Another explanation could be that there are variations in characterization of depth between the high resolution CSMP data and the lower resolution NOAA CRM data in the same areas and this may lead the NOAA CRM to overestimate the area of shallow depths offshore of state waters.

The three charts in Figure E below provide an example of the differences between depth classified by NOAA CRM versus CSMP off Crescent City on the northern California coast. The chart on the left is NOAA CRM depth data grouped into 10 m intervals. The chart in the middle is 2 m resolution CSMP data grouped into the same 10 m intervals. The chart on the right displays the CSMP data over top of the NOAA CRM data. The maps show that NOAA CRM classified a large area as being in the 0 to 10 m depth range while very little area was classified in this depth range under the higher resolution CSMP classification. In addition, the areas classified in the 10 to 30 m depth range extend much further offshore under the NOAA CRM classification than under the CSMP classification. The likely explanation for this is that depths are more generalized under the lower resolution NOAA CRM data resulting in the offshore reef being captured as a smooth depth contour. Conversely, the high resolution CSMP data can capture the detail in variation in depth of the offshore reef since features are mapped at a 2 by 2 m resolution.

The data quality differences highlighted by this example must be carefully considered when deciding whether to use 90 m resolution NOAA CRM data to estimate rockfish abundance using the same methods applied to 2 m resolution CSMP data.

In summary data exists to fill the gaps both shoreward and offshore of the CSMP data but at a much lower resolutions or in vector format and at a lower quality for both depth and substrate classifications offshore of California state waters line.

Table D. Area in square kilometer from the dataset that best represents depth and habitat between the California Oregon border and Point Conception by 10 m depth intervals.

Depth (m)	Area (km) from Each Resolution of Data			
	Depths defined by NOAA CRM	**Depths defined by CSMP**		**Depths defined by NOAA CRM**
	White zone (km)	**2 meter (km)**	**5 meter (km)**	**84 meter (km)**
0 - 10	407.551	147.539	2.119	0.413
10 - 20	29.052	936.584	1.143	5.431
20 - 30	4.268	1,112.596	1.038	123.660
30 - 40	3.180	1,018.159	1.275	283.376
40 - 50	0.590	745.098	2.113	512.509
50 - 60	0.235	583.114	11.898	862.793
60 - 70	0.299	565.900	17.485	798.643
70 - 80	0.299	450.771	15.777	999.902
80 - 90	0.043	288.837	81.484	1,125.865
90 - 100	0.078	72.239	162.696	1,316.198
100 - 110	0.050	14.486	139.386	1,314.431
110 - 120	0.007	7.954	86.204	1,115.110
Total 0 to 120 meters	**445.651**	**5,943.277**	**522.617**	**8,458.333**

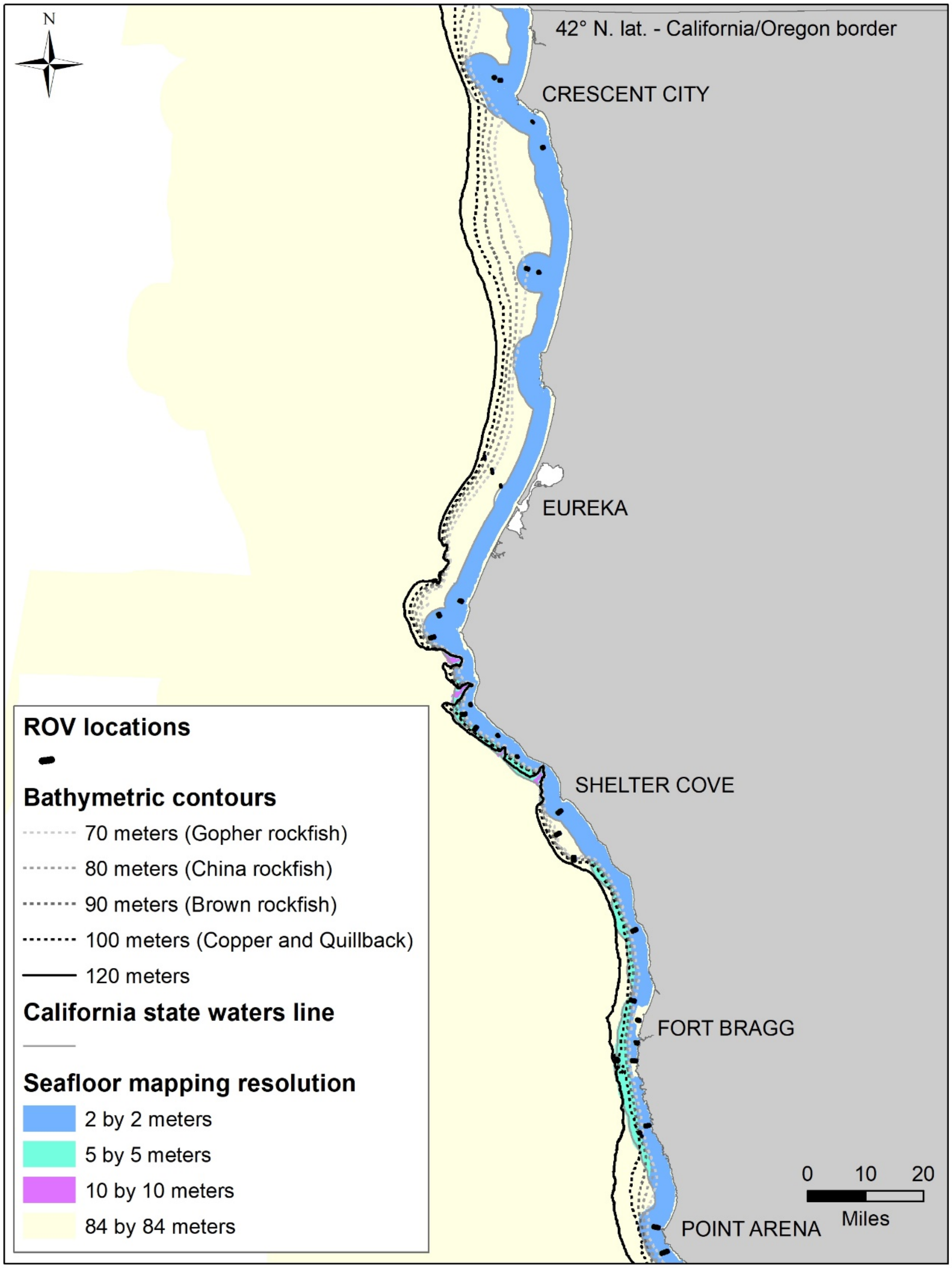

Figure B. The Northern California coast from the California/Oregon border to Point Arena showing the resolution of available seafloor mapping data compared to the maximums several of nearshore rockfish species' common depth ranges.

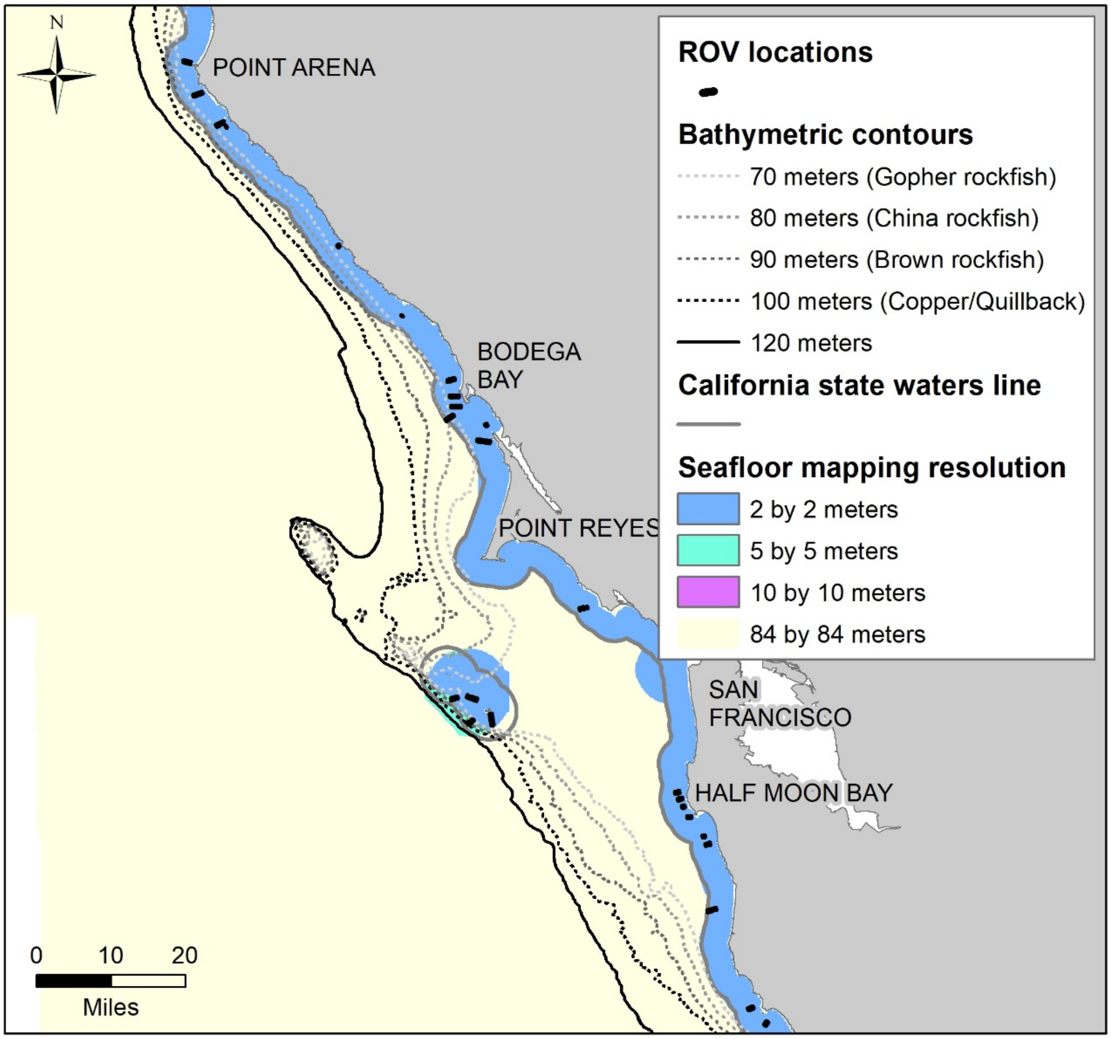

Figure C. The North-Central California coast from Point Arena to Pigeon Point showing the resolution of available seafloor mapping data compared to the maximums of several nearshore rockfish species' common depth ranges.

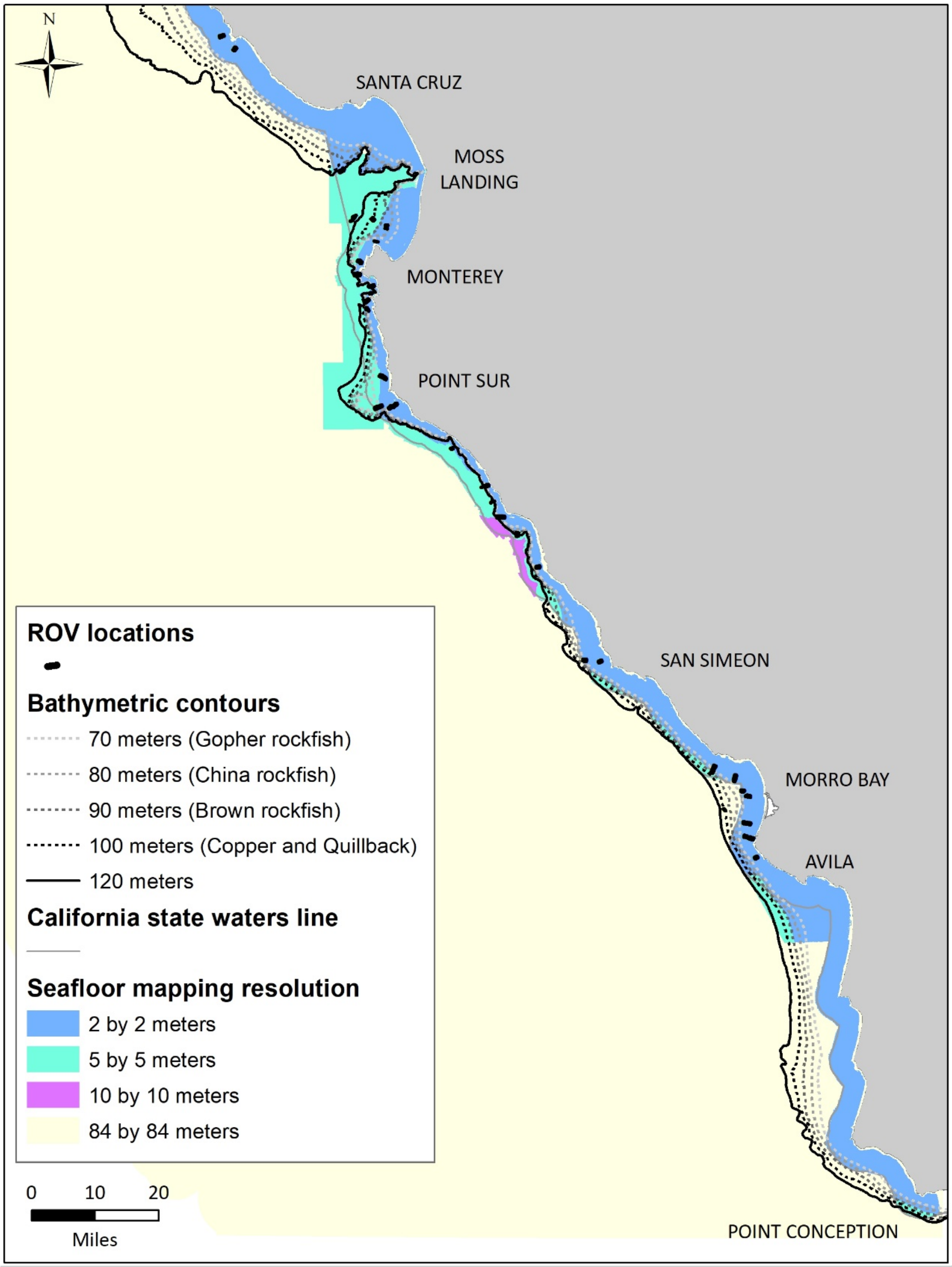

Figure D. The South-Central California coast from Pigeon Point to Point Conception showing the resolution of available seafloor mapping data compared to the maximums of several nearshore rockfish species' common depth ranges.

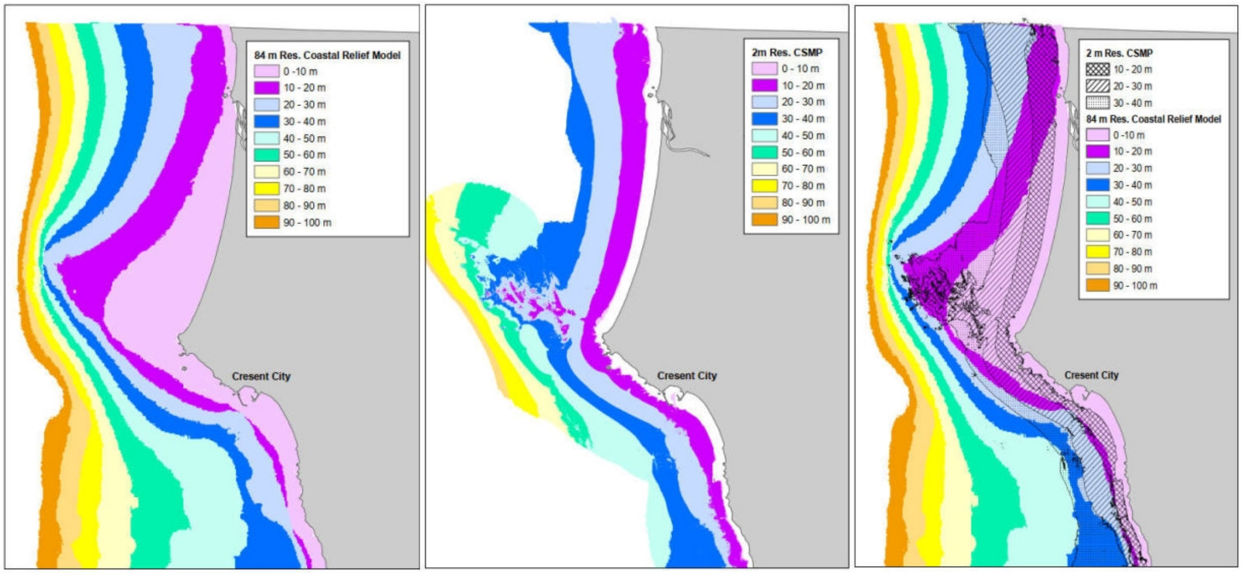

Figure E. NOAA Coastal Relief Model data (map on left), California Seafloor Mapping project data (map in middle), and NOAA CRM overlaid with CSMP (map on right) with depths displayed in 10 m intervals in the area of Crescent City off of the California north coast.

Analytical Methods

Data Sets Analyzed

The ROV study collected data on the depth and seafloor characteristics as well as the counts of various species in each of the approximately 20 m subunits. To prevent loss of a data from fragments that were not exactly 20 m in length, tolerances of 12 m or 60% of the target unit size were implemented. Seafloor attributes were available for a subset of the segments due to limitations in the spatial extent of the seafloor mapping data, though the exclusion of the remainder resulted in a substantial reduction in the number of available segments from 10,248 segments to 8,601 segments reducing the number of encounters with which to evaluate associations or indices of abundance provided in Table 3. As a result, for evaluation of GLM distributions and variable selection indices of abundance, both data sets were analyzed; one evaluating only variables available from the ROV survey and another with only those segments for which terrain attributes were available. The number of fish encountered in each 10 m depth bin for each candidate groundfish species is provided in Table 4 so that this information can be considered for future analysis of indices of abundance or abundance estimates.

Table 3. Number of positive segments and fish encountered with the full dataset from the ROV survey and the more limited data set including terrain attributes (TA).

Species	Positive Segments ROV	Number Encountered ROV	Positive Segments TA	Number Encountered TA
UI fish	330	8872	237	8505
UI rockfish	2256	6382	1541	3825
Cabezon	25	25	20	20
Brown Rockfish	203	246	196	239
Quillback Rockfish	231	256	127	142
China Rockfish	112	119	91	98
Kelp Greenling	1385	1683	1134	1395
Copper Gopher Rockfish Complex	14	16	14	16
Copper Rockfish	286	286	208	217
Gopher Rockfish	351	429	317	391
Blue Deacon Rockfish Complex	1303	13165	1102	10992
Black Rockfish	175	515	162	496
Black Blue Deacon Rockfish Complex	12	59	12	59
Vermilion Rockfish	624	1293	501	813
Sunset Rockfish	1	3	0	0
Canary Rockfish	795	2057	590	1654
Canary Vermillion Rockfish Complex	39	50	28	32
Yelloweye Rockfish	154	183	72	81
Lingcod	1775	2495	1276	1714

Table 4. Number of fish of each species encountered at each 10m depth increment.

Depth	Segments	Blue Deacon Rockfish Complex	Lingcod	Kelp Greenling	Copper Rockfish	Vermilion Rockfish	Yelloweye Rockfish	Gopher Rockfish	Quillback Rockfish	Black Rockfish	Brown Rockfish	China Rockfish	Canary Rockfish
0-10	2	1		1									
10-20	154	599	25	50		5		11		34		2	2
20-30	1719	3446	276	398	17	107	2	81	15	268	20	37	150
30-40	2571	3910	481	517	44	159	10	148	76	201	35	46	580
40-50	1461	3003	382	290	45	135	9	101	46	8	35	12	392
50-60	1387	1625	349	238	57	209	19	85	47		92	15	201
60-70	1310	248	285	106	43	188	13	3	37	2	47	6	255
70-80	835	326	207	52	32	157	47		20	1	13	1	215
80-90	759	5	292	27	41	81	57		13	1	4		190
90-100	473	1	143	3	27	229	13		2				40
100-110	104		11	1		17	6						15
110-120	26		12			4							16
120-130	17		7										
130-140	15		1			1							
140-150	11		1										
150-160	15												
160-170	17						3						
170-180	22		2										
180-190	7		1										
190-200	10						1						
200-210	8		1										
210-220	10		3			1	2						1
220-230	7	1											
230-240	6						1						
Grand Total	11190	13165	2491	1683	306	1293	182	429	256	515	246	119	2057

Indices of Abundance and Density Estimates

Methods: Variables relevant to poststratification informing design-based estimates of abundance and pertinent to normalizing indices of abundance were analyzed using GLMs. The models were applied to two data sets, one derived purely based on variables available from the ROV survey and the second a reduced data set reflecting only segments that terrain attributes could be derived for allowing further analysis. Segments from depths greater than 150 m were excluded from analysis since the number of segments collected beyond this depth declined and the focal species were not found in deeper depths. Count based GLMs, were used to test for significant correlations of density with depth, latitude, proportion hard or mixed habitat as observed from the ROV survey as well as "take" reflecting whether the segments were from MPAs for which harvest is prohibited or areas open to fishing. Area observed was included as an offset to account for differences in the area observed in determining the densities in each segment. Most of the MPAs have been in place for five to ten years, which is unlikely to be long enough for there to be a detectable difference due to recruitment or differential exploitation since they have been established.

Segment distance, width and time are expected to be correlated with area sampled in the denominator of density estimates. Similarly, time and distance are expected to be correlated with one another. A second GLM analysis including these variables in addition to the depth, latitude and proportion hard or mixed habitat and take was conducted to evaluate the potential for their influence on density estimates. A third GLM analysis was conducted including variables from the ROV and terrain attributes derived from the CSMP to evaluate correlations pertinent to index development including measures of relief. Lastly, a fourth GLM analysis was conducted with only those derived from the CSMP that can be applied in model-based estimation of abundance.

Tests for overdispersion were conducted using the AER library in R to evaluate the assumption of equal mean and variance in Poisson and binomial distribution models necessitating use of quasi-Poisson or negative binomial models capable of accounting for overdispersion due to greater than expected variance. Comparison of AIC values for zero-inflated Poisson distributions to standard Poisson distribution methods were analyzed to evaluate whether the high number of zero values from small subunits selected to evaluate associations with environmental variables, or inclusion of depths, or habitat where species are absent or uncommon resulted in skew greater than that represented by a regular Poisson distribution.

Five distributions were analyzed including Poisson, zero-inflated Poisson distribution, quasi-Poisson model, zero-inflated quasi-Poisson model and the negative binomial model. In addition, the binomial model was evaluated with counts converted to binomial data to evaluate association based on presence/absence. The models with alternative distributions were compared using AIC values, though they were not available for quasi-Poisson models precluding comparison to other models. Variables were tested for significant correlation with density for gopher rockfish, copper rockfish, vermilion rockfish, kelp greenling and China rockfish. In addition, deviance and selection in backward stepwise model selection were evaluated for gopher rockfish to evaluate consistency of identification of pertinent variables between methods since it is used as the focal species for evaluation of expansion methods. Deviance and backward stepwise variable selection could not be conducted with the zero-inflated models and thus was not provided.

Results: The results of GLM runs with the full and reduced data set are provided in tabular form for each species with each of the following sets of variables:

- Variables derived from the ROV
- Variables derived from the ROV and effort variables
- Variables derived from the ROV and CSMP
- Variables derived from the CSMP

Analysis of the results for each set of variables is provided below.

GLM Using Variables Derived from the ROV

For variables available from the ROV survey including latitude of the centroid of the observed segment, proportion hard/mixed bottom type observed along the transect, average depth along the segment and whether take was allowed in the segment, results varied by species and are provided in Tables 5 through 9. Across species the negative binomial distribution provided appreciably lower AIC scores except for copper rockfish for which the zero-inflated Poisson model had marginally lower AIC values (2404 vs 2391). Significant overdispersion was observed for all species, thus use of the quasi-Poisson or negative binomial distributions is preferable to the Poisson. The zero-inflated Poisson model resulted in a lower AIC value than the Poisson indicating zero-inflated models may be preferable or that reduced data sets excluding segments from depths greater than the distribution of the species would be beneficial. The binomial model had appreciably lower AIC values, though direct comparisons are not viable considering differences in the input data.

Latitude and proportion hard or mixed bottom type along the transect were significant for all species. Whether take was allowed in the segment was not significant for gopher and copper rockfish, but was significant across models for kelp greenling, and varied between distributions for vermilion and China rockfish. Given the recent implementation of marine protected areas, correlations are likely the result of siting of the locations rather than accumulation of abundance. Depth was significant for all models for all species except for vermilion rockfish for which depth was not significant for all distributions, which may be a result of its deeper depth distribution making it present across all depths examined.

The deviance measures for gopher rockfish were highest for latitude followed by proportion hard/mixed bottom type, then depth, while deviance was zero for take. Backward stepwise variable selection identified latitude, proportion hard or mixed bottom and depth as resulting in the lowest AIC values. These results were consistent with the significant correlation of the variables with density. These results indicate that all four variables available for analysis with the ROV survey observations alone should be considered in deriving indices of abundance. In addition, overdispersion should be addressed by using the quasi-Poisson or negative binomial models, and zero-inflated models or filtering may be beneficial to eliminate segments outside the distribution of the species in question.

Table 5. Results of GLM with various distributions for gopher rockfish density with variables derived from the ROV. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. The variables selected in backward stepwise regression are highlighted in yellow. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	**	***	**	***	***
Latitude	*** (338)	***	*** (338)	***	*** (299)	*** (271)
Proportion Hard/Mix ROV	*** (253)	***	*** (253)	***	*** (214)	*** (204)
Depth	*** (113)	***	*** (113)	***	*** (96)	*** (94)
Take	(0)		(0)			(0)
AIC	2864	2744	NA	2744	2690	2317
Overdispersion	1.25 (0.039)					

Table 6. Results of GLM with various distributions for copper rockfish density with variables derived from the ROV. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept		***		***		**
Latitude	***		***		***	***
Proportion Hard/Mix ROV	***	***	***	***	***	***
Depth	*	***	*	***	*	*
Take						
AIC	2559	2391	NA	2391	2534	2404
Overdispersion	1.13 (0.012)					

Table 7. Results of GLM with various distributions for vermilion rockfish density with variables derived from the ROV. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	***	**	***	**	***
Latitude	***	***	***	***	***	***
Proportion Hard/Mix ROV	***	***	*	***	***	***
Depth	***	***		***		
Take	***	***		***		***
AIC	9662	7017	NA	7017	5525	4103
Overdispersion	16.96 (0.02984)					

Table 8. Results of GLM with various distributions for kelp greenling density with variables derived from the ROV. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***		***		***	***
Latitude	***	**	***	**	***	***
Proportion Hard/Mix ROV	***	***	***	***	***	***
Depth	***	***	***	***	***	***
Take	***	**	***	**	***	**
AIC	8846	8557	NA	8557	8720	7111
Overdispersion	1.19 (2.69^-12)					

Table 9. Results of GLM with various distributions for China rockfish density with variables derived from the ROV. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	**	***	**	***	***
Latitude	***	***	***	***	***	**
Proportion Hard/Mix ROV	***	*	***	*	***	***
Depth	***	***	***	***	***	***
Take	**		**		**	**
AIC	1129	1085	NA	1085	1115	1067
Overdispersion	1.04(0.035)					

GLM Using Variables Derived from the ROV and Effort Variables

Variables available from the ROV survey in addition to the distance covered over the transect, average width of the segment and duration of time over the course of the segment were evaluated to examine whether there were significant correlations. The results varied by species and are provided in Tables 10 through 14. Though correlations with distance and width varied across species and model distribution within a species in some cases, time was consistently significant. Given that unit of effort is area observed, it is not surprising that distance and width would be significant. The time over which the segments were observed varies given the tolerances for speed and length of the segment observed, thus correlation with time is not unexpected either. We considered basing segments on fixed time, but due to the need for consistent spatial coverage for analysis of correlation with habitat characteristics, fixed distance was determined the unit of observation. Future development of indices of abundance may consider accounting for time, width or distance in normalizing indices. The relationships between duration of time over which a segment was sampled, area, distance and width are provided in Figures 8 through 12.

Table 10. Results of GLM with various distributions for gopher rockfish density with variables derived from the ROV and effort variables. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. The variables selected in backward stepwise regression are highlighted in yellow. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	***	***	***	***	***
Latitude	*** (338)	***	*** (338)	***	*** (304)	*** (237)
Proportion Hard/Mix ROV	*** (253)	***	*** (253)	***	*** (219)	*** (196)
Depth	*** (113)	*	*** (113)	*	*** (98)	*** (94)
Take	(0)		(0)		(0.2)	(0.09)
Time	*** (41)	***	*** (41)	***	*** (46)	*** (38)
Distance	(2)		. (2)		. (2.3)	* (3)
Width	*** (32)	***	*** (32)	***	*** (16)	*** (16)
AIC	2794	2622	NA	2622	2633	2287
Overdispersion	1.18 (0.017)					

Table 11. Results of GLM with various distributions for copper rockfish density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept		*		*		
Latitude	***		***		***	***
Proportion Hard/Mix ROV	***	***	***	***	***	***
Depth		***		***	.	
Take						
Time	***	***	***	***	***	**
Distance						
Width						**
AIC	2553	2388	NA	2388	2528	2395
Overdispersion	1.12 (0.012)					

Table 12. Results of GLM with various distributions for vermilion rockfish density with variables derived from the ROV and effort variables. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of theta and significance of overdispersion are indicated at the bottom of the table. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson Lo	Negative Binomial	Binomial
Intercept	**	***				***
Latitude	***	***	***	***	***	***
Proportion Hard/Mix ROV	***	***		***	*	***
Depth	.	***		***		
Take	***	***		***		***
Time	***	***	***	***	**	***
Distance		*		*		
Width	***	***	**	***		**
AIC	9205	6839	NA	6839	5394	4025
Overdispersion	16.13 (0.091)					

Table 13. Results of GLM with various distributions for kelp greenling density with variables derived from the ROV and effort variables. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***		***		***	***
Latitude	***	***	***	***	***	***
Proportion Hard/Mix ROV	***	***	***	***	***	***
Depth	***	***	***	***	***	***
Take	***	.	***	.	***	*
Time	***	***	***	***	***	***
Distance	.		.		.	
Width	***	***	***	***	***	*
AIC	8665	8417	NA	8417	8556	7038
Overdispersion	1.15 (5.3^-12)					

Table 14. Results of GLM with various distributions for China rockfish density with variables derived from the ROV and effort variables. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	**	***	**	**	**
Latitude	**	***	**	***	*	*
Proportion Hard/Mix ROV	***	*	***	*	***	***
Depth	***	***	***	***	***	***
Take	**		**		*	*
Time	***		***		***	***
Distance					.	
Width		**		**		
AIC	1105	1055	NA	1055	1080	1046
Overdispersion	1.03 (0.043)					

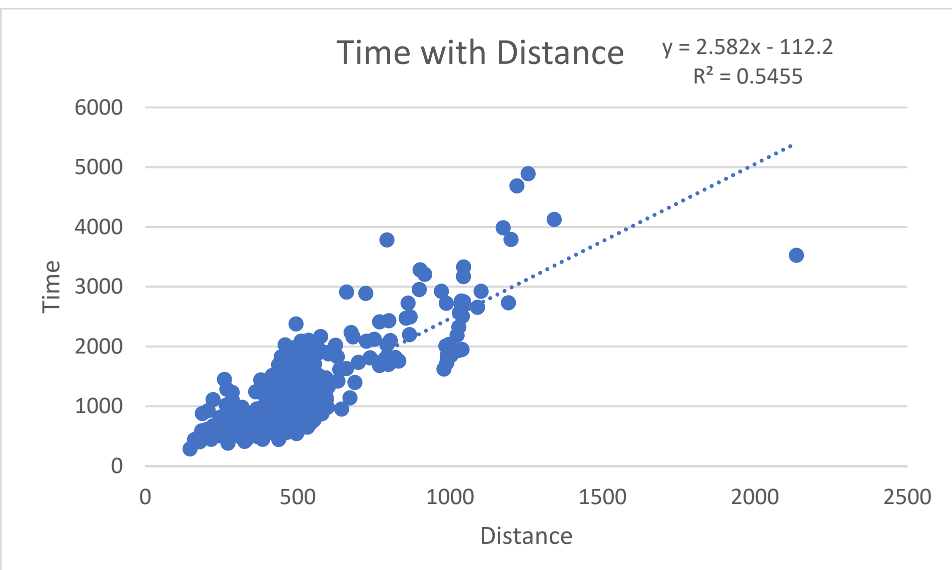

Figure 8. Duration of time (seconds) the segment was sampled with segment distance (m).

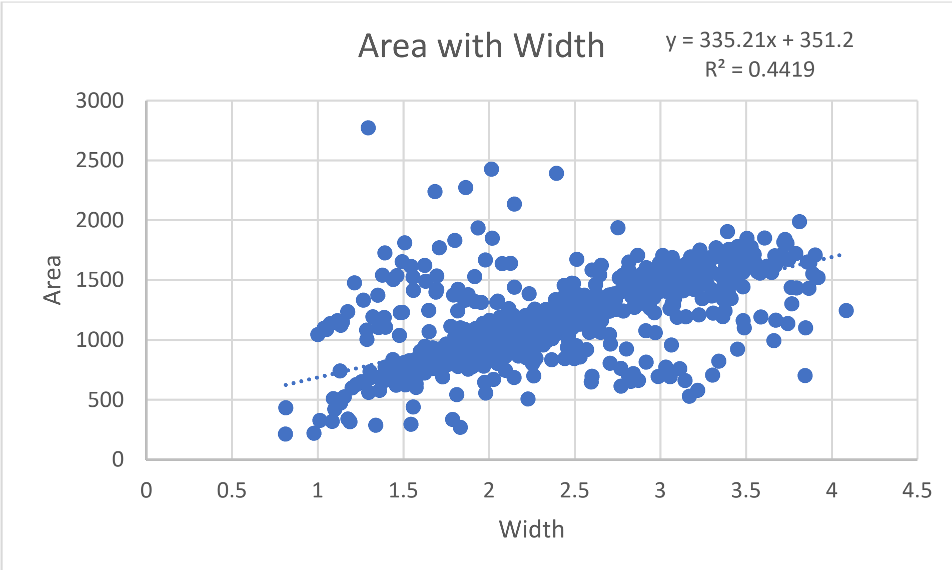

Figure 9. Segment area (m^2) with segment width (m).

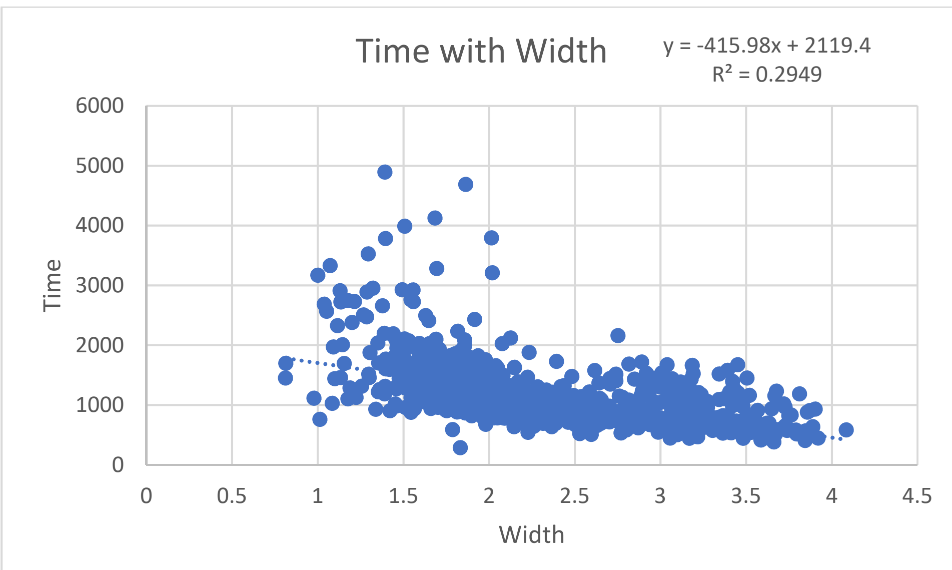

Figure 10. Segment width (m) with duration of time (seconds) the segment was sampled.

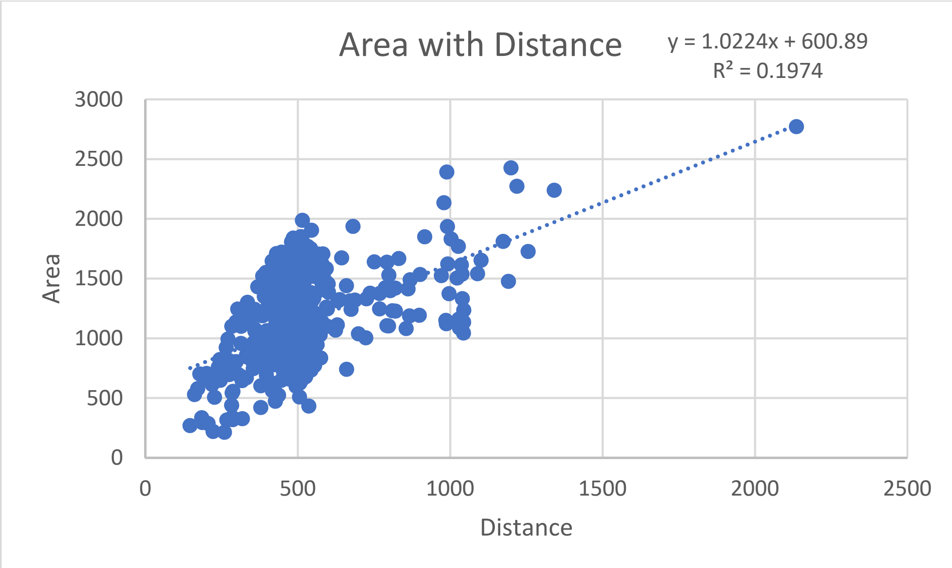

Figure 11. Segment area (m^2) with segment distance (m).

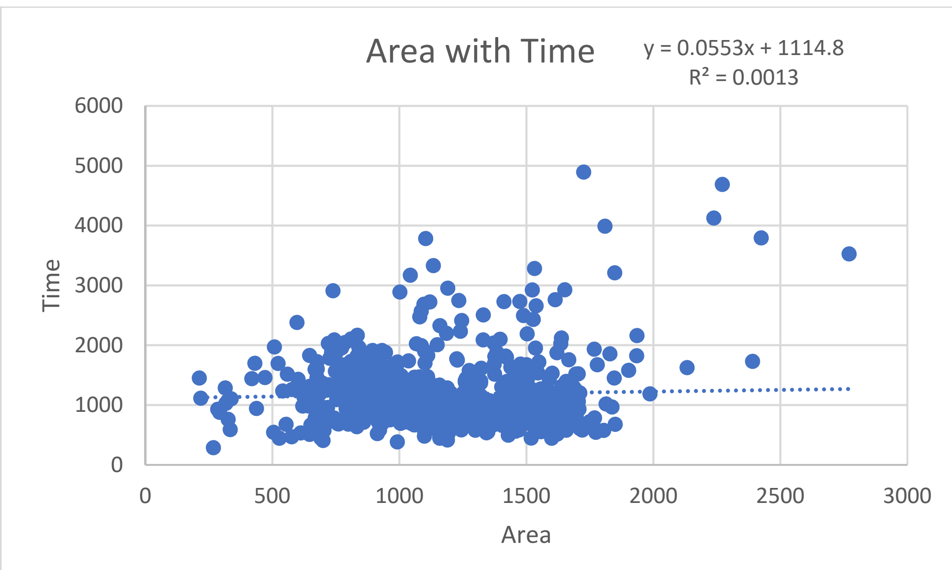

Figure 12. Duration of time (seconds) the segment was sampled with segment area (m^2).

GLM Using Variables Derived from the ROV and CSMP

Evaluation of the full suite of variables available from ROV as well as the terrain attributes derived from the CSMP, was conducted to evaluate correlations for all available variables. Results varied by species and results are provided in Tables 15 through 19. Across species the negative binomial distribution provided appreciably lower AIC scores. Significant overdispersion was observed for all species, thus use of the quasi-Poisson or negative binomial distributions was preferable to the Poisson. The zero-inflated Poisson model resulted in lower AIC values than the Poisson indicating zero-inflated models may be preferable or that reduced data sets excluding segments from depths greater than the distribution of the species would be beneficial. The binomial model had appreciably lower AIC values, though direct comparisons are not viable considering differences in the input data.

Results for correlation of density with variables available from the ROV survey were largely consistent with the previous analysis with the exception of depth for copper rockfish, which was no longer significant, while take was significant for the zero-inflated models when this was not the case previously. Correlations with terrain attributes were seldom consistent between distributions for a given species, though the depth range, RDMV, standard deviation of slope and surface to planar area were frequently significant. The proportion of rock bottom type was consistently highly significant across species and distributions. The lack of consistent significant correlations with terrain attributes across distributions within a species and between species as well as the lower level of significance may be in part attributable to the spatial error in the pairing of centroids of segments with CSMP raster grids and the aggregated scale at which terrain attributes were determined to address spatial error. There may be microscale habitat characteristics related to relief, which fish associate with, that the terrain attributes were unable to capture. An additional consideration is the sample size of presence for each species and the ability to identify associations. Consistency of correlation variables across space may provide an indication of whether apparent correlations are spurious.

The deviance measures for gopher rockfish were highest for latitude followed by proportion hard/mixed bottom type, then depth despite no longer being significant, while deviance was zero for take and the deviance for terrain attributes were substantially lower. Backward stepwise variable selection identified latitude, proportion hard or mix and depth, and several terrain attributes varying across distributions as resulting in the lowest AIC values.

These results indicate that variables from the ROV survey are more consistently significantly correlated to density, and terrain attributes correlated with density are inconsistent across species and distributions for a given species making identification of pertinent variables inconclusive. For a given model one should consider evaluation of the deviance measures and backward stepwise model variable selection to identify candidate variables, and more refined analysis of initially identified variables to identify candidate terrain attributes to represent relief in the habitat. Further evaluation of alternative scales of derivation for terrain attributes may allow causal aspects of relief that fish are responding to be better reflected in the respective metrics leading to more consistent results.

Table 15. Results of GLM with various distributions for gopher rockfish density with variables derived from the ROV and CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. The variables selected in backward stepwise regression are highlighted in yellow. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Binomial	Negative Binomial
Intercept	***	***	***	***	**	***
Latitude	*** (395)	***	*** (395)	***	*** (286)	*** (330)
Proportion Hard/Mix ROV	*** (119)	***	*** (119)	***	*** (88)	*** (107)
Depth	(79)		(79)		(62)	(73)
Take	(0.3)		(0.3)		(0.7)	(1)
DepthRange_3By3	(5)		(4.5)		(2.8)	(4)
DepthRange_5By5	(12)		(12)		(3.4)	(6)
DepthMean_3By3	(0.9)		(1)		(0.7)	(0.2)
DepthMean_5By5	(2.6)		(2.6)		(1.8)	(2.5)
RDMV_3By3	*(5)		* (4.9)		(5)	(3.1)
RDMV_5By5	(2)		(2)		(0)	(0.5)
Slope_3By3	(0.4)		(0.4)		(0.1)	(0.4)
STDofDepth_3By3	(1.1)		(1.1)		(2.8)	(1.2)
STDofDepth_5By5	(0.5)		(0.5)		(0.8)	(0.3)
STDofSlope_3By3	(3.5)		(3.4)		(2.1)	(2.3)
STDofSlope_5By5	(2.2)		(2.2)		(0.1)	(1.3)
SurfaceAreaTo PlanarArea_3By3	(3.4)	*	(3.4)	*	(1.5)	(2.1)
Proportion Hard CSMP	(2.7)	***	(2.7)	***	(1.8)	(1.6)
AIC	2748	2636	NA	2636	2256	2596
Dispersion	1.22 (0.031))					

Table 16. Results of GLM with various distributions for copper rockfish density with variables derived from the ROV and CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Binomial	Negative Binomial
Intercept		.		.		
Latitude	*	*	*	*	**	*
Proportion Hard/Mix ROV	***		***		***	***
Depth	.		.		.	.
Take		**		**		
DepthRange_3By3	.		.		.	.
DepthRange_5By5	.		.			.
DepthMean_3By3		*		*		.
DepthMean_5By5		.		.		
RDMV_3By3	*		*		*	*
RDMV_5By5	.		.			.
Slope_3By3		*		*		
STDofDepth_3By3	.	*	.	*	.	
STDofDepth_5By5						
STDofSlope_3By3						
STDofSlope_5By5	***		***		**	**
SurfaceAreaTo PlanarArea_3By3		*		*		
Proportion Hard CSMP	*	***	*	**		*
AIC	2019	1986	NA	1986	1913	2005
Dispersion	1.09 (0.09)					

Table 17. Results of GLM with various distributions for vermilion rockfish density with variables derived from the ROV and CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Binomial	Negative Binomial
Intercept	*					
Latitude	***	**	***	**	***	***
Proportion Hard/Mix ROV	***		***		***	***
Depth	***	***	***	***	**	***
Take	***	*	***	*	**	***
DepthRange_3By3	**	.		.		.
DepthRange_5By5		*		*		
DepthMean_3By3						
DepthMean_5By5	**	.		.		
RDMV_3By3	***		.			
RDMV_5By5	**					
Slope_3By3		***		***		
STDofDepth_3By3		**		**		
STDofDepth_5By5	**	.	.	.		
STDofSlope_3By3	*					
STDofSlope_5By5	***	***	***	***		***
SurfaceAreaTo PlanarArea_3By3	.					
Proportion Hard CSMP	***	***	**	***		
AIC	5604	4757	NA	4757	3525	4451
Dispersion	3.09 (0.037)					

Table 18. Results of GLM with various distributions for kelp greenling density with variables derived from the ROV and CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Binomial	Negative Binomial
Intercept	***		***		***	***
Latitude	***	***	***	***	***	***
Proportion Hard/Mix ROV	***	***	***	***	***	***
Depth	***	***	***	***	***	***
Take	***		***		**	***
DepthRange_3By3						
DepthRange_5By5	***	**	**	**	**	**
DepthMean_3By3						
DepthMean_5By5						
RDMV_3By3	*				*	.
RDMV_5By5	*	.	.	.	.	.
Slope_3By3		*	*	*		
STDofDepth_3By3		.		.		
STDofDepth_5By5	.					
STDofSlope_3By3					.	
STDofSlope_5By5						
SurfaceAreaTo PlanarArea_3By3	*		.		.	.
Proportion Hard CSMP	***		**		*	**
AIC	8118	7871	NA	7871	6541	8005
Dispersion	1.20 (6.9^-12)					

Table 19. Results of GLM with various distributions for China rockfish density with variables derived from the ROV and CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Binomial	Negative Binomial
Intercept		**		**		
Latitude	***	***	***	***	***	***
Proportion Hard/Mix ROV	**	*	*	*	*	*
Depth	**		**		**	**
Take	**	***	**	***	**	**
DepthRange_3By3	*		.		*	
DepthRange_5By5	.		.		*	.
DepthMean_3By3						
DepthMean_5By5						
RDMV_3By3						
RDMV_5By5						
Slope_3By3						
STDofDepth_3By3	**		*		.	.
STDofDepth_5By5					***	.
STDofSlope_3By3					*	
STDofSlope_5By5						
SurfaceAreaTo PlanarArea_3By3						
Proportion Hard CSMP	**		*		**	*
AIC	1033	988	NA	988	963	1023
Dispersion	1.07 (0.029)					

GLM Using Variables Derived from the CSMP

We tested for correlation of density with latitude, whether take was allowed in a segment, and terrain attributes derived for each segment from the CSMP, to identify variables for model-based expansion of density estimates using the CSMP data, providing estimates of abundance. Results are provided in Tables 20 through 24. For gopher rockfish and vermilion rockfish, the negative binomial distribution provided appreciably lower AIC scores, while the zero-inflated Poisson distribution resulted in lower AIC values for kelp greenling, China rockfish and copper rockfish. Significant overdispersion was observed for gopher rockfish, kelp greenling and China rockfish, thus use of the quasi-Poisson or negative binomial distributions was preferable to the Poisson. The zero-inflated Poisson model resulted in lower AIC values than the Poisson indicating zero-inflated models may be preferable or that reduced data sets excluding segments from depths greater than the distribution of the species would be beneficial. The binomial model had appreciably lower AIC values, though direct comparisons are not viable considering differences in the input data.

Latitude, proportion hard bottom from CSMP, and take were consistently significant variables across species and distributions. Whether take was allowed was significant for vermilion rockfish, kelp greenling and China rockfish, though this may be spurious or associated more with site selection than accrual of differences given the recent establishment of MPAs. Correlations with terrain attributes were seldom consistent between distributions for a given species, and none were significant across all species, though the depth range, RDMV, standard deviation of slope and surface to planar area were significant in some cases. The lack of consistent significant correlations with terrain attributes across distributions within a species and between species as well as the lower level of significance may be in part attributable to the spatial error in pairing of centroids of segments with CSMP raster grids and the aggregated scale at which terrain attributes were determined to address spatial error, as mentioned previously. There may be microscale habitat characteristics related to relief, which fish associate with that the terrain attributes were unable to capture. An additional consideration is the sample size of presence for each species and the ability to identify associations. Consistency of correlation variables across space may provide an indication of whether apparent correlations are spurious or the result of consistent associations.

The deviance measures for gopher rockfish was highest for latitude, followed by the mean depth on the three by three cell neighborhood scale, then proportion hard bottom type from CSMP. While the deviance for terrain attributes were substantially lower, the rugosity (surface to planar area), RDMV, and standard deviation of slope on a three by three cell neighborhood size provided the highest deviance. Backward stepwise variable selection identified latitude, mean depth, proportion hard bottom and a number of terrain attributes varying across distributions highlighted in yellow as resulting in the lowest AIC values. For a given model, one should consider evaluation of the deviance measures and backward stepwise model variable selection to identify candidate variables and more refined analysis of initially identified variables to identify candidate terrain attributes to represent relief in the habitat.

Table 20. Results of GLM with various distributions for gopher rockfish density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. The variables selected in backward stepwise regression are highlighted in yellow. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***	***	***	***	***	***
Latitude	*** (316)	***	*** (316)	***	*** (204)	***(222)
Take	(3.2)		(3.2)		(0.3)	(3.1)
DepthRange_3By3	(25)		(25)		(1.2)	(17.3)
DepthRange_5By5	(8)		(8)		(0.1)	(1.6)
DepthMean_3By3	(133)		(133)		(0.3)	(101.6)
DepthMean_5By5	(1)		(1)		(0.4)	(0.45)
RDMV_3By3	* (3.8)		* (3.8)		. (2.5)	(4)
RDMV_5By5	. (0.8)		(0.8)		(0.9)	(0.5)
Slope_3By3	(0.1)		(0.1)		(0)	(0.01)
STDofDepth_3By3	(3.8)		(3.8)		(0.7)	(3.7)
STDofDepth_5By5	(1.2)		(1.2)		(0.02)	(0.001)
STDofSlope_3By3	(3.8)		(3.8)		(0.2)	(2.6)
STDofSlope_5By5	(1.2)		(1.2)		(0.6)	(0.003)
SurfaceAreaToPlanarArea_3By3	. (8.5)	.	(8.5)	.	(2.1)	(5.1)
Proportion Hard CSMP	*** (30.5)	***	*** (30.5)	***	***(20)	*** (23.5)
AIC	2838	2717	NA	2717	2672	2326
Overdispersion	1.29 (0.021)					

Table 21. Results of GLM with various distributions for copper rockfish density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept		***		***		
Latitude	*	***	*	***	*	**
Take						
DepthRange_3By3	.	.	.	.	.	.
DepthRange_5By5						
DepthMean_3By3		*		*		
DepthMean_5By5		*		*		
RDMV_3By3	*	.	*	.	*	*
RDMV_5By5	*	*	.	*	*	
Slope_3By3		*		*		
STDofDepth_3By3	.	*	.	*	.	.
STDofDepth_5By5						
STDofSlope_3By3						
STDofSlope_5By5	**		**		**	**
SurfaceAreaTo PlanarArea_3By3		.		.		
Proportion Hard CSMP						
AIC	2075	2014	NA	2014	2058	1964
Overdispersion	1.11 (0.095)					

Table 22. Results of GLM with various distributions for vermilion rockfish density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	**		.		.	.
Latitude	***	***	***	***	***	***
Take	***		***		***	**
DepthRange_3By3	**	.		.		
DepthRange_5By5		***		***		
DepthMean_3By3	.					
DepthMean_5By5	.					
RDMV_3By3	***		.			
RDMV_5By5	**					
Slope_3By3		***		***		
STDofDepth_3By3		**		**		
STDofDepth_5By5	***		*			
STDofSlope_3By3	**					
STDofSlope_5By5	***	***	***	***	***	
SurfaceAreaTo PlanarArea_3By3	*					
Proportion Hard CSMP		***		***	*	***
AIC	3.24					
Overdispersion	5955 (0.12)	4904	NA	4904	4529	3648

Table 23. Results of GLM with various distributions for kelp greenling density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept	***		***		***	***
Lat	***	**	***	**	***	***
Take	***		***	.	***	**
DepthRange_3By3						
DepthRange_5By5	***	**	**	**	**	**
DepthMean_3By3						
DepthMean_5By5						
RDMV_3By3	.		.			*
RDMV_5By5	*		.			.
Slope_3By3		.		.		
STDofDepth_3By3						
STDofDepth_5By5	.	.	.	.	.	
STDofSlope_3By3	.					.
STDofSlope_5By5						
SurfaceAreaTo PlanarArea_3By3	.		.			
Proportion Hard CSMP	***		***	.	***	***
AIC	1.21					
Overdispersion	8164 (3.77e-13)	7919	NA	7919	8045	6580

Table 24. Results of GLM with various distributions for China rockfish density with variables derived from the CSMP. The *** is significant at the 0.001 level, ** is significant at the 0.01 level and * is significant at the 0.05 level, . <0.1 is nearly significant. Values in brackets are the deviance for the variable in question. AIC values are indicated for each model and overdispersion test result provided. Estimates of overdispersion (>1 indicating overdispersion) and significance of overdispersion are indicated at the bottom of the table.

Variables / Factors	Poisson	Zero-Inflated Poisson	Quasi-Poisson	Zero-Inflated Quasi-Poisson	Negative Binomial	Binomial
Intercept		**		**		
Latitude	***	***	***	***	***	***
Take	**	***	**	***	**	*
DepthRange_3By3		*		*		.
DepthRange_5By5	*	*	.	*	*	*
DepthMean_3By3						
DepthMean_5By5						
RDMV_3By3						
RDMV_5By5						
Slope_3By3						**
STDofDepth_3By3	*	*	*	*	.	*
STDofDepth_5By5					.	
STDofSlope_3By3						
STDofSlope_5By5						
SurfaceAreaTo PlanarArea_3By3						
Proportion Hard CSMP	***	***	***	***	***	***
AIC	1052	1005	NA	1005	1041	982
Overdispersion	1.09 (0.039)					

Expansion of Density to Estimate Abundance

Expansion of Estimates using Habitat Mapping

Both design-based and model-based methods of expanding density estimates to estimates of abundance using CSMP data were explored using gopher rockfish in Central California as a test case. A range of methods could be employed each with their own data requirements, pros/cons, complexity and assumptions described in the overview of the five methods outlined below. We explored method two as an example of a design-based approach and method five which is a more complex model-based method for review. A generalized description of the design-based and model-based methods we pursued that fall into these respective methods is provided for perspective.

The poststratification applied design-based method was informed by significant correlations between variables and density identified using GLM. Distributions of average density across significant variables were examined to inform poststratification of estimates. Then the density estimate, and area estimate for each stratum were calculated and the sum of the product across all strata provided an estimate of abundance.

For the model-based method, CSMP data was paired based on location to the centroid of each segment of the ROV survey and terrain attributes associated with each segment were estimated to provide variables to be evaluated for correlation to density using GAMs. Coefficients for the variables found to be significantly correlated in the GAM were then derived. The Marine Geospatial Ecology Tool was then employed to estimate density of fish in each raster grid cell of the CSMP using correlations derived by the GAM. The densities for each grid cell were then multiplied by the area of each grid cell and summed to provide an estimate the abundance across the mapped area consistent with the methods of Young and Carr (2013). A more detailed description of the design-based and model-based methods and the results for the test case application to gopher rockfish are provided in the following section.

Method 1. Design-based estimate with statewide densities and hard bottom habitat area.
Requirements and Data Needs:

- Rates: Currently available as average rates of fish per meter squared. Density estimates from bootstrap methods provide mean density and variance estimates.
- Hard Bottom Area: Estimates of the square meters of hard bottom habitat from CSMP.

Pros: Easy to produce. Can be estimated without pairing ROV observations to the CSMP data to derive terrain attributes.
Cons: Does not account for variation in abundance by area or depth. Assumes fish are not encountered in areas not classified as hard bottom habitat.

Method 2. Design-based estimate accounting for differences in densities by depth and latitude.
Requirements and Data Needs:

- Density Estimates: Currently available as average rates of fish per meter squared. Density estimates from bootstrap methods provide mean density and variance estimates.
- Area Estimates: Estimates of the square meters of hard bottom habitat from CSMP by depth and latitude.

Pros: Relatively easy to produce. Addresses basic trends in abundance by depth and latitude. Can be estimated without pairing ROV observations with the CSMP data to derive terrain attributes. May be more accurate and reduce variance in estimates compared to option 1.
Cons: Assumes no fish are encountered in areas not classified as hard bottom habitat. Does not account for differences in density with relief captured in terrain attributes from the CSMP.

Method 3. Model-based estimate with coefficients from GLM using Latitude and Depth data available from the ROV alone implemented in R or MGET.
Requirements and Data Needs:

- Density Estimates: GLM coefficients for models fitted to depth, latitude and rock present.
- Area Estimates: Latitude, depth, hard/soft classification and area for each raster cell from the CSMP data.

Pros: Addresses basic trends in abundance by depth and latitude at a finer scale than option 1. Can be estimated without pairing ROV observations to the CSMP data to derive terrain attributes. May be more accurate and reduce variance in estimates compared to option 1 and 2.
Cons: More computationally intensive than the design-based estimates. Assumes no fish are encountered in areas not classified as hard bottom habitat. Does not account for differences in density with relief captured in terrain attributes from the CSMP.

Method 4. Model-based estimate from GLM using data available from ROV observations paired with CSMP with derived terrain attributes implemented in MGET.
Requirements and Data Needs:

- Density Estimates: Coefficients from GLM for proportion rock, depth, latitude, take or terrain attributes for each model.
- Area Estimates: Latitude, depth, proportion rock, take and terrain attributes for each raster cell from the CSMP data.

Pros: Addresses trends in density with bottom relief as well as depth and latitude as a result of pairing to the CSMP data to derive terrain attributes. May account for attributes associated with abundance increasing accuracy of estimates of abundance. Estimation based on proportion hard bottom accounting for abundance on soft and hard substrate.
Cons: More computationally intensive than the design-based estimates. May be subject to spurious correlation with terrain attributes. Cannot account for non-linear relationships between density and variables.

Method 5. Model-based estimate from GAM using data available from ROV paired to CSMP with derived terrain attributes implemented in MGET.
Requirements and Data Needs:

- Density Estimates: Coefficients from GAM for proportion rock, depth, latitude, take or terrain attributes for each model.
- Area Estimates: Latitude, depth, proportion rock, take and terrain attributes for each raster cell from the CSMP data.

Pros: Addresses trends in density with bottom relief as well as depth and latitude as a result of pairing to the CSMP data to derive terrain attributes. May account for attributes associated with abundance increasing accuracy of estimates of abundance. Estimation based on proportion hard bottom accounting for abundance in soft and hard substrate. Better able to capture non-linear relationships than the GLM i.e. increase then decline in density of China rockfish with depth.
Cons: More computationally intensive than the design-based estimates. May be subject to spurious correlation with terrain attributes.

Are estimated size compositions consistent with input data and population biological characteristics? (Unaddressed Review Performance Topic)
Estimation of average weights and application to expansions to estimate biomass

Comparison of lengths available from each method
The difficulties presented by the orientation of the fish and costs associated with processing the samples limit the availability of precise stereo camera-based estimates of lengths of fish encountered by the ROV as discussed previously. Given this limitation there is a need for data from other methods to provide enough length data to allow for stratification to address variability over the strata. An analysis conducted by Kline et al. (2014) indicated that the visual approximation of lengths using paired lasers as a reference systematically underestimate length of observed fish relative to the stereo camera estimates. This bias was proportional to the size of the fish observed and a correction can be made using the results of Kline et al. (2014) where results were available or analogous methods applied to newly derived data. Our preliminary analysis used length data from the recreational fishery for application in converting the number of fish to provide an estimate biomass since length data from the ROV survey was unavailable at

the time. To derive average weight estimates, lengths would be converted to weights using length-age conversions developed from previous research, stock assessments or for use in approximating average weights in sampling programs.

To evaluate the suitability of length compositions from the recreational fishery as a proxy for length compositions from the ROV survey we plotted length frequency distributions of species of interest from the ROV and the retained catch from boat-based modes of the CRFS sampling program from the Oregon border to Point Conception for comparison. The length compositions were very similar (Figures F through J), with the distribution from CRFS biased slightly lower for all species except gopher rockfish for which the distribution was higher for the CRFS lengths.

In addition, we estimated the average weight from each data source and estimated the percent difference in weight between data sources. The percent difference between average weights from each data source was biased high for gopher rockfish by 21% as compared to a negative bias of less than 11.3% for the remaining species (Table F). Further analysis of the comparison of distributions for bias adjusted lengths of paired laser length estimates from the ROV to account for the overestimation of length with paired laser approximation may result in reduced percent difference relative to CRFS lengths for all but gopher rockfish. The sample size for the CRFS sampling was far greater than that for paired laser estimates from the ROV survey (Table F) and points to the potential benefit of being able to use these lengths as proxy estimates of average weight. The results suggest that the CRFS lengths provide a reasonable proxy estimate of average weight providing a sufficient sample size to allow stratification.

Application of average weights in estimates of biomass estimates

The sampling scheme employed was not random with respect to depth or latitude and as a result, variation in average weight may not be adequately represented in the sample data from the ROV or proxy estimates from the recreational fishery. Should significant differences in average weight exist among strata used in design-based estimates or correlations with variables in model-based methods, these relationships may need to be addressed to provide unbiased estimates. Tests for significant differences in the average weight of gopher rockfish among depths and latitudes or other pertinent strata can be conducted using an ANOVA or non-parametric Kolmogorov Smirnov tests and appropriate stratification implemented before applying the average weights For model-based estimates,   GLM or GAM can be used to test for significant correlations of weight with depth and latitude or other variables and the relationship used to apply a representative average weight in converting estimates of the number of fish to weights.

We examined the average weights from paired laser approximations of lengths from the ROV survey for gopher rockfish across depth and latitude (Figure K.). While some variability was observed, it occurred in strata with low sample sizes, otherwise the average weight was consistent across depths and latitudes. As a result, no efforts were made to address the average weights employed in our previous analysis. This is a first step that should be taken to examine trends and the ANOVA or GLM used to rigorously evaluate significant differences between strata or correlations to be addressed in converting numbers of fish to biomass.

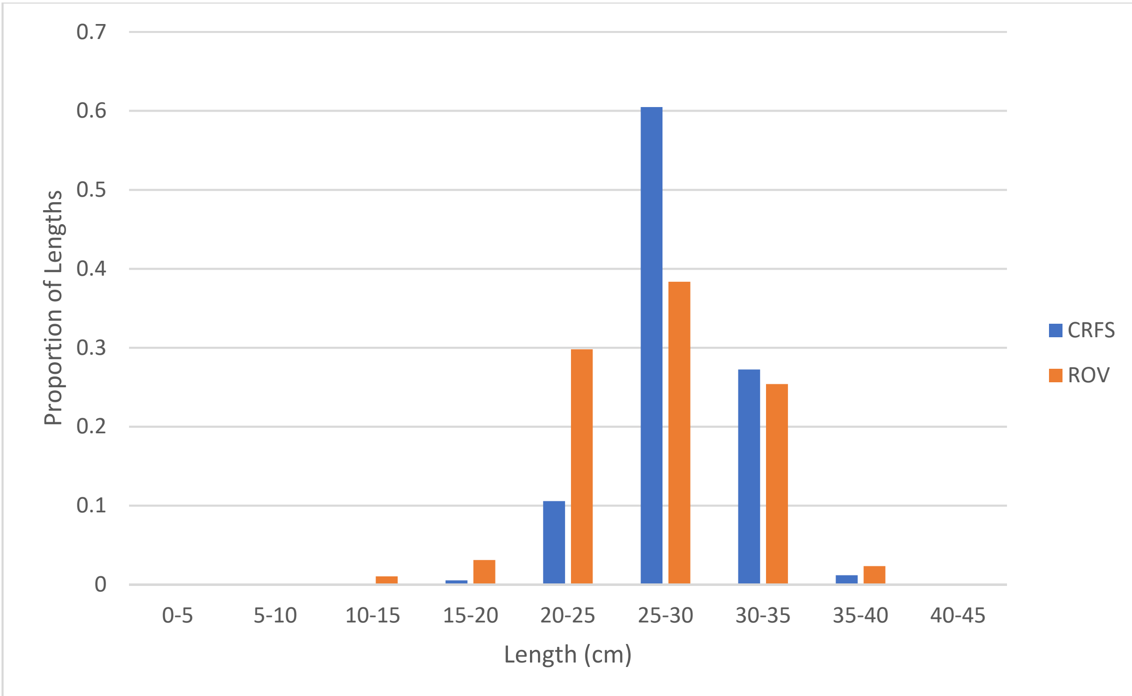

Figure F. Length composition of gopher rockfish from the CRFS sampling and the ROV observations approximated with paired lasers.

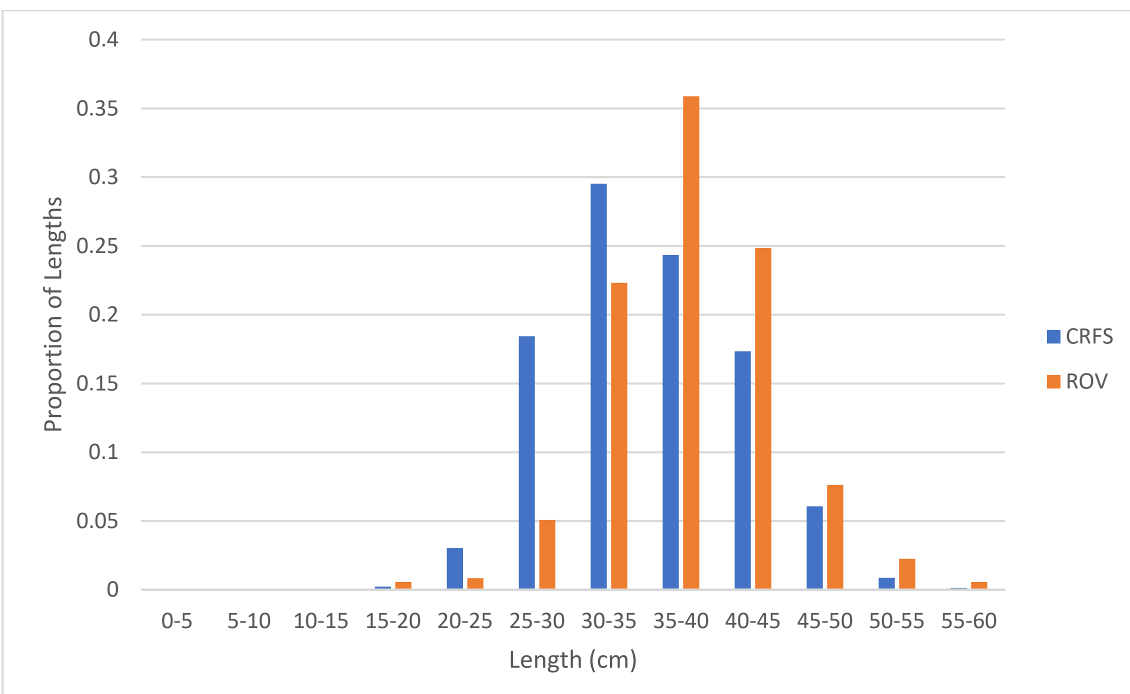

Figure G. Length composition of copper rockfish from the CRFS sampling and the ROV observations approximated with paired lasers.

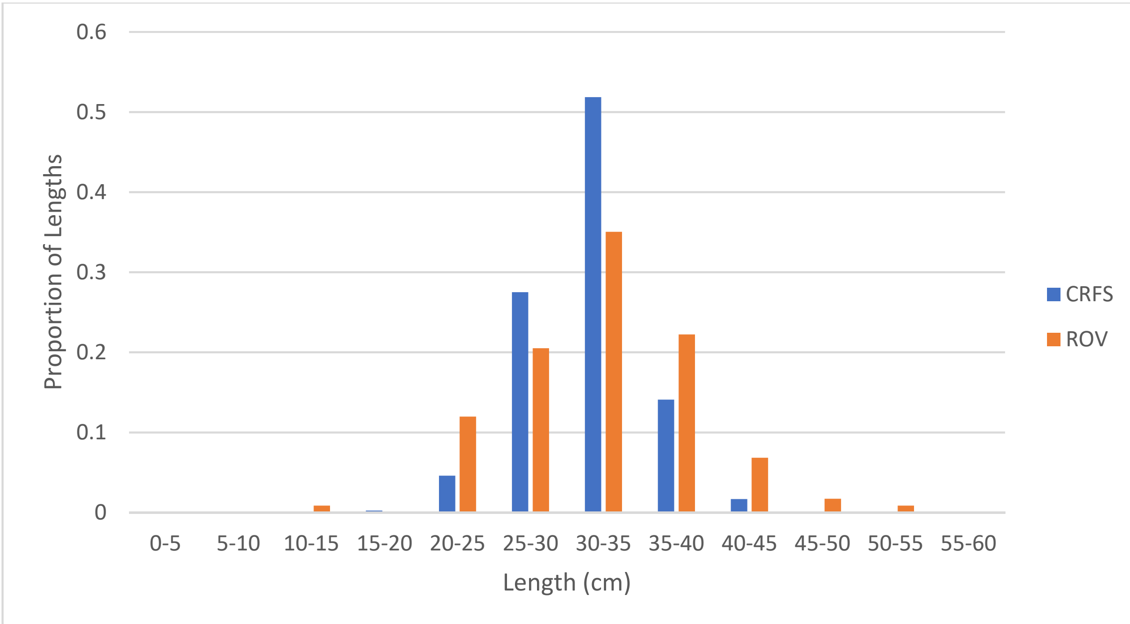

Figure H. Length composition of China rockfish from the CRFS sampling and the ROV observations approximated with paired lasers.

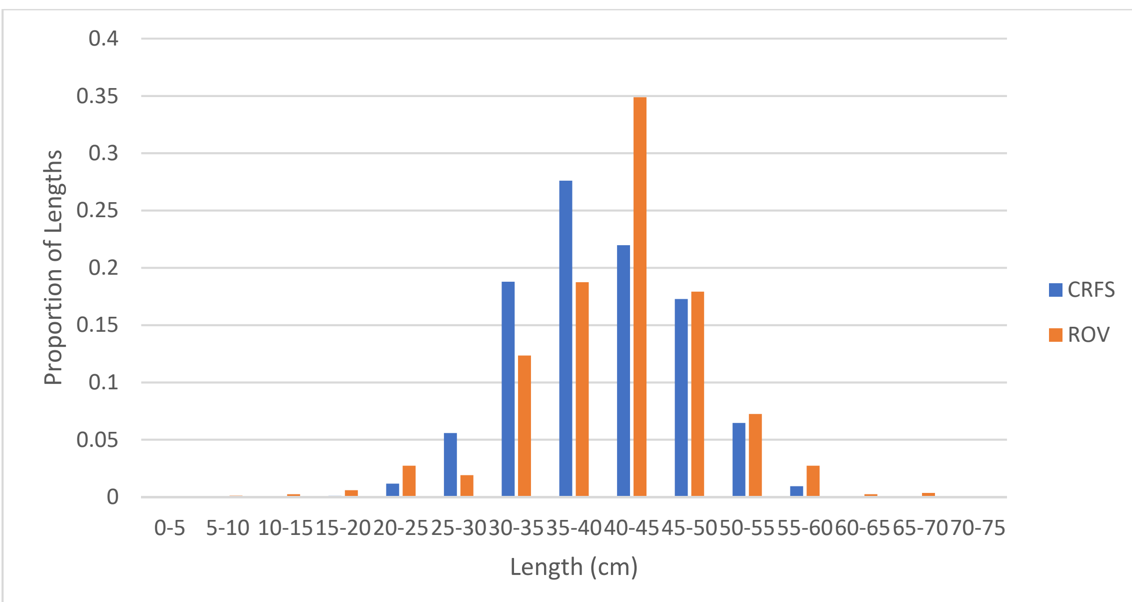

Figure I. Length composition of vermilion rockfish from the CRFS sampling and the ROV observations approximated with paired lasers.

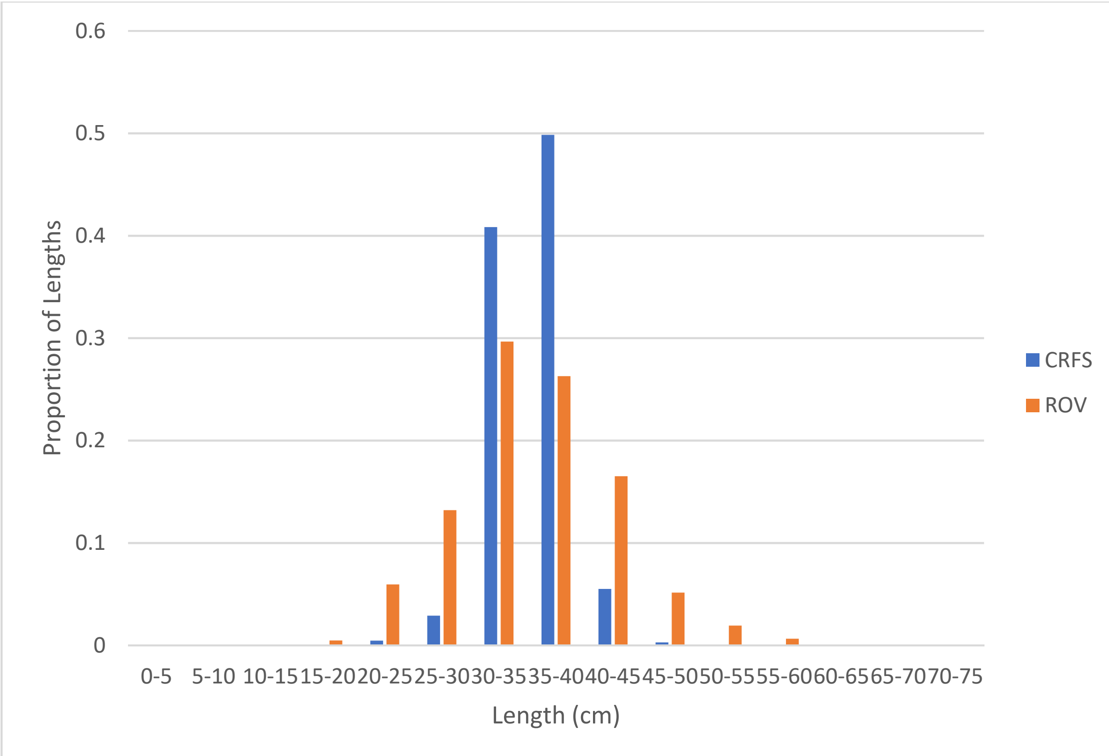

Figure J. Length composition of kelp greenling from the CRFS sampling and the ROV observations approximated with paired lasers.

Table F. Number of lengths, average weight in kilograms and percent difference between CRFS and ROV estimates of average weight in kilograms for species of interest.

Species	Number of Lengths CRFS	Number of Lengths ROV	Average Weight CRFS (kg)	Average Weight ROV (kg)	Percent Difference CRFS vs. ROV
Gopher Rockfish	19423	386	0.377	0.311	21.2%
Copper Rockfish	9451	354	1.499	1.690	-11.3%
China Rockfish	3635	117	0.657	0.697	-5.8%
Vermilion Rockfish	14301	843	1.108	1.165	-5.0%
Kelp Greenling	3031	1864	0.654	0.659	-0.7%

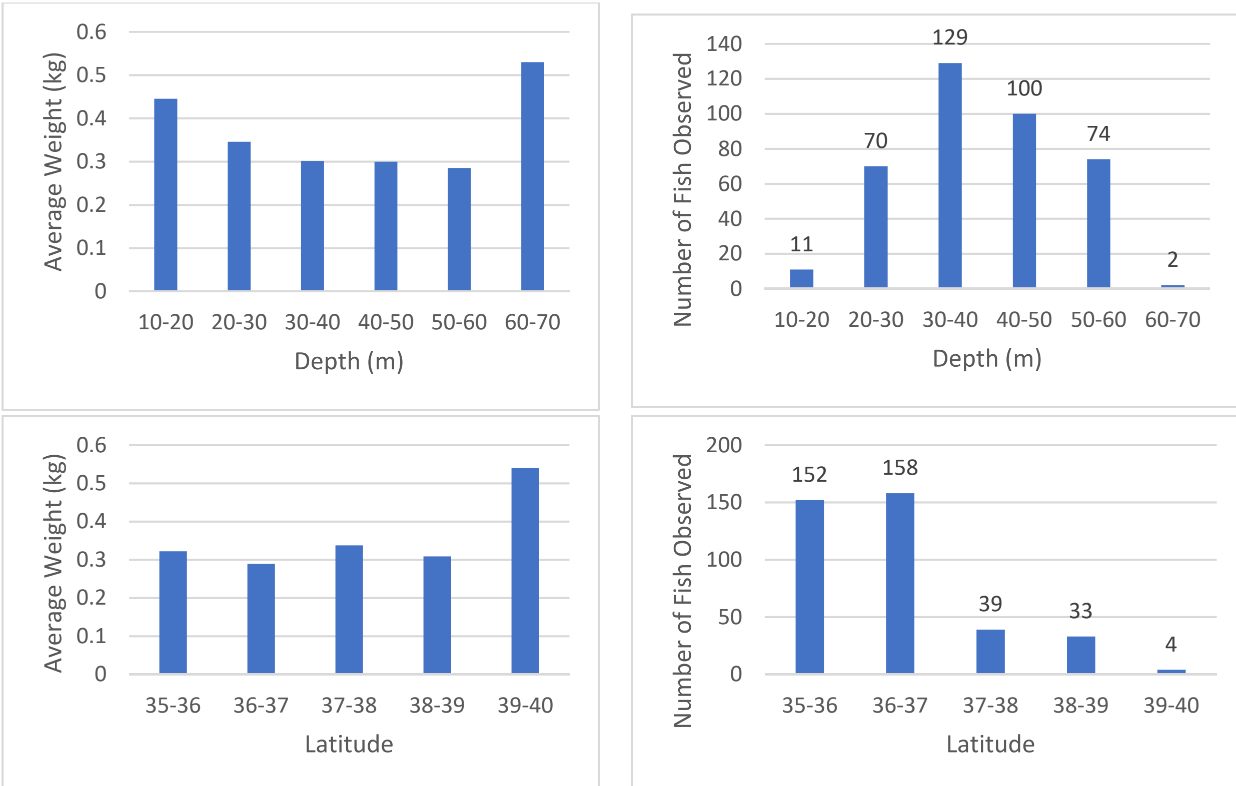

Figure K. The average weight of gopher rockfish and number of fish observed with latitude and depth.

Detailed Description of the Design-based Expansion Methods

Methods: A step by step description of the design-based expansion method is provided below.

1. A GLM was conducted with gopher rockfish density data for the 20 m segments to examine correlations with the associated variables including depth, latitude and proportion hard reef from the ROV study as well as whether take was allowed in the segment. The results provided in Table 20 above indicated that all but take were significant.

2. Distributions of average density in fish/squared meter across observed segments by depth (Figure 13), latitude (Figure 14) and proportion hard substrate from the ROV observations (Figure 15) were examined to identify differences amongst bins indicative of the need to poststratify density estimates across bins.

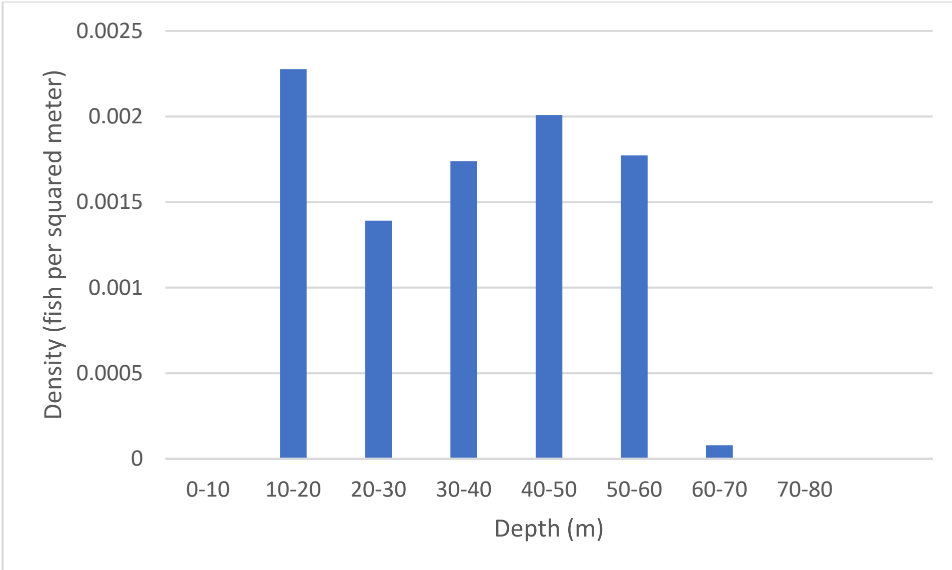

Figure 13. Density of gopher rockfish with depth.

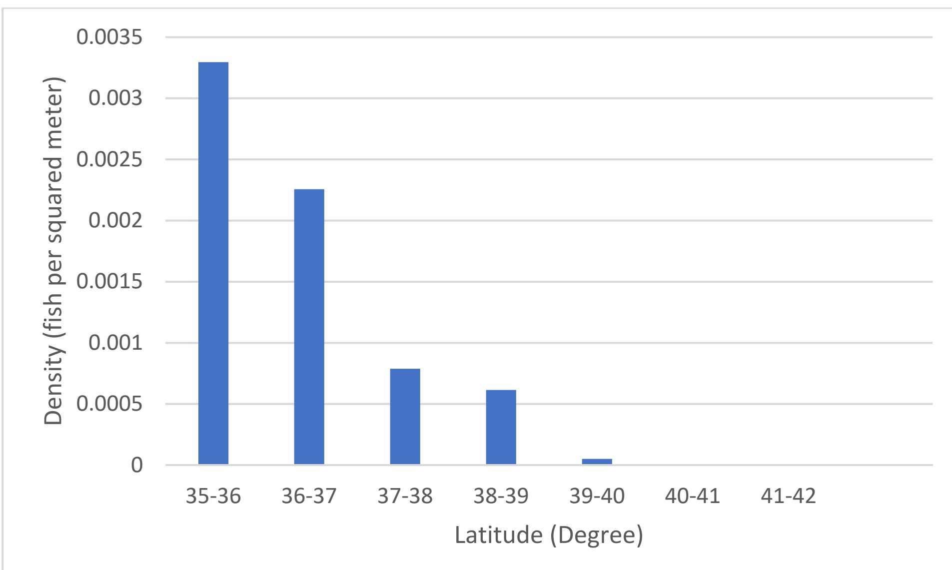

Figure 14. Density of gopher rockfish with whole degree latitude.

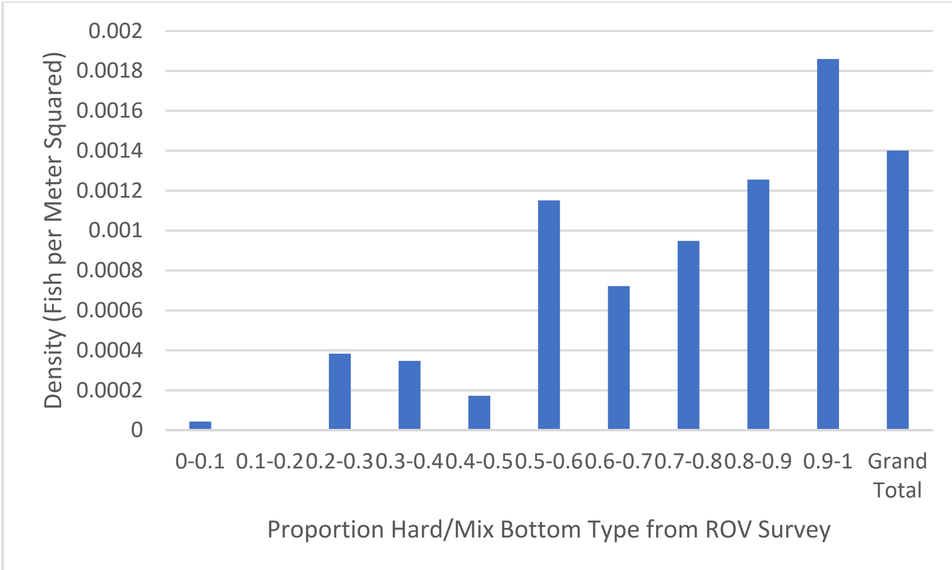

Figure 15. Density of gopher rockfish with proportion hard habitat from CSMP.

3. To determine the habitat area to which estimates should be expanded, we examined the density of gopher rockfish with the proportion of hard bottom from the CSMP in Figure 15. While the proportion hard or mixed habitat along the ROV transect observations was significant in the GLM results in Table 5 above, the percent hard bottom from the CSMP in Table 20 was not, though a slight trend can be observed below. The CSMP data forms the basis for the habitat area estimate in the expansion, thus we are limited to its use to inform stratification. Two alternatives were considered to bracket the uncertainty presented by the habitat classification informing expansions. The first was to include only those raster cells classified as hard bottom in the CSMP, which is expected to provide an estimate that is biased low given the presence of gopher rockfish in areas with less than 100% hard bottom in the ROV based observations. To bracket the high end of the uncertainty, we used a threshold method based on the proportion of raster grid cells within a 30 by 30 m neighborhood that were classified as hard bottom habitat with a threshold of 0.10 at which the density approached zero in Figure 16.

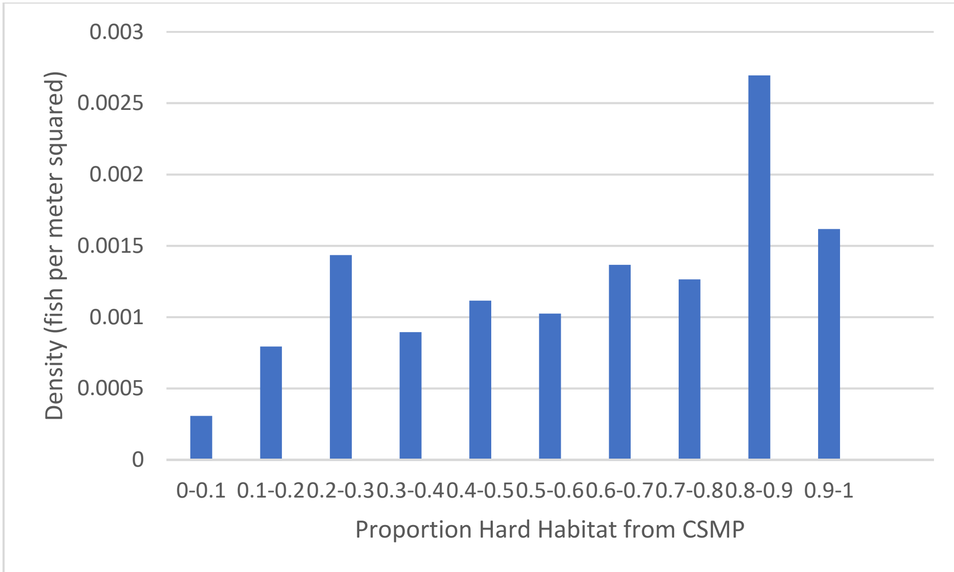

Figure 16. Density of gopher rockfish with proportion hard bottom habitat from CSMP.

4. While depth was significant in the GLM, the distribution of density with depth in Figure 13 did not reflect apparent trends at a 10 m resolution. Thus, we analyzed two levels of stratification to evaluate the effect of stratifying by depth. The data set was reduced to only segments from the 10-60 m depth range were retained as a result of the relative absence of fish in shallower and deeper depths as not to bias low estimates of density in the stratification aggregating depths. The alternative method of stratifying by depth was conducted to allow comparison of estimates.

5. The product of density and area estimates for each stratum were summed across all strata to provide an estimate of abundance in numbers of fish. The numbers of fish was multiplied by the estimate of average weight in kilograms for retained and discarded fish from the California Recreational Fishery Survey and divided by 1000 to provide an estimate of abundance in metric tons.

6. Variance estimates for density were derived for each stratum of postratification applied and criterion for inclusion of habitat in expansions. The habitat estimates for the respective strata were assumed known with 100% certainty based on the resolution of the CSMP seafloor mapping. The product of the variance estimate and the respective area for a given stratum from the seafloor mapping were used to provide an estimate of variance in the estimated number of fish, which was summed across all strata to provide an aggregate estimate of variance. The result was multiplied by a proxy average weight for gopher rockfish of 0.48 kg/fish from the retained and discarded fish for the region from the California Recreational Fishery Survey.

The variance estimates for density are hard to follow. Please provide an equation for these calculations. Recall that if scaling a variance estimate, the equation is Var(aX) = a^2 * Var(X). If I interpret what is being done correctly, I think you are not squaring the scaler. Also recall that when summing across aggregate estimates of variance you are assuming independence (covariance among strata = 0). (Dr. Berger)

The variance equation used in excel is the sample variance provided below estimated for the density in each of the latitude and depth strata. The variance was then multiplied by the area in the respective strata to scale it to the absolute number of fish it represents, then the resulting variance in total fish was summed to provide the total variance. This assumes the area was known without error. The variance in number of fish was multiplied by the average weight to provide an estimate variance in kilograms, which was converted to metric tons. Since the value variance in density was converted to absolute number of fish using the area of the respective stratum, the estimate accounts for the relative area contributing to the total and applying squared weights as suggested is not necessary. These estimates do not account for potential spatial autocorrelation between segments discussed further below and thus assume independence and the covariance among strata is equal to zero and likely underestimate variance as a result.

$$\text{Variance} = \frac{\sum (x-\bar{x})^2}{(n-1)}$$

Results: A summary comparing the resulting estimates for each stratification by depth and habitat area criterion is provided in Table 25. Tables reflecting the values used in the calculations for stratification by 10 m depth bins vs. aggregated estimates in the 10-60 m depth bin with alternative criteria for habitat included are provided in Tables 26 through 29.

The design-based estimation method accounting for depth using hard bottom from the CSMP as the basis for expansion provided an estimated abundance of 578,621 fish or 281.0 mt assuming an average weight of 0.48 kg/fish with a variance of 42,776 fish or 20.8 mt. The design-based estimation method combining depths from 10-60 m using hard bottom from the CSMP as the basis for expansion provided an estimated abundance of 549,995 fish or 266.9 mt assuming an average weight of 0.48 kg/fish with a variance of 46,939 fish or 22.8 mt.

The design-based estimation method accounting for depth using 10% hard bottom in a 30 by 30 m neighborhood from the CSMP as the basis for expansion provided an estimated abundance of 957,227 fish or 464.8 mt assuming an average weight of 0.48 kg/fish with a variance of 75,749 fish or 36.8 mt. The design based estimation method combining depths from 10-60 m using 10% hard bottom in a 30 by 30 m neighborhood from the CSMP as the basis for expansion provided an estimated abundance of 915,161 fish or 444.4 mt assuming an average weight of 0.48 kg/fish with a variance of 80,162 fish or 38.9 mt.

Table 25. Summary of design-based abundance estimates in numbers of fish and metric tons from alternative stratification of depth and habitat selection criteria.

Habitat Criteria for Area	Depth Resolution	Fish (#)	Fish (mt)
CSMP Hard Only	Combined 10-60 m	549995	266.9
CSMP Hard Only	Stratified by 10 m bins	578621	281.0
>10% Hard Threshold	Combined 10-60 m	915161	444.4
>10% Hard Threshold	Stratified by 10 m bins	957227	464.8

Table 26. Estimates of abundance based on stratification with combined depth for 10-60 m and hard substrate only from the CSMP included in the area estimate.

	Area (m Sq.)		Density (Fish/m Sq.)			
	Latitude (Deg.)		Latitude (Deg.)			
Depth (m)	35-36	36-37	35-36	36-37	Fish (#)	Fish (mt)
0-10	9762613	4765384	NA	NA	NA	NA
10-60	89978743	49427739	0.0045	0.0029	549027	266.6
60-70	7448067	6170170	0.0000	0.0002	967	0.2
				Total	549995	266.9

Table 27. Estimates of abundance based on stratification by depth and hard substrate only from the CSMP included in the area estimate.

	Area (m Sq.)		Density (Fish/m Sq.)			
	Latitude (Deg.)		Latitude (Deg.)			
Depth (m)	35-36	36-37	35-36	36-37	Fish (#)	Fish (mt)
0-10	9762613	4765384	NA	NA	NA	NA
10-20	25775809	14432924	0.0000	0.0121	174016	84.5
20-30	24269648	11737685	0.0033	0.0026	110091	53.5
30-40	16585925	8874023	0.0080	0.0017	147752	71.8
40-50	11843589	6942529	0.0049	0.0033	80987	39.3
50-60	11503772	7440577	0.0023	0.0051	64808	31.5
60-70	7448067	6170170	0.0000	0.0002	967	0.5
70-80	5555413	6701693	0.0000	0.0000	0	0.0
80-90	1323569	6190817			0	0.0
				Total	578621	281.0

Table 28. Estimates of abundance based on stratification with combined depth for 10-60 m and 10% threshold for hard substrate from CSMP.

	Area (m Sq.)		Density (Fish/m Sq.)			
	Latitude (Deg.)		Latitude (Deg.)			
Depth (m)	35-36	36-37	35-36	36-37	Fish (#)	Fish (mt)
0-10	12138070	6917543	NA	NA	NA	NA
10-60	157151169	70802864	0.0045	0.0029	914551	444.1
60-70	14898466	3890020	0.0000	0.0002	610	0.3
				Total	915161	444.4

Table 29. Estimates of abundance based on stratification by depth and 10% threshold for hard substrate from CSMP.

	Area (m Sq.)		Density (Fish/m Sq.)			
	Latitude (Deg.)		Latitude (Deg.)			
Depth (m)	35-36	36-37	35-36	36-37	Fish (#)	Fish (mt)
0-10	12138070	6917543	NA	NA	NA	NA
10-20	40058156	22807898	0.0000	0.0121	274993	133.5
20-30	41221462	19258477	0.0033	0.0026	185217	89.9
30-40	31679358	14181320	0.0080	0.0017	277455	134.7
40-50	22506449	8972328	0.0049	0.0033	139828	67.9
50-60	21685743	5582841	0.0023	0.0051	79124	38.4
60-70	14898466	3890020	0.0000	0.0002	610	0.3
70-80	11993427	4047605	0.0000	0.0000	0	0.0
80-90	2985162	4314689			0	0.0
				Total	957227	464.8

Conclusion: The results of the design-based approach point out a few limitations. The lack of sampling in the 0-10 m depth bin would require a proxy value to be used from the most proximate depth bin or for gopher rockfish to be assumed to be absent from depths less than 10 m. We know that the abundance is not zero from fishery data and descriptions of the distribution of this species (Love et al 2002), but assuming the density is the same as the 10-20 m depth bin may not be a valid assumption either. Also problematic is the lack of CSMP seafloor mapping in portions of the shallow depths in the 0 - 10 m depth range despite the presence of gopher rockfish and other nearshore species that would go unaccounted for. CDFW GIS analysts have developed proxy information to fill in absent depth information. Dive surveys have been conducted by PISCO and other organizations, which could be used to provide fish density data for these shallower depths and equivalent analyses conducted as in Young and Carr (2015) for portions of the coast in which proxy habitat data is available.

The zero density estimate for the 10 - 20 m depth bin in the 10 m stratification for the latitude of 35-36 degrees points out the potential for low sample size of observations in some strata at higher resolution of stratification. Increased sampling effort or decreased stratification may resolve such issues and a more optimal design may be preferable balancing the need to capture sources of variation and sample size.

As expected, inclusion of only two by two-meter resolution rasters classified as hard bottom habitat in the CSMP results in lower estimates of abundance than when using a 10% threshold for proportion hard bottom habitat in a 30 by 30 m habitat neighborhood size surrounding a given raster grid. Conversely, inclusion of only habitat classified as rock provides a biased low estimate of viable habitat compared to the sampled seafloor over the course of the transects used to derive the density estimates, which included both soft and hard bottom habitat. A higher threshold than 10% may be preferable to exclude habitat that may be isolated from suitable reef structure, though the example bears out the level of potential bias and uncertainty from lower threshold values based on CSMP data. The proportion hard bottom habitat from the transect observations would suggest a threshold of 20% or more may be more appropriate.

Future refinements should also include estimates of density and variance generated using bootstrap analyses of randomly selected segments of transects in each stratum. Use of ANOVA or other means of identifying strata in a more rigorous fashion would also be preferable in future efforts. Coefficients of variation (CVs) can be compared between alternative stratification schemes to best account for sample variability. Unfortunately, time constrained our ability to pursue further refinements prior to the review.

Detailed Description of the Model-based Expansion Method using MGET

Methods: MGET allows for application of GLMs and GAMs in projecting abundance. GAM models have the advantage of being able to use flexible splines to fit non-linear relationships. The GAM function also offered the negative binomial distribution model in MGET, which the GLMs run in R identified as providing the lowest AIC. This was in part due to the ability of the negative binomial distribution to address overdispersion through the variable theta estimated by the mgcv library in R. We developed a GAM model to evaluate correlations with terrain attributes derived from the CSMP data, latitude of the segments, whether take was allowed in each segment and the proportion hard bottom habitat in a 30 by 30 m neighborhood around the centroid of a segment. A step by step description of the model-based expansion method is provided below.

Explore removal of data outside species' ranges to reduce over-dispersion in data. (Dr. Williams)

Data used in the GLM for gopher rockfish were from to the extent of their depth and latitudinal distribution. Since the survey was limited to areas where rocky reef was known to exist, with intermittent sections of soft bottom that in part were informing correlations with habitat variables, no further reduction of the data set was undertaken to exclude soft bottom habitat where they are not expected to occur. The primary source of overdispersion for gopher rockfish appeared to be the sporadic encounters with multiple individuals in a segment taken to be representative and the resulting overdispersion was addressed using the negative binomial distribution.

1. Model Selection. Given the AIC results of the GLMs in Table 20, we used the negative binomial distribution for variable selection in the GAM. The criteria for variable selection included whether the variables were significantly correlated with density, the inclusion of the variables in the backward stepwise model selection and the deviance explained by each variable. Candidate variables were reanalyzed to determine whether they remained significant with the refined model and those that remained significant were included in the final model including the depth, latitude, proportion hard bottom from the CSMP and the standard deviation of the slope as seen in Figure 17 with diagnostics in Figure 18 and partials with each variable in Figure 19.

   We evaluated the GAM model with linear fit and the Poisson distribution (results in Figure 20 and diagnostics in Figure 21) as well as the quasi-Poisson (results in Figure 22 and diagnostics in Figure 23), Tweedie (results in Figure 24 and diagnostic in Figure 25) and negative binomial distribution models (results in Figure 17 and diagnostics in Figure 18) to address overdispersion present in the data. We found the negative binomial distribution to provide the best fit as evidenced by the fit to deviance residuals with theoretical quantiles seen in Figure 18 and the partial plots are provided in Figure 19.

   The GAM model with a linear fit and the negative binomial distribution explained 22.1% of the deviance, which was more than 3.8% higher than the remaining models, the closest of which was the Tweedie distribution at 18.3%. In discussion with Dr. Jason Roberts, the developer of MGET he indicated that this was considered a good fit for a habitat model, and the residual pattern with linear predictors in Figure 18 is common given that not all variables explaining the distribution of gopher rockfish can be accounted for. This is expected for count based data with a limited range of outcomes and randomized quantile residuals are the residuals of choice for generalized linear models in large dispersion situations when the deviance and Pearson residuals can be grossly non-normal as noted in Dunn and Sythe (1996). He recommended providing the randomized quantile residuals with the linear predictor as the residual pattern observed in Pearson and deviance residuals will often show a pattern that is resolved by this method. Since randomized quantile residuals cannot be applied to GAMs, we applied them to a GLM with the same variables and the negative binomial distribution and provide a figure of the residuals with the linear predictor as well as the randomized quantile residuals showing no pattern reflective of a well fit model leaving no residual pattern (Figure 26).

We also evaluated the application of flexible splines using the REML method with the negative binomial distribution and the aforementioned selected variables, which increased the deviance explained to 27.5% (Figure 27). The deviance residuals with theoretical quantiles are provided in Figure 28 and the partials with each of the variables are provided in Figure 29. Despite explaining a greater percentage of the deviance, the fit to the deviance residuals did not appear as close as the linear model (Figure28). While this may be the case for gopher rockfish, other species with dome shaped relationships of abundance with depth or other non-linear relationships that are not well represented by a linear model fit may be better modeled with a flexible spline. We proceed with the GAM model with a linear fit and the negative binomial distribution in the remainder of the example for gopher rockfish, though the spline-based results could be applied in an analogous fashion.

```
MODEL SUMMARY:
==============

Family: Negative Binomial(0.265)
Link function: log

Formula:
GopherRF ~ offset(log(UsableArea)) + AvgY + DepthMean_3By3 +
    STDofSlope_3By3 + HardSoft_Proportion

Parametric coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          2.339e+01  2.189e+00  10.683  < 2e-16 ***
AvgY                -7.298e-06  5.330e-07 -13.691  < 2e-16 ***
DepthMean_3By3       2.897e-02  4.732e-03   6.122 9.23e-10 ***
STDofSlope_3By3      9.910e-02  3.228e-02   3.070  0.00214 **
HardSoft_Proportion  1.082e+00  2.370e-01   4.567 4.95e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.0274   Deviance explained = 22.1%
-REML = 1275.5  Scale est. = 1         n = 6786


Method: REML   Optimizer: outer newton
full convergence after 3 iterations.
Gradient range [-8.504637e-05,-8.504637e-05]
(score 1275.532 & scale 1).
Hessian positive definite, eigenvalue range [37.74915,37.74915].
Model rank =  5 / 5
```
Figure 17. Result of GAM model with a linear fit and negative binomial distribution for gopher rockfish proving coefficients used in MGET to predict density.

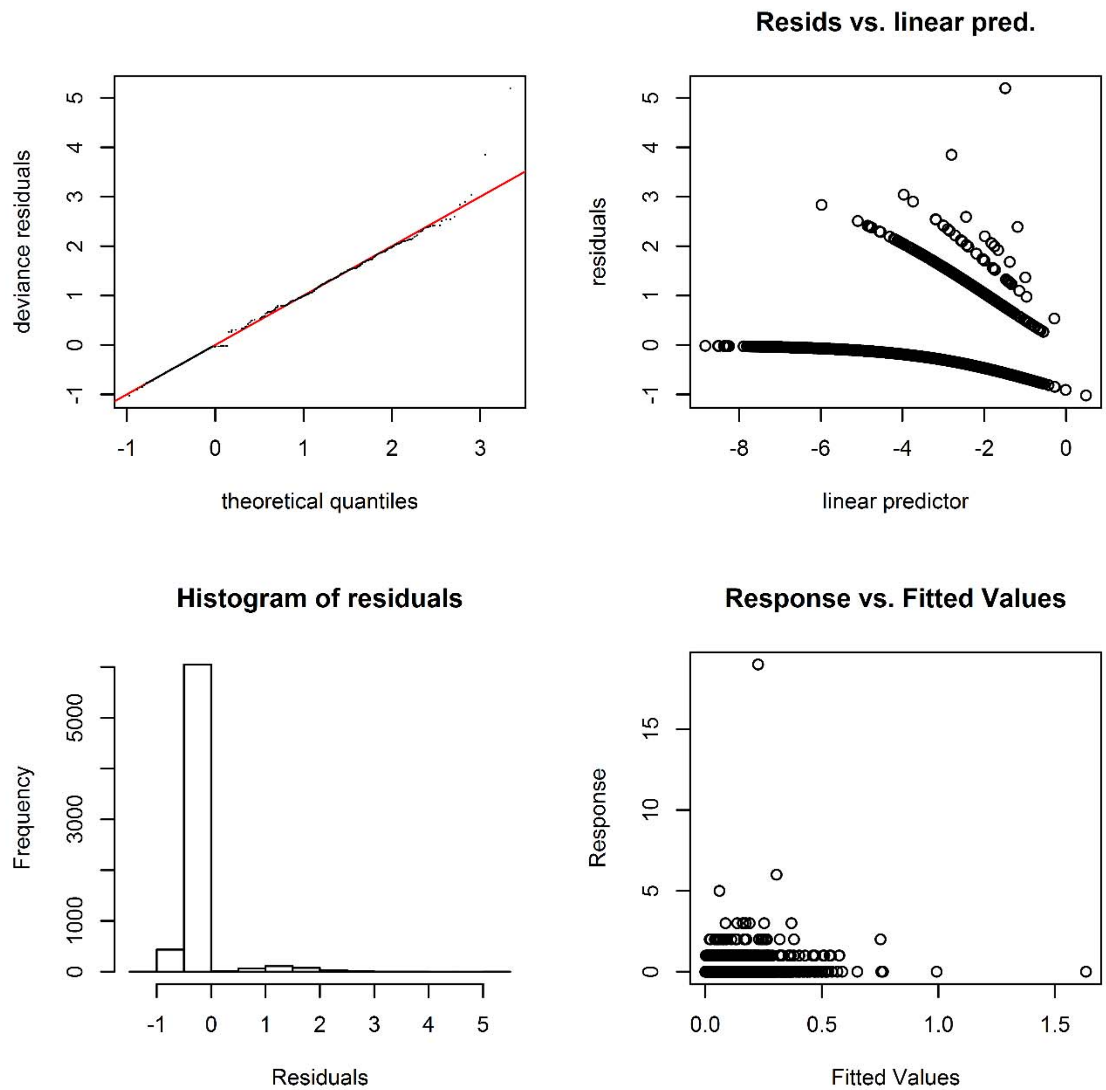

Figure 18. Diagnostics for model fit from R for a GAM model with linear fit and negative binomial distribution for gopher rockfish density with depth, latitude, proportion hard bottom and standard deviation of slope.

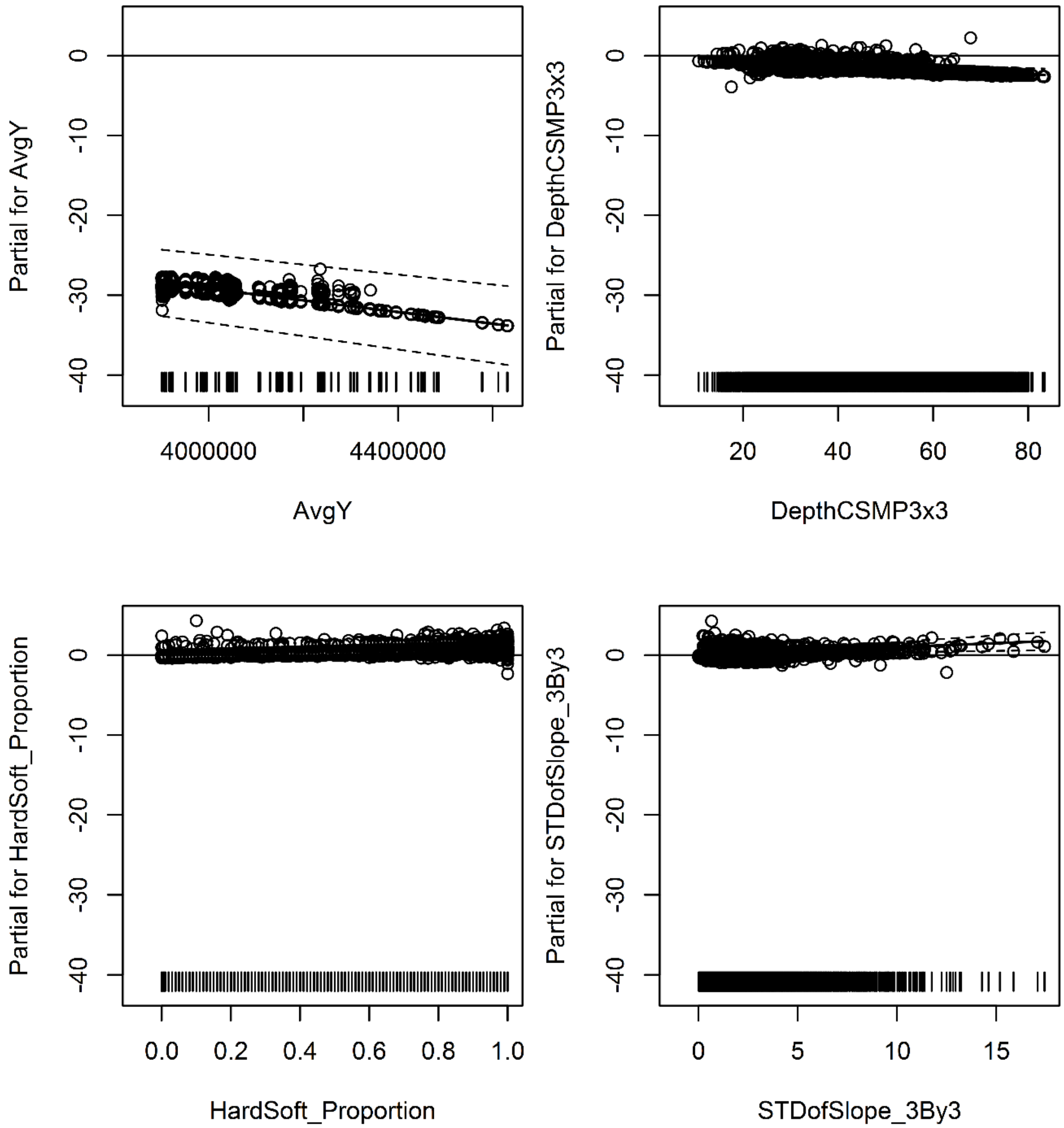

Figure 19. Partial plots for each of the variables included in the GAM model using linear fit and the negative binomial distribution including the proportion of hard bottom, standard deviation of slope, depth and latitude.

```
MODEL SUMMARY:
==============

Family: poisson
Link function: log

Formula:
GopherRF ~ offset(log(UsableArea)) + AvgY + DepthMean_3By3 +
    HardSoft_Proportion + STDofSlope_3By3

Parametric coefficients:
                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)          2.369e+01  1.864e+00  12.711  < 2e-16 ***
AvgY                -7.421e-06  4.569e-07 -16.243  < 2e-16 ***
DepthMean_3By3       2.636e-02  3.911e-03   6.740 1.58e-11 ***
HardSoft_Proportion  1.143e+00  2.024e-01   5.649 1.62e-08 ***
STDofSlope_3By3      8.540e-02  2.448e-02   3.489 0.000485 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.0306   Deviance explained = 18.1%
UBRE = -0.7302  Scale est. = 1         n = 8028


Method: UBRE   Optimizer: outer newton
Model required no smoothing parameter selectionModel rank =  5 / 5
```

Figure 20. Result from a GAM model with linear fit and Poisson distribution for gopher rockfish for comparison to the results of the negative binomial distribution.

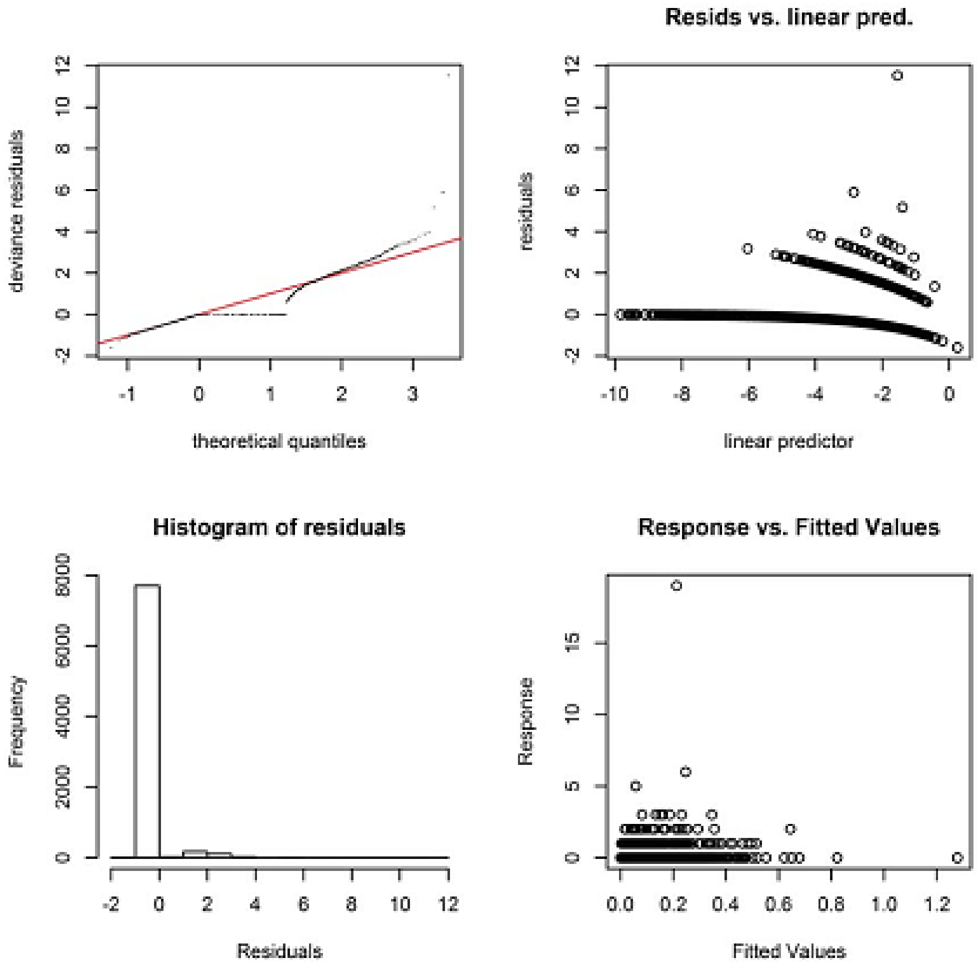

Figure 21. Diagnostics for model fit from R for a GAM model with a linear fit and Poisson distribution for gopher rockfish density with depth, latitude, proportion hard bottom and standard deviation of slope.

```
MODEL SUMMARY:
==============

Family: quasipoisson
Link function: log

Formula:
GopherRF ~ offset(log(UsableArea)) + AvgY + STDofSlope_3By3 +
    DepthMean_3By3 + HardSoft_Proportion

Parametric coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.369e+01  2.204e+00  10.749  < 2e-16 ***
AvgY                -7.421e-06  5.403e-07 -13.736  < 2e-16 ***
STDofSlope_3By3      8.540e-02  2.894e-02   2.950  0.00318 **
DepthMean_3By3       2.636e-02  4.625e-03   5.700 1.24e-08 ***
HardSoft_Proportion  1.143e+00  2.394e-01   4.777 1.81e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.0306   Deviance explained = 18.1%
GCV = 0.26889  Scale est. = 1.3983    n = 8028


Method: GCV   Optimizer: outer newton
Model required no smoothing parameter selectionModel rank =  5 / 5
```

Figure 22. Result from a GAM model for linear fit and quasi-Poisson distribution for gopher rockfish for comparison to the results of the negative binomial distribution.

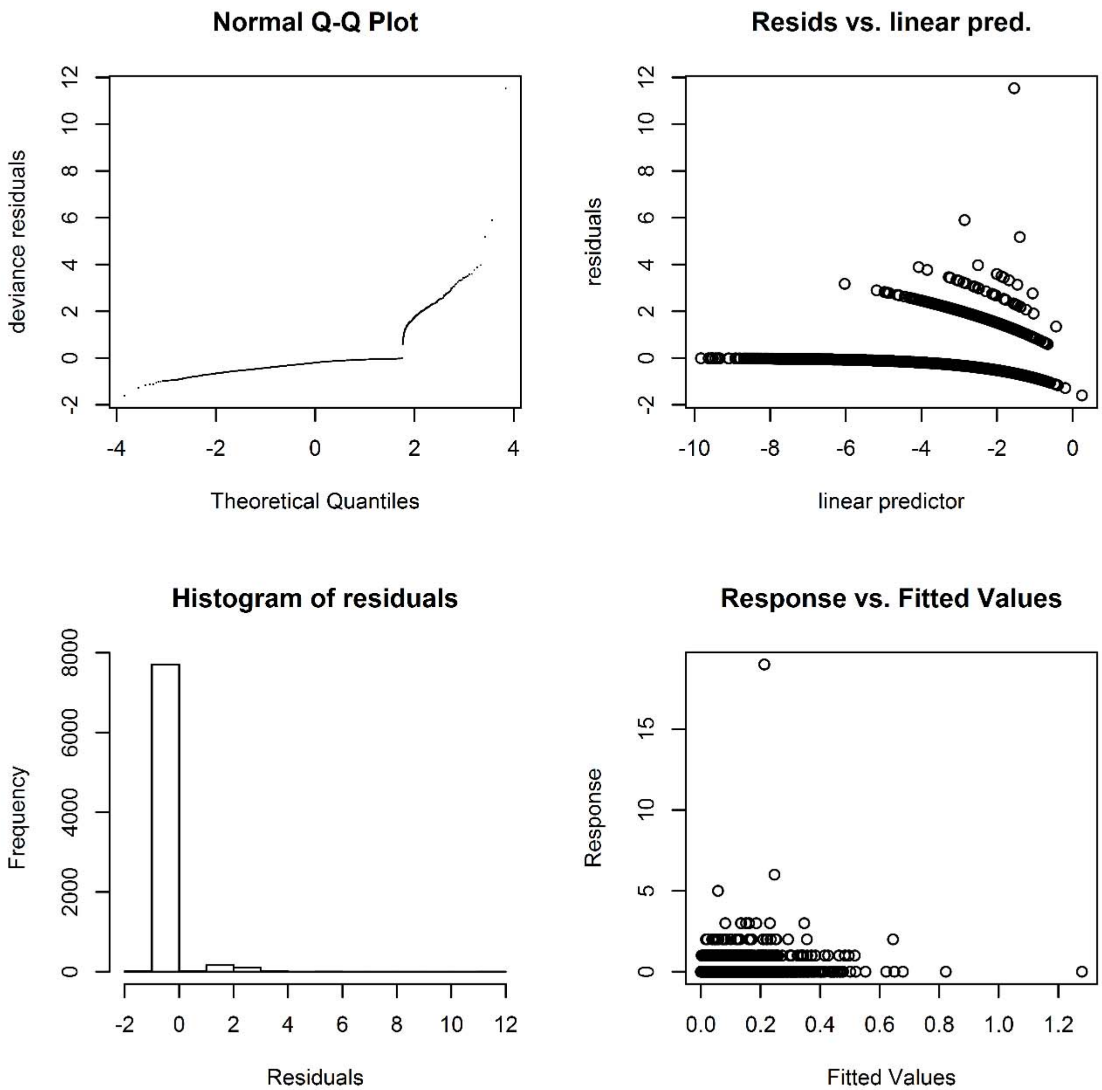

Figure 23. Diagnostics for model fit from R for a GAM model with a linear fit and quasi-Poisson distribution for gopher rockfish density with depth, latitude, proportion hard bottom and standard deviation of slope.

```
MODEL SUMMARY:
==============

Family: Tweedie(p=1.01)
Link function: log

Formula:
GopherRF ~ offset(log(UsableArea)) + AvgY + DepthMean_3By3 +
    STDofSlope_3By3 + HardSoft_Proportion

Parametric coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.371e+01  1.867e+00  12.698  < 2e-16 ***
AvgY                -7.424e-06  4.576e-07 -16.226  < 2e-16 ***
DepthMean_3By3       2.647e-02  3.934e-03   6.729 1.83e-11 ***
STDofSlope_3By3      8.553e-02  2.469e-02   3.464 0.000534 ***
HardSoft_Proportion  1.142e+00  2.033e-01   5.615 2.03e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


R-sq.(adj) =  0.0306   Deviance explained = 18.3%
-REML = 1016.9  Scale est. = 1.0374    n = 8028


Method: REML   Optimizer: outer newton
full convergence after 20 iterations.
Gradient range [-0.002105265,0.0002441918]
(score 1016.868 & scale 1.037391).
Hessian positive definite, eigenvalue range [0.0002441914,37643.13].
Model rank =  5 / 5
```

Figure 24. Result from a GAM model for linear fit and Tweedie distribution for gopher rockfish for comparison to the results of the negative binomial distribution.

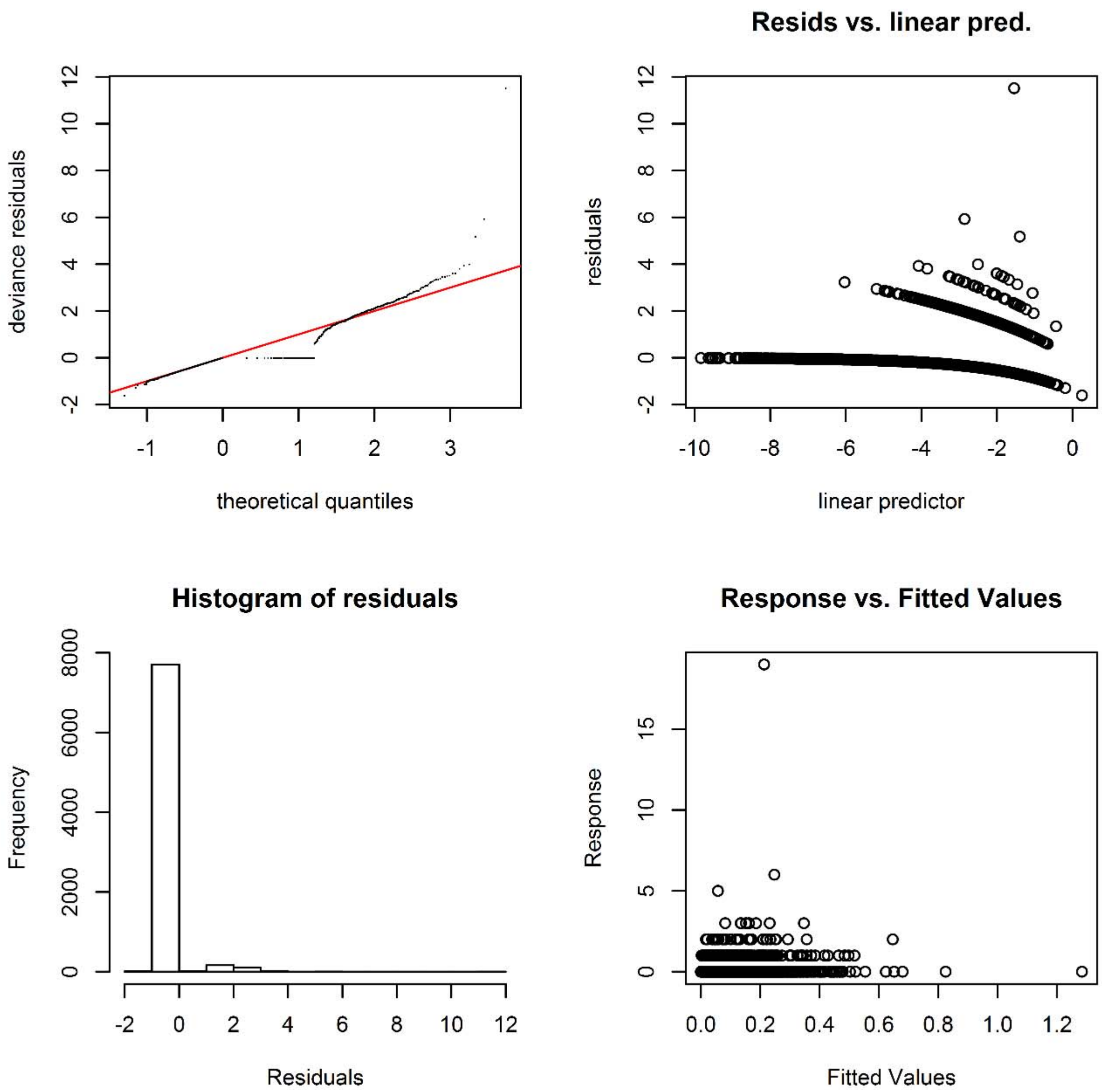

Figure 25. Diagnostics for model fit from R for a GAM model with a linear fit and Tweedie distribution for gopher rockfish density with depth, latitude, proportion hard bottom and standard deviation of slope.

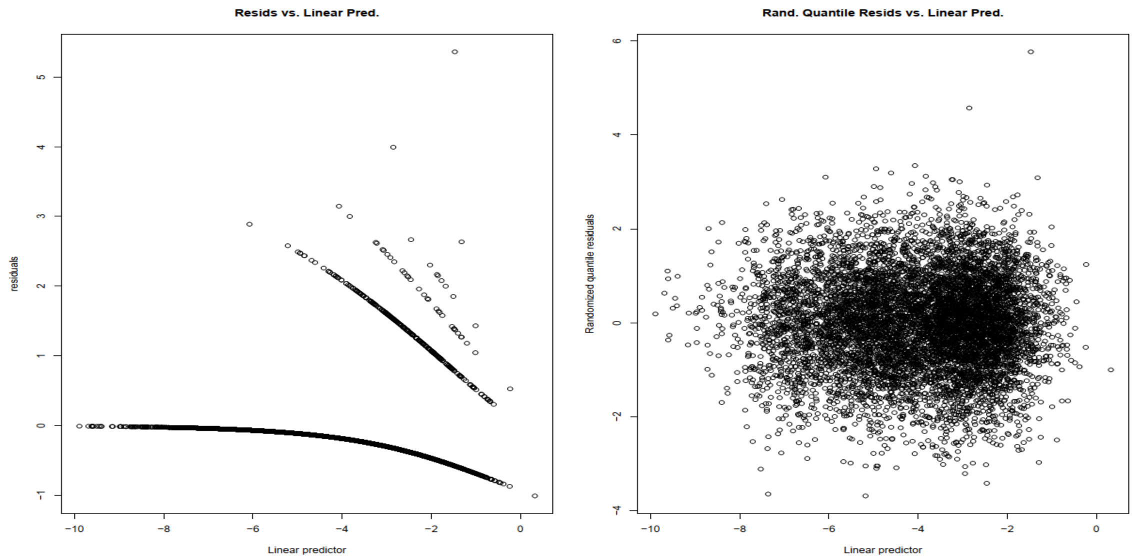

Figure 26. The residuals with the linear predictor and the randomized quantile residuals showing no pattern reflective of a well fit model leaving no residual pattern.

```
MODEL SUMMARY:
==============

Family: Negative Binomial(0.33)
Link function: log

Formula:
GopherRF ~ offset(log(UsableArea)) + s(DepthMean_3By3, bs = "ts") +
    s(STDofSlope_3By3, bs = "ts") + s(HardSoft_Proportion, bs = "ts") +
    s(AvgY, bs = "ts")

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.0669     0.1623   -49.7   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                        edf Ref.df Chi.sq  p-value
s(DepthMean_3By3)      5.1021      9  58.40 7.31e-13 ***
s(STDofSlope_3By3)     0.9082      9   9.34  0.00111 **
s(HardSoft_Proportion) 1.0946      9  28.37 4.19e-08 ***
s(AvgY)                1.5919      9 191.25  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0414   Deviance explained = 27.5%
-REML = 1300.2  Scale est. = 1         n = 8028


Method: REML   Optimizer: outer newton
full convergence after 5 iterations.
Gradient range [-4.792778e-06,2.641704e-05]
(score 1300.197 & scale 1).
Hessian positive definite, eigenvalue range [0.2517562,35.50325].
Model rank =  37 / 37

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                          k'   edf k-index p-value
s(DepthMean_3By3)      9.000 5.102    0.84    0.01 **
s(STDofSlope_3By3)     9.000 0.908    0.87    0.64
s(HardSoft_Proportion) 9.000 1.095    0.85    0.12
s(AvgY)                9.000 1.592    0.78  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 27. Result from a GAM model for gopher rockfish using a negative binomial distribution with REML flexible spline fit.

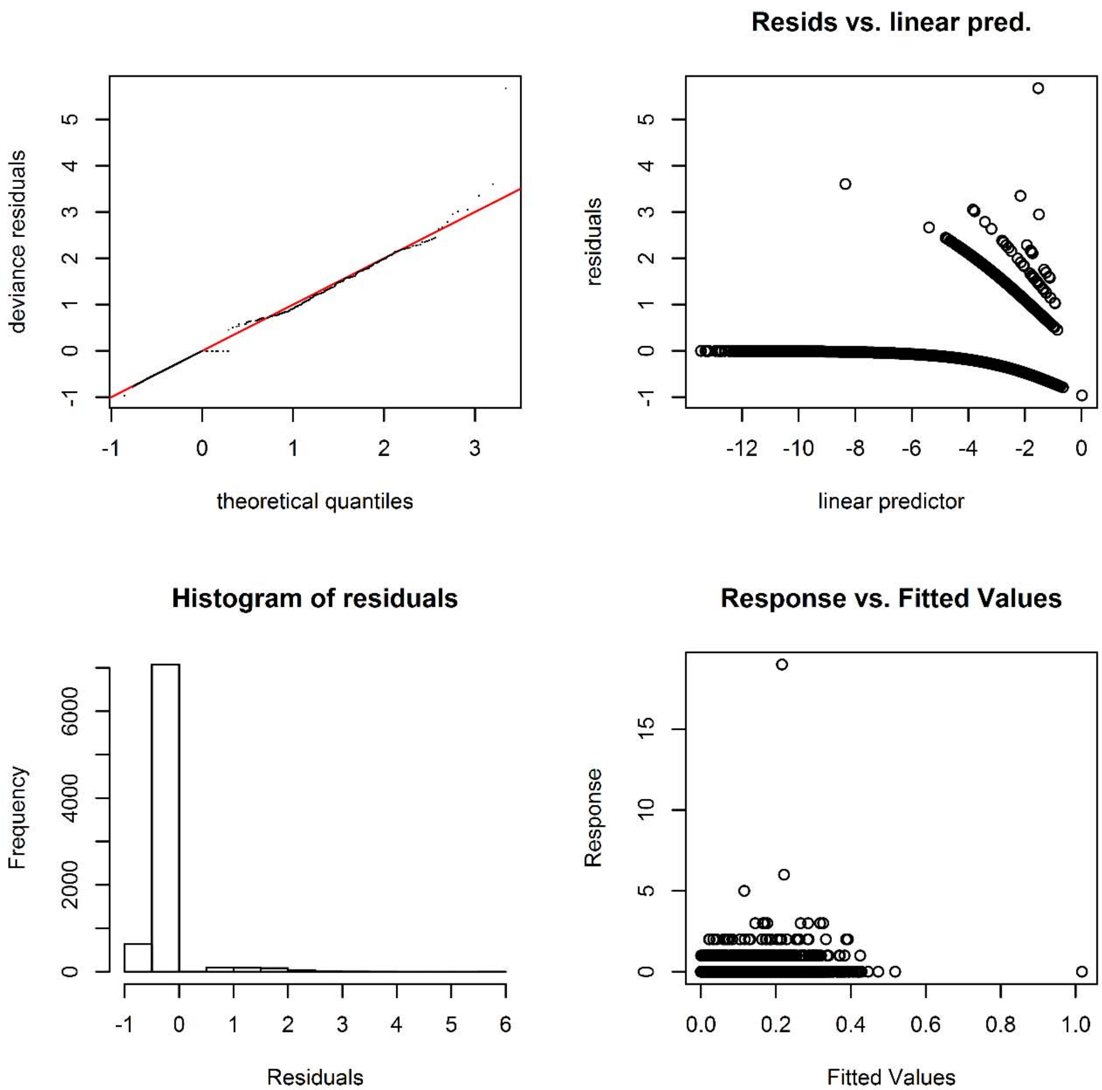

Figure 28. Diagnostics for model fit from R for a GAM model with REML flexible spline fit and a negative binomial distribution for gopher rockfish density with depth, latitude, proportion hard bottom and standard deviation of slope.

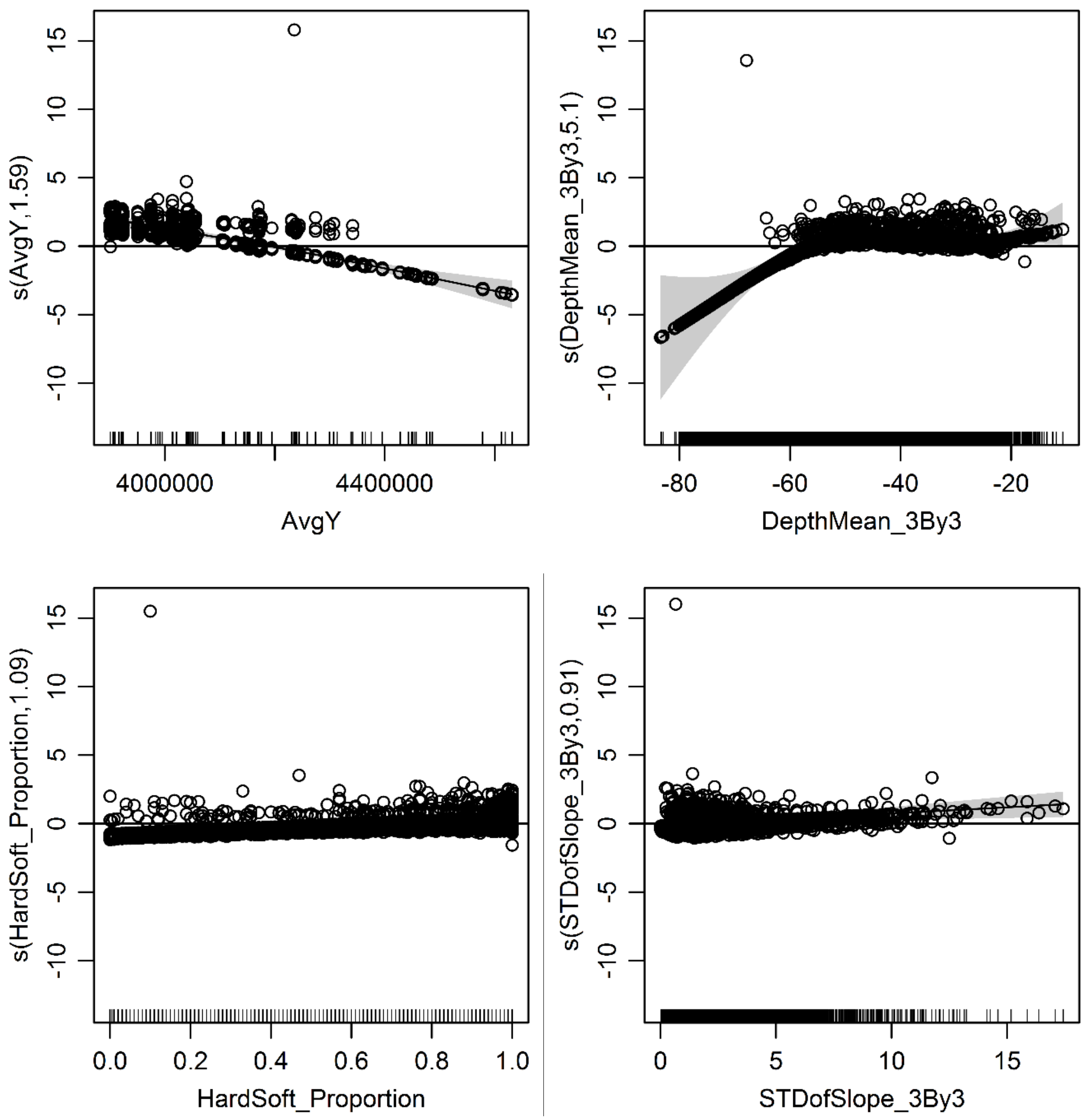

Figure 29. Partial plots for each of the variables included in the final model including, the proportion of hard bottom, standard deviation of slope, depth and latitude from the negative binomial model.

Cross-validation is particularly important for GAMs model development and also care must be taken in the interpretation of results for ranges of certain parameters that exceed most of the sampling, for instance estimates in depth ranges that are not frequently sampled. Evaluate utility of k-fold cross-validation and use of RMSE other tools. (Dr. Trembanis)

A repeated k-fold cross-validation was conducted with k equal to 10 and 5 resulting in training/testing data sets composed of 90%/10% and 80%/20% of the data, respectively, to evaluate models from differing data sources. These k levels have been shown to yield test error rate estimates that do not suffer from excessively high bias or high variance. The repeated analysis ten times allowed validation across the full data set reducing bias from only using partitioning of a single subset of the data. The R^2, root mean squared error (RMSE) and mean absolute error (MAE) were estimated and compared for alternative model configurations.

First a comparison was conducted for models with variables from CSMP, CSMP variables excluding terrain attributes and variables from data collected by the ROV alone (Table G). The results indicated that R squared value was low for all models of gopher rockfish density with either 5 or 10 partitions. The R-squared was lower for CSMP derived variables when the terrain attribute for variability in relief (STDofSlope_3By3) was included than with the model with only latitude (Avg Y), depth (DepthMean_3By3) and bottom type (HardSoft_Proportion). The equivalent model with only variables from the ROV itself resulted in a marginally higher R squared value and lower RMSE and MAE indicating better predictions.

The second comparison conducted was for models derived from the CSMP when one or more variables were excluded (Table H). The results indicate that depth and latitude are the most important variables to include in the model, while exclusion of bottom type variable (PropHardMix) had had a lesser effect. Interestingly, while the model was improved by the exclusion of the terrain attribute for variability in relief (STDofSlope_3By3, it was marginally improved when it was included in the absence of the bottom type (HardSoft_Proportion).

Table G. Results of repeated k-fold cross validation with 5 and 10 folds for models with derived variables from CSMP, CSMP variables excluding terrain attributes and variables from data collected by the ROV alone.

Variables	Partitions/Repeats	RMSE	R Squared	MAE
AvgY, DepthMean_3By3, HardSoft_Proportion, STDofSlope_3By3	10/10	0.289	0.060	0.086
AvgY, DepthMean_3By3, HardSoft_Proportion	10/10	0.289	0.062	0.086
Lat, Depth, PropHardMix	10/10	0.288	0.068	0.085
AvgY, DepthMean_3By3, HardSoft_Proportion, STDofSlope_3By3	5/10	0.301	0.054	0.086
AvgY, DepthMean_3By3, HardSoft_Proportion	5/10	0.301	0.057	0.086
Lat, Depth, PropHardMix	5/10	0.300	0.062	0.085

Table H. Results of repeated k-fold cross validation with 10 folds for models from data derived from CSMP excluding one or more variables.

Variables	Partitions/Repeats	RMSE	R Squared	MAE
AvgY, DepthMean_3By3, HardSoft_Proportion, STDofSlope_3By3	10/10	0.289	0.060	0.086
AvgY, DepthMean_3By3, HardSoft_Proportion	10/10	0.289	0.062	0.086
AvgY, DepthMean_3By3, STDofSlope_3By3	10/10	0.292	0.051	0.087
AvgY, DepthMean_3By3	10/10	0.291	0.048	0.088
AvgY, HardSoft_Proportion, STDofSlope_3By3	10/10	0.292	0.040	0.088
DepthMean_3By3, HardSoft_Proportion, STDofSlope_3By3	10/10	0.296	0.010	0.092

2. Projection of density to rasters using selected variables. Density maps were generated in MGET using the "predict GAM from rasters" function which uses the formula from the GAM and variable values from the CSMP in raster format as the basis for expansion to estimate the density of gopher rockfish in each raster grid. The density map for the northern and southern portion of the central California study region from the GAM with a linear fit and a negative binomial distribution are provided in Figures 30 and 31.

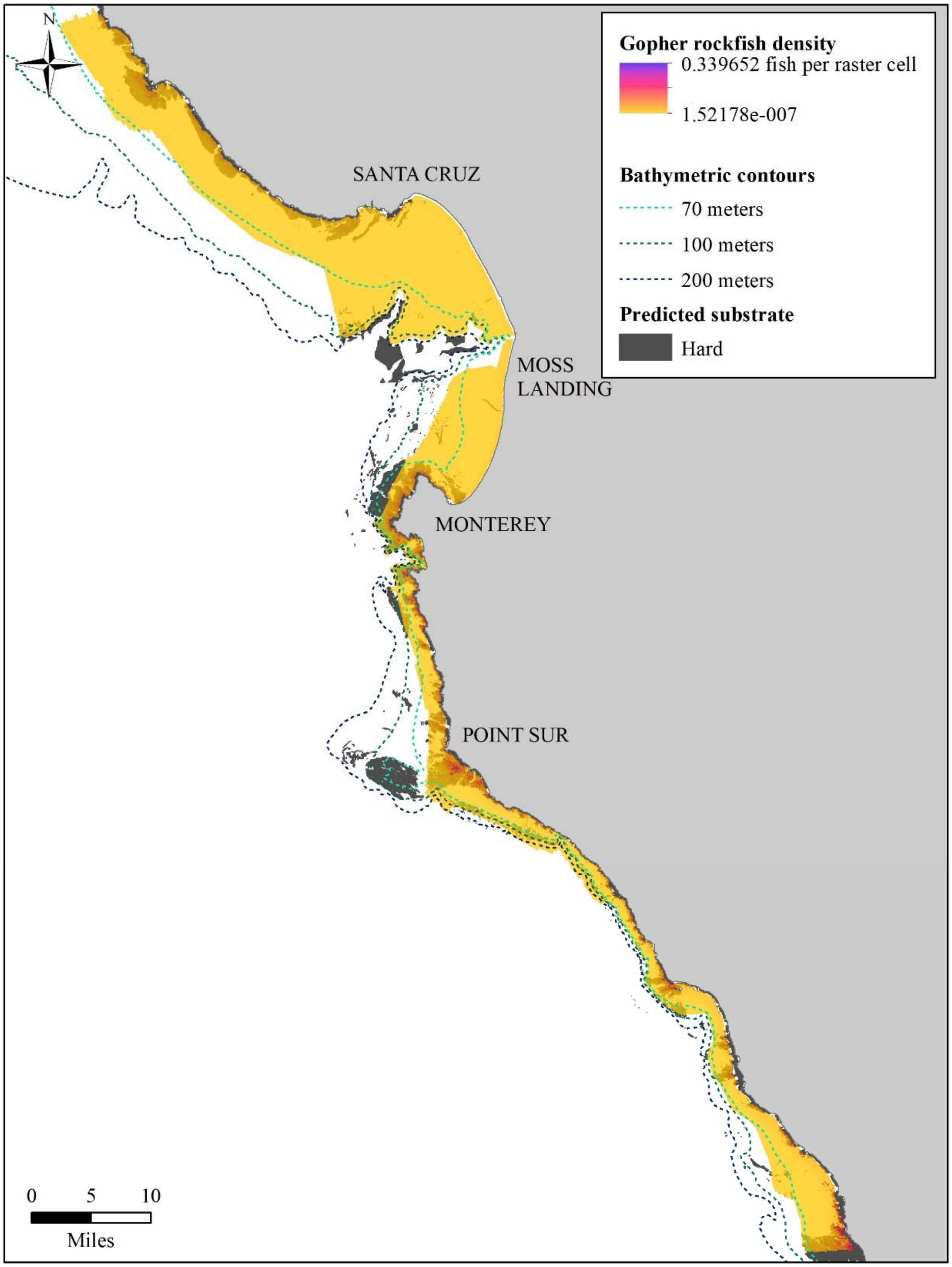

Figure 30. Density map of gopher rockfish per two by two meter raster grid for the northern portion of the study region generated from the GAM using a linear model with a negative binomial distribution including depth, latitude, standard deviation of slope and proportion hard bottom from the CSMP expanded using CSMP mapping in MGET.

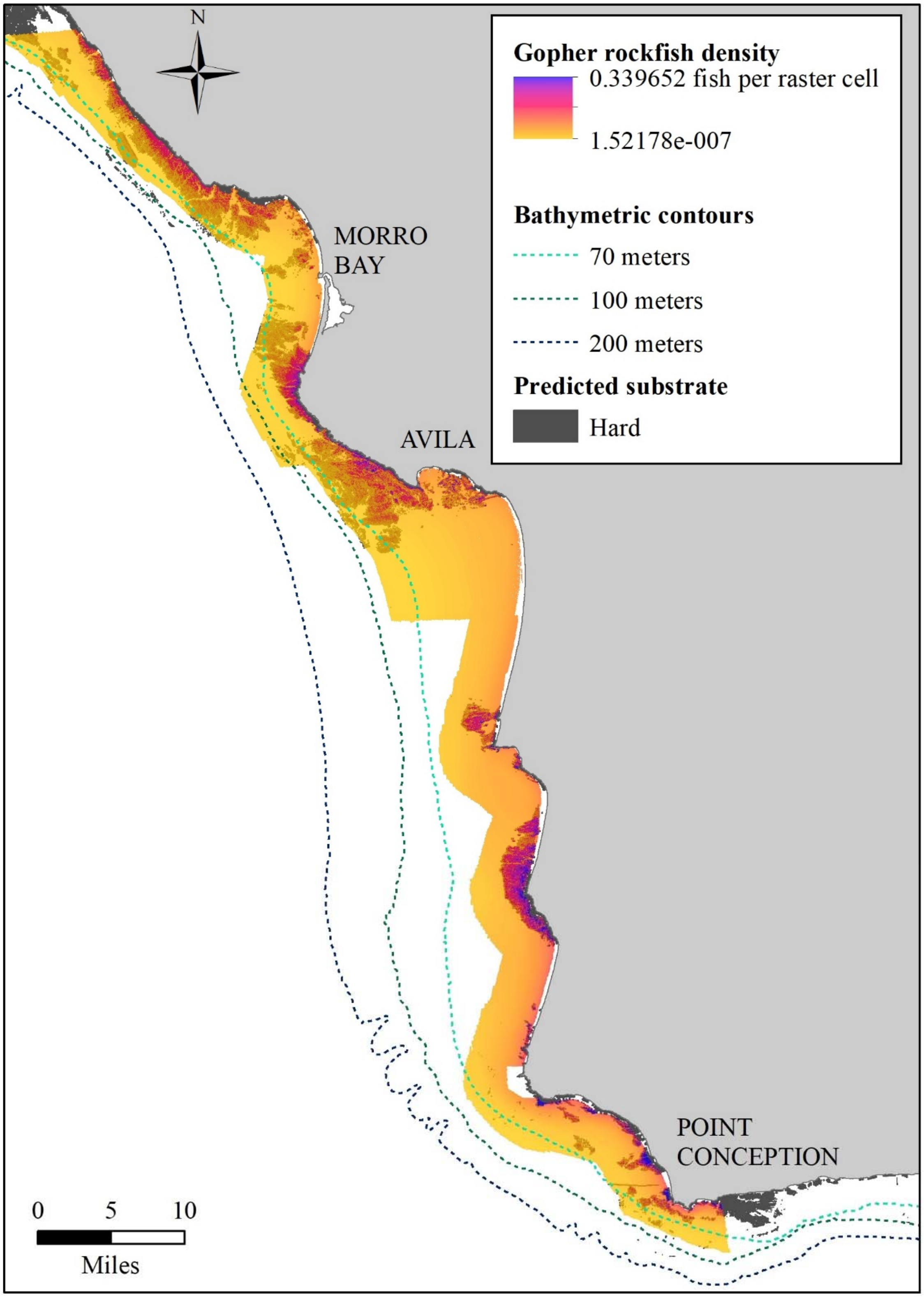

Figure 31. Density map of gopher rockfish per two by two meter raster grid for the southern portion of the study region generated from the GAM using a linear model with a negative binomial distribution including depth, latitude, standard deviation of slope and proportion hard bottom from the CSMP expanded using CSMP mapping in MGET.

3. Summation of projected densities providing abundance estimates. The zonal statistics tool in ArcGIS Spatial Analyst was used to sum the density estimates for each two by two meter raster grid cell across the mapped area to provide an estimate of the abundance of gopher rockfish in numbers of fish across the study area (Figure 31). The estimated number of gopher rockfish over the study area was estimated to be 539,759 fish.

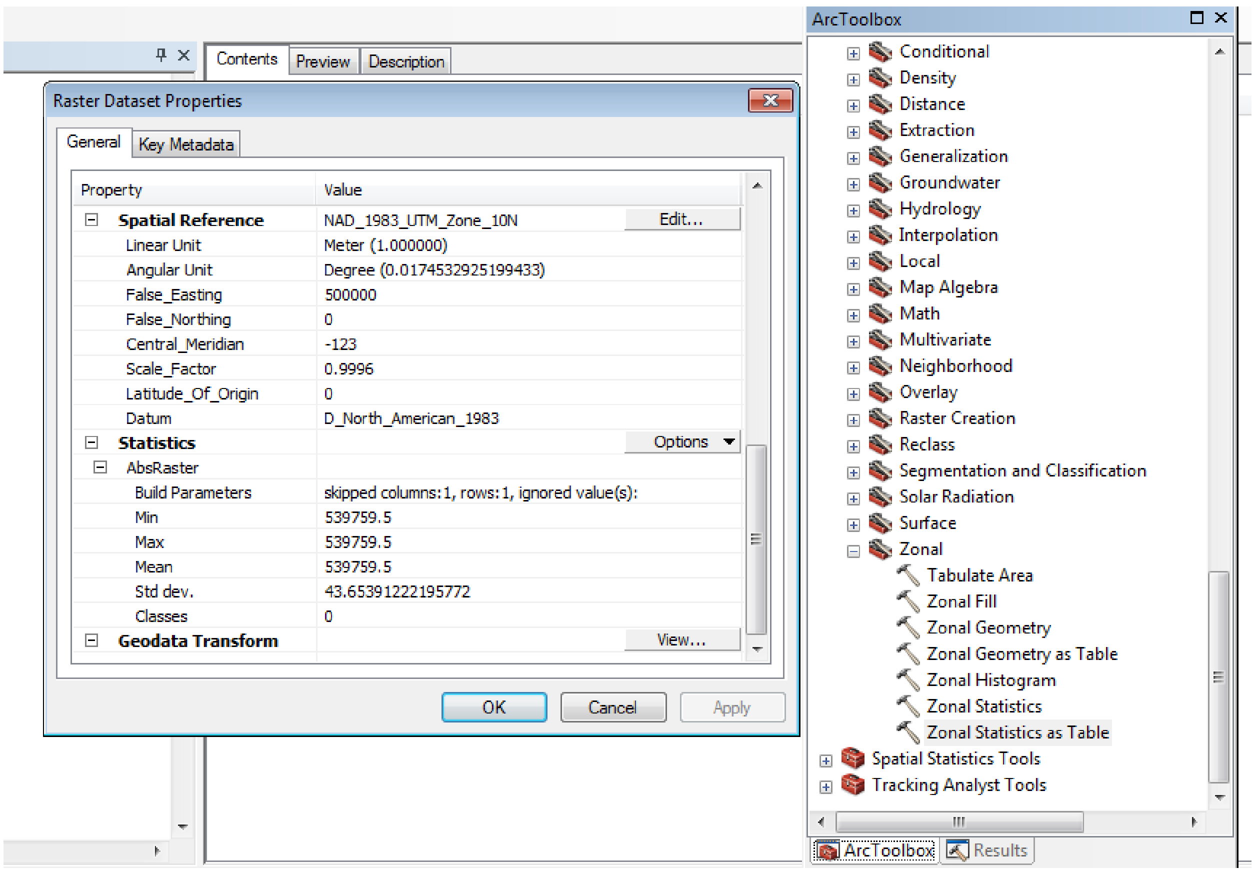

Figure 31. Output of estimated number of gopher rockfish from summation of raster estimates of density per raster cell from the "zonal statistics as table" function in the ArcGIS Spatial Analyst package.

4. Conversion of numbers of fish to metric tons: Average weights were proxied from CRFS estimates of average weight from kept and discarded gopher rockfish in district 3 corresponding to the study area resulting is an estimate of 0.48 kg per fish. The average weight was multiplied by the estimated number of fish from the study area and divided by 1000 to provide an estimate of 259 mt in the study area.

   For the sake of comparison, an expansion using a GLM as described in method 4 outlined above, was conducted using the same variables a with a quasi-Poisson distribution analogous to the GAM approach in steps 1 to 4. This provided a comparable estimate of 498,714 fish or 239.4 mt indicating that the results are robust to the model and distribution used.

5. Variance Estimate: The Math function in ArcGIS Spatial Statistics was used to square all of the standard deviation outputs for each raster-based estimate of density per raster grid to provide an estimate of variance in each grid. The sum of the variance across all raster grids was provided using the Zonal Statistics as Table function in ArcGIS Spatial Statistics resulting in a variance estimate of 13,143 fish or 6.3 mt assuming an average weight of 0.48 kg/fish.

   The variance estimates for density (#6 on page 51) are hard to follow. Please provide an equation for these calculations. Recall that if scaling a variance estimate, the equation is Var(aX) = a^2 * Var(X). If I interpret what is being done correctly, I think you are not squaring the scaler. Also recall that when summing across aggregate estimates of variance you are assuming independence (covariance among strata = 0). (Dr. Berger)

   The standard deviation was estimated as described in the documentation from ArcGIS seen below. The standard deviation for each raster grid was squared to convert it to variance. These values were then summed. Multiplying the variance by the square of the weighted area was not necessary for the model-based method given that each of the raster grid cells had the same area and thus the same weight in the suggested equation making weighting unnecessary. These estimates do not account for potential spatial autocorrelation between segments discussed further below and thus assume independence and the covariance among strata is equal to zero and likely underestimate variance as a result. Given that the variance in density among sites captured in the model is likely greater than the unaccounted-for variance due to spatial autocorrelation, the degree of underestimation of variance may be within the noise of the estimate.

   Standard deviation

   - The standard deviation of the values in each zone is assigned to all cells in that zone.
   - The formula for the standard deviation is as follows:

   $$\text{Std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

   Note: Note that the standard deviation is calculated on the entire population (the "N" method), not estimated based on a sample (the "N-1" method). For comparison, the calculation for standard deviation is equivalent to the STDEVP, not STDEV, method in Microsoft Excel.

6. To validate the geographic distribution of gopher rockfish indicated by the density maps, we superimposed the location and frequency of encounters in each positive segment. The result indicates that the locations where gopher rockfish are projected to be found in high density are consistent with the locations where gopher rockfish were encountered (Figure 32 and Figure 33).

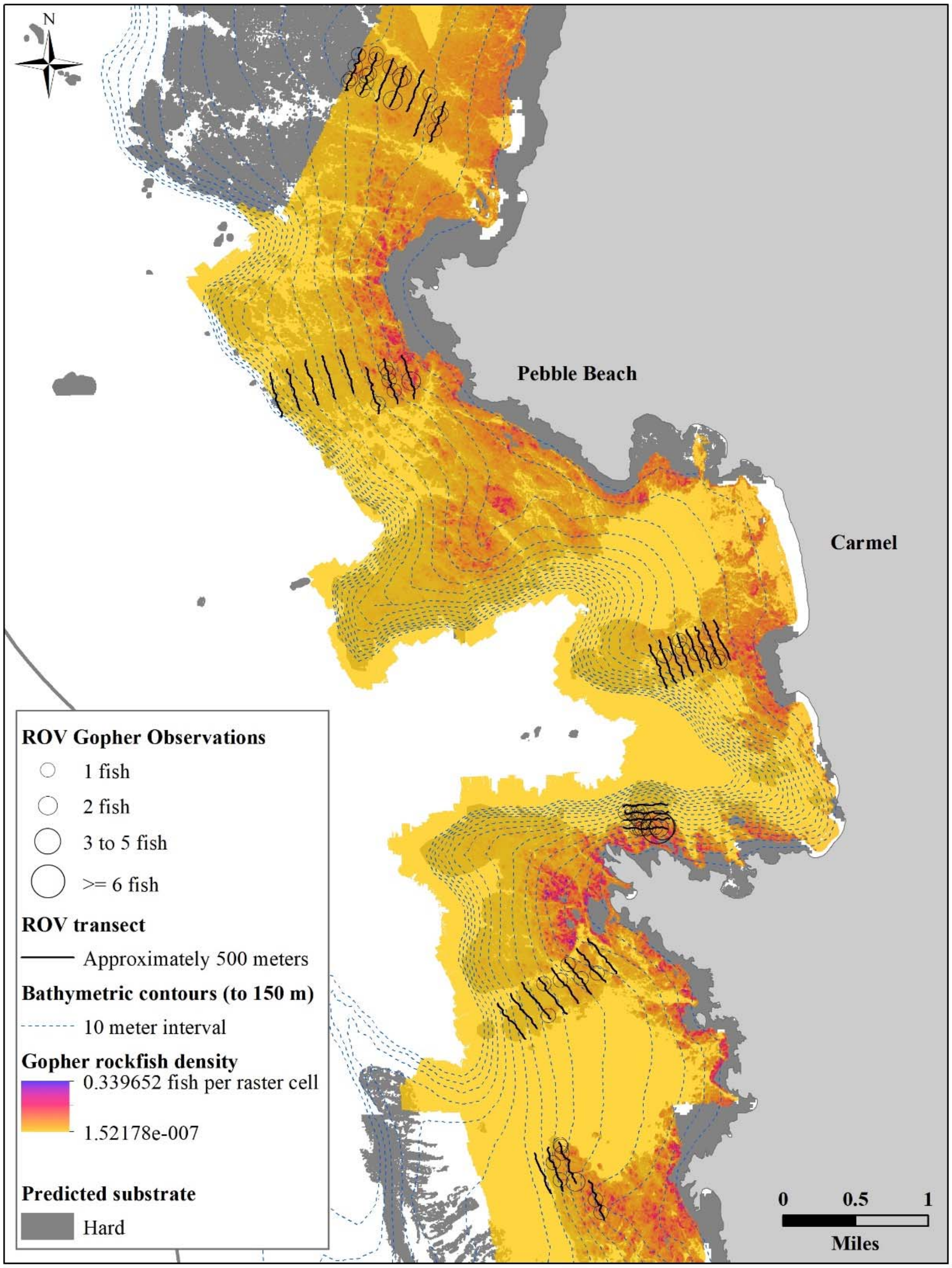

Figure 32. Density map of gopher rockfish per two by two meter raster grid for Carmel Bay area generated from the GAM using a linear model with a negative binomial distribution including depth, latitude, standard deviation of slope and proportion hard bottom from the CSMP expanded using CSMP mapping in MGET. The location of centroids of segments positive for presence of gopher rockfish and the frequency in the segment are identified for validation of density mapping.

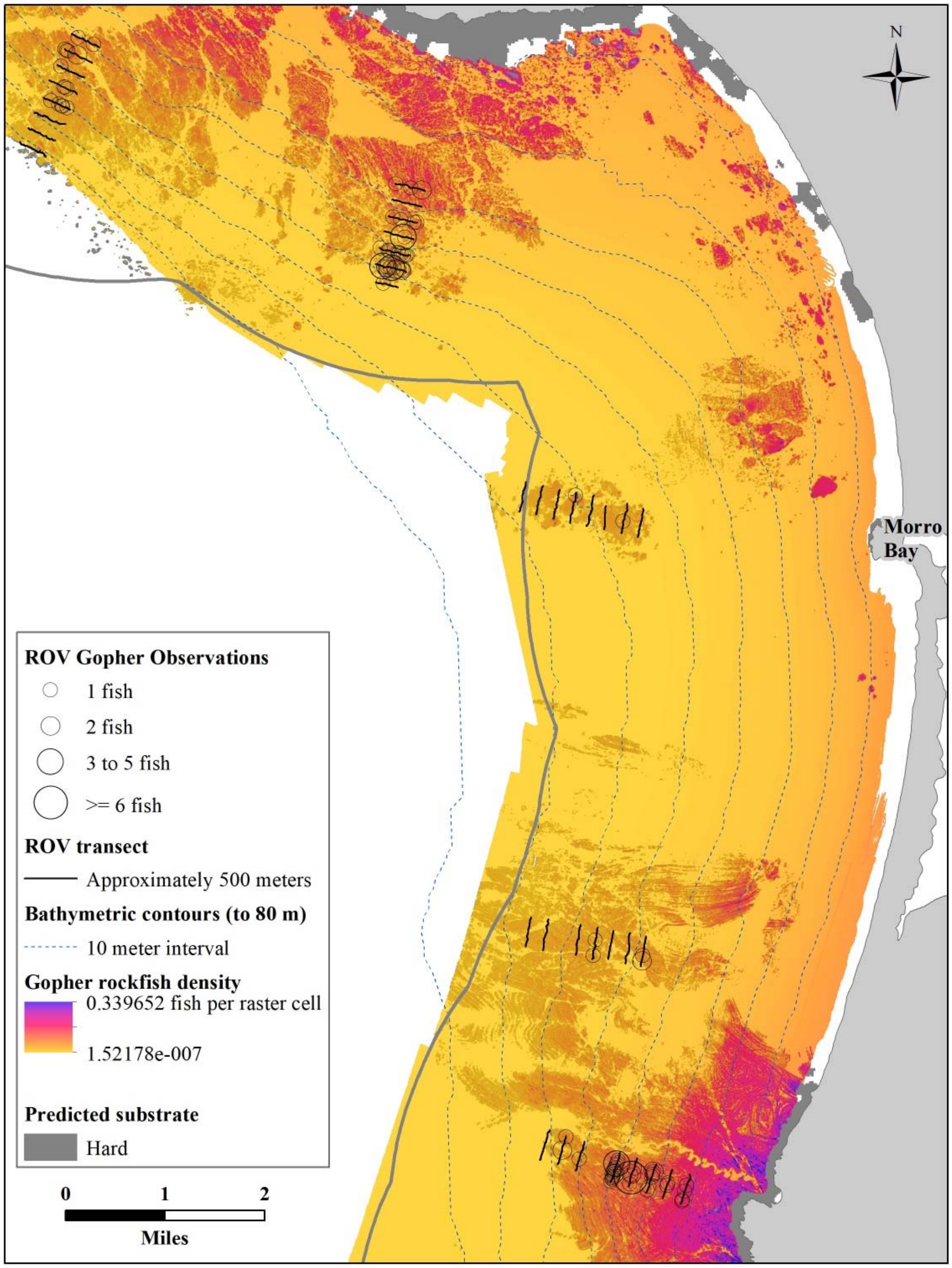

Figure 33. Density map of gopher rockfish per two by two meter raster grid for the Morro Bay area generated from the GAM using a linear model with a negative binomial distribution including depth, latitude, standard deviation of slope and proportion hard bottom from the CSMP expanded using CSMP mapping in MGET. The location of centroids of segments positive for presence of gopher rockfish and the frequency in the segment are identified for validation of density mapping. with centroids of segments positive for presence of gopher rockfish and the frequency in the segment.

There are gaps in mapping coverage that limit where data can be expanded. Future designs should consider prioritization for gap-filling. (Dr. Williams)

Overview of the data available for estimation of biomass for nearshore rockfish species outside of the 2x2 m resolution CSMP

The methods discussed thus far have focused on estimation of biomass in the area with CSMP mapping at a resolution of 2x2 m for which terrain attributes were derived in the vicinity of segments sampled for use in modeling and to inform expansion of density estimates. While this area contains most of the habitat of nearshore rockfish species noted in Table 4, there are areas shoreward and seaward for which habitat and associated biomass of these stocks reside Figure B-D. Habitat categorization and depth data is available seaward of the 2x2 m resolution data at 5x5 m resolutions and from polygons derived for use in mapping EFH as discussed previously. Shoreward of the CSMP mapping, estimates of rocky reef habitat is available from CDFW in collaboration with University of California Santa Cruz and the California Ocean Science Trust. The amount of area in each habitat mapping source is provided in Table D. Unfortunately, the highest resolution of depth data available shoreward and seaward of the 2x2 m resolution CSMP coverage is the U.S. Coastal Relief Model that has an approximate resolution of 90 m. Density estimates from the ROV survey are limited in the shoreward area. Some data from dive surveys or proxy information from proximate depths extrapolated to these areas can be used to inform density there. Estimates of the density seaward of the 2x2 m resolution CSMP data coverage is available from the ROV survey. Below is a description of potential methods that can be used to estimate biomass in the shoreward and seaward area.

Estimation of biomass shoreward of the CSMP data

In the area shoreward of the 2x2 m resolution CSMP data coverage, referred to as the “white zone”, limited data is available from the ROV survey in depths from 0-10 m to provide estimates of density due to constraints posed by kelp and surge preventing access to sample in shallow waters of the subtidal zone. The data available from the most proximate depth bin of 10 to 20 m can be extrapolated into the shallower waters for the design-based methods or the model-based method used to estimate density in this area. In the future, examination of data from the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO) dive surveys may provide density data that can be used to estimate density in shallower depths. Though the spatial coverage of this dive survey and distribution of sampling along the state is limited, it provides an alternative data source and a means of validating estimates from shallow depths where data from the ROV is sparse.

Contributors from The University of California Santa Cruz, California Ocean Science Trust, and California Department of Fish and Wildlife used CSMP and National Oceanic and the National Oceanic and Atmospheric Administration Environmental Sensitivity Index (ESI) shoreline habitat categorizations to generate predictive maps of substrate characteristics in this “white zone” through interpolation. Spatial interpolation methods were used to create a raster that is 30 m in resolution and that predicted the proportion of rock substrate in each pixel. Depth data available from the U.S. Coastal Relief Model that has an approximate resolution of 90 m can be used to inform depth in the shoreward area, though the resolution of the data relative to the rates of change in depth in shallow depths with distance limit the accuracy depth determination.

The estimates of density from the most proximate depth bin can be applied as a proxy estimate of density to apply in the 0-10 m depth bin in design-based estimates. For model-based estimates, the density from a simple model with bottom type, depth and latitude as covariates would be applied to a 2x2m grid superimposed on white zone seafloor classification polygons allowing depth and latitude to be used to estimate abundance in each grid cell and summed across the rocky reef to provide a total estimate. The average weight from the most proximate depth bin would be applied to provide estimates of biomass in the white zone. The estimate of biomass in the white zone from model or design-based methods would be combined with estimates from the remaining areas to provide a complete estimate of biomass.

Estimation of biomass seaward of the 2x2m resolution CSMP data.
In the area seaward of the 2x2 m resolution data in deeper water within state waters3 miles from shore where there is coarser scale 5x5 m resolution CSMP bottom classification data, but scale issues complicate estimation of terrain attributes on a comparable resolution. Further from shore, bottom type classifications derived from high resolution multibeam data are not available, though bottom type can be derived from polygons developed for use in the evaluation of Essential Fish Habitat in the Council process. EFH habitat classification data results from compilation of several data sources and is subject to greater uncertainty as described previously. Depth data available from the U.S. Coastal Relief Model with an approximate resolution of 90 m can be used to inform depth in the seaward area, though the resolution of the data limits the accuracy depth determination. This area contributes a substantial proportion of the total seafloor in nearshore rockfish habitat in depths less than 120 m where species of interest were encountered in this study (Table D). The uncertainty in depth and bottom classification as a result of the proportion of habitat that resides in the low-resolution area varies between species, increasing with depth distribution as depicted in Figures B through D for species of interest.

A grid at the 10x10 m resolution can be superimposed on the polygons describing habitat from the EFH mapping project and depths from the U.S Coastal Relief Model for use in estimating rocky reef habitat area in each depth stratum in the seaward area. A simple model of density with bottom type, depth and latitude from the ROV survey data can be used to estimate density for application to the habitat and depth classifications at a 10x10 m resolution to estimate abundance seaward. For the design-based method, estimates of density for each respective depth and latitude stratum will be multiplied by the area of rocky reef in the respective stratum and average weight of the observed species for that stratum applied, which would then be summed across all strata to provide estimates of the biomass in the seaward area.

The biomass for each respective area will be combined with those from other areas to provide a comprehensive estimate of biomass and variance.

Expansion Methods Considered but not Explored
Additional model-based approaches could be explored. Time constraints and other considerations noted below prohibited us from pursuing them at this time.

- Point process models implemented using the igcp and geostatsp packages in R (Chakraborty et al. 2011, Hedley and Buckland 2004). This method is promising and is currently being pursued for reef level estimates of abundance by a postdoctoral research collaborating with the CDFW Marine Protected Area project.

- Maximum entropy models in the program Maxtent (Philips and Dudik 2008) relying on presence data was considered for use in estimation of density and expansion across strata. There was concern about spatial error in georeferencing observations from the ROV with the two by two-meter grid cells as opposed to characterizing larger areas around segments of transect as was done in MGET.

Comparison of Design-based and Expansion-based Expansions

Neither the design-based nor expansion-based expansions account for abundance in the unmapped portion of the seafloor shoreward or seaward of the CSMP, though alternative methods may be available to estimate abundance shoreward from CDFW mapping efforts and data from dive surveys. The relative proportion of the total abundance unavailable shoreward and seaward of the CSMP within the range of a given species should be considered for either method. For gopher rockfish a relatively small proportion of the total habitat has been omitted, however the core of its depth distribution is represented in the CSMP data. The design-based expansion method does have the advantage of having accounted for area within the five by five meter and 10 by 10 m resolution CSMP mapping in deeper depths where the resolution was limited due to transducer cone width, in addition to the two by two meter resolution data that the model-based method was able to account for. The two by two-meter resolution data is the only resolution of seafloor captured in the model-based method due to the scale at which the density relationships in the GAM were defined. This may make the model-based estimation method biased low relative to the design-based estimation method. Conversely, the larger latitudinal extent of the area encompassed in the model-based method should be considered when comparing results.

The results of design-based methods were dependent on stratification applied and the basis for the habitat area included in expansions. The use of only seafloor categorized as rock provided an estimate of 269 mt to 281 mt depending on the stratification, which is comparable to the estimate from the model-based expansion method which indicated 259 mt. If the criteria for habitat area are loosened to include habitat meeting the threshold of a minimum of 10% of hard bottom habitat in a 30 by 30 m neighborhood, then abundance estimates of 444 mt to 465 mt result depending on the stratification by depth. A more conservative threshold may yield a design-based estimate intermediate to the results observed between these relatively extreme bottom habitat area inclusion criteria.

The model-based method accounts for relief in the vicinity in addition to the proportion of hard bottom as well as latitude and depth, accounting for more dimensions of variability in the estimate in a more systematic framework, which provides merit worthy of consideration. Weighting schemes or averaging of results of these methods or alternative methods could be evaluated in the future to balance considerations. The model-based method also resulted in lower variance estimates than the design-based method. This may in part be the result of having used data from all segments across the state to define the density model for the GAM in the model-based method as opposed to the limited number of examined segments in each latitude and depth of the design-based method, as well as accounting for more of the variables contributing to the density of gopher rockfish.

Percent Reef Area Sampled, Usable Area, Segments, Encountered Fish, Standard Deviation and Variance for Density Estimates of Gopher Rockfish

To provide some perspective on the spatial coverage of sampling and representativeness of the estimates of density we provide the following. The total area sampled within 70 meters where gopher rockfish were encountered contributing to density estimates in the design-based method between 35 and 37 Deg. N. Lat. was 107,504 square meters, as compared with the model-based inference that used density information in all areas north of Point Conception over which 255,396 square meters were sampled. While this study represents one of the most comprehensive ROV surveys on the Pacific Coast, the percent of habitat for gopher rockfish sampled was less than a tenth of 1% of the potential habitat, making clear that future efforts to increase the sampled area would be beneficial. The total number of square meters of usable area sampled over transect lines in each depth and latitude north of Point Conception is provided in Table 30.

The total number of 20 m segments of transect sampled in each depth and latitude north of Point Conception is provided in Table 32. The number of 20 m segments of transect within 70 meters where gopher rockfish were encountered contributing to density estimates in the design based method between 35 and 37 Deg. N. Lat. was 2,505, as compared with the model-based inference that used density information in all areas north of Point Conception informing the model-based methods over which 6,326 segments were sampled.

The total number of gopher rockfish encountered in each depth and latitude north of Point Conception is provided in Table 33. The number gopher rockfish encountered contributing to density estimates in the design-based method between 35 and 37 Deg. N. Lat. was 230 fish, as compared with the model-based inference that used density information in all areas north of Point Conception over which 301 fish were sampled. The density of gopher rockfish (fish/square meter) encountered in each depth and latitude north of Point Conception is provided in Table 34, the corresponding standard deviation is provided in Table 35 and the variance in Table 36.

This information is provided to give an indication of the relative availability of data, sample size of segments, sampled usable area and their implications for estimates of density and associated uncertainty in application to producing design-based or model-based estimates abundance estimates for gopher rockfish and other species in the remainder of the sampled range in the future.

Table 30. Proportion of gopher rockfish habitat within 70 m in CSMP data sampled with the design or model-based methods in the study area for each method and north of Point Conception.

Expansion Method	Habitat Criteria	Region	Usable Area Sampled in Gopher Rockfish Habitat	Total Gopher Rockfish Habitat in CSMP	Percent Sampled
Design	Classified Rock in CSMP	35-37 Deg. N. Lat	107505	179809822	0.060%
Design	10% Rock in 30x30m	35-37 Deg. N. Lat	107505	281839163	0.038%
Model Expansion Area	Classified Rock in CSMP	34 Deg. 30 Min. N. Lat. - 37 Deg. 11 Min. N Lat	112295	226867423	0.049%
Model Expansion Area	10% Rock in 30x30m	34 Deg. 30 Min. N. Lat. - 37 Deg. 11 Min. N Lat	112295	361295549	0.031%
Model Sample Area	Classified Rock in CSMP	Pt. Conception-OR/CA Border	255397	595672765	0.043%
Model Sample Area	10% Rock in 30x30m	Pt. Conception-OR/CA Border	255397	908799115	0.028%

Table 31. Total square meters of usable area sampled over transect lines in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			86					86
10-20	84	153	2328	1067	69			3701
20-30	6762	6597	21238	11406	3395	2549	1267	53213
30-40	10629	20778	14703	12707	7068	6616	6502	79003
40-50	9612	11316	3936	12649	4056	2021	1014	44604
50-60	14584	7941	4169	9105	2291	2631	34	40755
60-70	10383	8667	1945	6683	3604	2356	395	34033
70-80	4926	4169	4282	1603	2256	1970	44	19249
Total	56980	59620	52688	55219	22739	18144	9256	274646

Table 32. Total number of 20 m segments sampled over transect lines in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			2					2
10-20	2	4	56	28	2			92
20-30	158	151	526	300	83	70	36	1324
30-40	257	461	363	329	176	177	182	1945
40-50	239	245	96	338	105	53	29	1105
50-60	348	181	95	253	60	70	1	1008
60-70	262	197	45	182	92	58	14	850
70-80	115	93	102	46	56	46	2	460
Total	1381	1332	1285	1476	574	474	264	6786

Table 33. Total number gopher rockfish encountered in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			0					0
10-20	0	1	6	1	0			8
20-30	19	10	20	13	0	0	0	62
30-40	47	27	11	8	0	0	0	93
40-50	35	28	0	8	1	0	0	72
50-60	31	31	1	1	0	0	0	64
60-70	0	1	0	1	0	0	0	2
70-80	0	0	0	0	0	0	0	0
Total	132	98	38	32	1	0	0	301

Table 34. Average density gopher rockfish encountered (fish per meter squared) in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			0.0000					0.0000
10-20	0.0000	0.0121	0.0024	0.0009	0.0000			0.0023
20-30	0.0033	0.0026	0.0011	0.0012	0.0000	0.0000	0.0000	0.0014
30-40	0.0080	0.0017	0.0008	0.0008	0.0000	0.0000	0.0000	0.0017
40-50	0.0049	0.0033	0.0000	0.0006	0.0003	0.0000	0.0000	0.0020
50-60	0.0023	0.0051	0.0002	0.0001	0.0000	0.0000	0.0000	0.0018
60-70	0.0000	0.0002	0.0000	0.0002	0.0000	0.0000	0.0000	0.0001
70-80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Total	0.0033	0.0023	0.0008	0.0006	0.0001	0.0000	0.0000	0.0014

Table 35. The standard deviation of the density gopher rockfish encountered (fish per meter squared) in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			0.0000					0.0000
10-20	0.0000	0.0241	0.0070	0.0050	0.0000			0.0078
20-30	0.0094	0.0144	0.0057	0.0056	0.0000	0.0000	0.0000	0.0074
30-40	0.0385	0.0078	0.0044	0.0054	0.0000	0.0000	0.0000	0.0150
40-50	0.0131	0.0109	0.0000	0.0039	0.0028	0.0000	0.0000	0.0085
50-60	0.0078	0.0124	0.0023	0.0017	0.0000	0.0000	0.0000	0.0073
60-70	0.0000	0.0022	0.0000	0.0027	0.0000	0.0000	0.0000	0.0016
70-80	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Total	0.0184	0.0096	0.0046	0.0043	0.0012	0.0000	0.0000	0.0098

Table 36. The variance of the density gopher rockfish encountered (fish per meter squared) in each depth and latitude north of Point Conception.

	Latitude Deg. N. Lat.							
Depth (m)	35°-36°	36°-37°	37°-38°	38°-39°	39°-40°	40°-41°	41°-42°	Total
0-10			0.000000					0.000000
10-20	0.000000	0.000581	0.000049	0.000025	0.000000			0.000061
20-30	0.000088	0.000209	0.000033	0.000031	0.000000	0.000000	0.000000	0.000055
30-40	0.001479	0.000061	0.000019	0.000029	0.000000	0.000000	0.000000	0.000224
40-50	0.000172	0.000119	0.000000	0.000015	0.000008	0.000000	0.000000	0.000073
50-60	0.000061	0.000153	0.000005	0.000003	0.000000	0.000000	0.000000	0.000053
60-70	0.000000	0.000005	0.000000	0.000007	0.000000	0.000000	0.000000	0.000003
70-80	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Total	0.000337	0.000092	0.000022	0.000018	0.000001	0.000000	0.000000	0.000096

Analysis of spatial autocorrelation and the potential effect on variance estimates. Could geostatistics (i.e., variograms) be used to identify a minimum segment distance to make a reasonable assumption about independent samples (where 'samples' are based on segment distance)? (Dr. Berger)

We used Moran's I to test for the presence of spatial autocorrelation of density between segments. The results were significant for all species examined at the analyzed scale of 1000 m indicating spatial autocorrelation at the site level (Table I). If estimates were being made at a reef level, spatial autocorrelation would be a greater concern, but they are based on data aggregated over hundreds of miles of coastline resulting in independence of most segments. Given the limited mobility of these species as adults, the aggregation of segments from much larger distances over a multitude of sample sites mitigates the effect of spatial autocorrelation between segments for both the model and design-based estimation methods.

In addition, we estimated Ripley's K at a scale of 500 m for gopher rockfish to evaluate how spatial autocorrelation varies with distance, which appeared to decrease at the scale of 400 m (Figure L). Selection of sufficiently large segments to reduce spatial autocorrelation within a given site would negate the ability to derive terrain attributes or other variables at a scale that allows inference regarding the variables correlated to the density of each species. It is preferable to incur spatial autocorrelation even if estimates were made at a site level and summed thereafter in order to account for the effect of correlated variables on density. In addition a review by Dorman et al. (2007) indicates that attempts to account for spatial autocorrelation in producing expansions is not recommended since the spatial autocorrelation structure for the observed areas may differ greatly from areas it is being expanded to due to differences in demographic or environmental conditions between locations. While we acknowledge that spatial autocorrelation exists and that this may cause variance to be underestimated, it cannot be adequately addressed in the current context.

Table I. Results of Moran's I test for spatial autocorrelation.

Species	Moran's I	Z-Score	P-value
Gopher Rockfish	0.0538	41.418	0.000000
Copper Rockfish	0.0060	4.760	0.000002
China Rockfish	0.0173	12.251	0.000000
Vermilion Rockfish	0.0171	15.188	0.000000

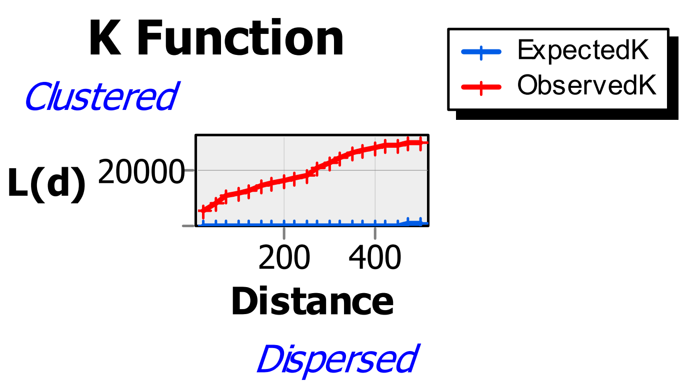

Figure L. Ripley's K analysis of spatial autocorrelation with distance for gopher rockfish within 500 m.

Future Refinements

- Estimation of average weights from lengths of observed fish in the ROV survey using length/weight relationships from the recreational fishery to convert lengths to weights. In addition, testing for trends in weights estimated from ROV by depth due to ontogenetic migration to deeper depth should be undertaken. This would inform whether summing density per raster to estimate abundance by depth is necessary to apply depth specific average weights.

- Evaluation of the consistency of significant correlation of densities with terrain attributes across areas would be of interest in determining whether the associations are consistent or due to spurious correlation.

- Explore the use of bootstrap methods to provide design-based estimates of abundance that are more robust to spatial autocorrelation as well as variance estimates.

- Evaluation of ANOVA or more rigorous methods of postratification of data for design-based estimation methods.

- Analysis of spatial autocorrelation and the potential effect on variance estimates. When developing confidence intervals for estimates, the variance may be underestimated if there is spatial autocorrelation between observed segments. In addition, spatial autocorrelation can also affect coefficients of estimates for covariates and their significance. The inclusion of segments from disparate areas in the estimates from design-based methods and in developing the model-based methods may decrease the potential for effects of spatial autocorrelation on variance estimates, but accounting for it explicitly would be informative. Use of bootstrap methods drawing across multiple reef areas may also reduce the effect of spatial autocorrelation.

- Use of flexible spline function to fit non-linear relationships in data. This may be essential for other species for which there are dome shaped relationships with variables that linear models do not adequately model. Though we used a GAM to model the correlation of variables with gopher rockfish density, we did not utilize the flexible spline function in expansions. Improved fits were possible for gopher rockfish as evidenced by the increase the deviance explained, though the linear model appears to provide a reasonable fit and results are consistent with the scale of the design-based methods. Alternatively, polynomials can also be tested to examine whether they would improve fit over linear relationships.

- Evaluation of alternative raster grid resolution in deriving terrain attribute variables from CSMP depth data to examine whether the correlations are scale-dependent and alternative resolutions result in more consistent correlations.

- Analysis of alternative habitat criteria for determining the proportion hard bottom threshold for inclusion of CSMP habitat in expansions would better inform the appropriate basis for design-based expansions.

- Provide estimates of abundance in the 0-10 m depth bin using proxy information from the next deeper depth bin in the design-based method. This proxy could also be applied to the area estimates for hard bottom habitat from CDFW GIS efforts for hard bottom habitat in the 0-10 m depth range unavailable from the CSMP.

- Provide estimates of abundance within 10 m unavailable from the CSMP using CDFW GIS efforts to estimate area for hard bottom habitat and estimates of density from dive surveys with model-based methods.

Additional research and data needs identified in the desk review and comments from the authors.

- ***Consider power analysis to identify sample sizes required to achieve lower variances. (Dr. Williams)***
  This is a survey design consideration that should be addressed prior to future iterations of the survey now that preliminary data is available to provide variance estimates needed to conduct the power analysis. Previous surveys were conducted with other goals related to long-term monitoring of MPAs and changes may require additional funding that has not yet been identified. Pending the results of this review and application of developed methods, further analysis will be warranted.

- ***Have additional focus on flexible splines in GAMs for area expansions. (Dr. Williams)***
  The shape of relationships identified for gopher rockfish did not necessitate the application of flexible splines. The value of the splines in modeling non-linear relationships with depth and latitude or terrain attributes is recognized and can be implemented if necessary, to capture non-linear relationships in the future.

- ***Considerable variation in total abundance estimates stem from the different ways MBS derived covariates (substrata) can be used to expand transect fish density data to total fish abundance over broader areas. The expansions are particularly sensitive to the way in which the proportion of rocky bottom is defined and mapped, and the association of fish density to this metric. Ways to more objectively assess these methods should be explored – noting that the underlying methods of bottom type classification (rugosity proxy vs geology) for rasters differs between states. A stronger link between relief/rugosity and 'hard bottom' may help – possibly by capture of additional relief attributes in ROV imagery. (Dr. Williams)***
  In the design-based method, use of thresholds for cutoffs on the proportion of rocky reef in a given area in determining the amount of rocky reef to expand densities to provided results closer to those of the model based methods, which are better able to capture the relationships between habitat as defined by the CSMP and density. Further exploration of these thresholds seems worthwhile in the future. Additional stratification by terrain attributes and estimation of densities based on these variables applied in the design-based methods can be pursued, but it may be preferable to pursue a model-based framework if these additional complexities are to be incorporated. We did not explore utilization of additional geological classifications available from the ROV since they were not available in the CSMP to facilitate expansions. Alternative resolutions of terrain attributes may be evaluated in the future to determine if stronger correlations with greater predictive power can be identified, but this was beyond the scope of the current analysis and the methods described herein would be applied in a similar fashion to identify variables.

- ***Evaluate repeated transects in a spatio-temporal model to address temporal variability. (Dr. Williams)***
  Repeated sampling of the same transect would facilitate use of these models but large-scale sampling may be cost prohibitive. If data from additional samples are available, temporal models could be evaluated to answer specific questions to validate assumptions.

- ***SDM models using additional environmental variables may provide alternative ways to improve prediction success or cross-validate expansions; state-scale (transboundary) analyses through time could include variables describing oceanographic variability. (Dr. Williams)***
  Once the proposed methods have been fully refined and additional data is available, these methods may be useful in addressing questions that remain or arise in the process.

- ***Model validation should be one objective of future sampling programs – by sampling across the gradient of predicted densities. (Dr. Williams)***
  The repeated k-fold cross validation is a step in this direction, but additional sampling explicitly directed to sampling in a way that better allows direct validation of particular trends observed in the data and captured in modeling would be advantageous to provide ground truthing.

- ***This review has focused on spatial expansion of fish density data to large spatial scales; information packaged at the scale of individual reefs or reef clusters will also have utility in stock assessments. (Dr. Williams)***
  Stock assessments for groundfish have been focused on larger areas due to management complexity arising from regional management. The reef level analyses of interest to MPA monitoring and management are likely to continue to be pursued using the ROV data in the future.

- ***One possibility for future survey designs is to consider spatially-balanced methods to generate transect designs using randomization where the probability of sampling each cell in a spatial grid is user-defined (the cell inclusion probabilities). This is a robust and efficient method to assess ecological patterns. (Dr. Williams)***
  Further evaluation of sampling design should be considered in the future but will be balanced against the requirements for the original purpose for the sampling in monitoring MPAs. Should additional funding be identified to supplement the current sampling for these purposes, this design consideration should be considered.

- ***It would benefit the program to incorporate multibeam and/or side-scan sonar into the ROV systems. (Dr. Trembanis)***
  The use of multibeam and side-scan sonar may allow collection of additional data on the environment surrounding the ROV and response of fish to the ROV. This also comes along with cost, logistic burden, additional data storage and processing demands, which must be considered and weighed against data needs.

- ***Sampling bias associated with any survey gear can result from many factors, including noise, light, motion and pressure waves generated by the gear. Such biases should be considered for any and all gear used in the stock and habitat surveys. Gear disturbance can result in avoidance by some mobile species, leading to underestimates in density, or in the attraction of other species, resulting in an overestimation of densities. It should be stressed that the more we can make the underwater vehicles "fish like", or stealthier, the closer we will be to accurately reflect the relationships that exist between marine animals and their habitats. The need for studies of bias underline the necessity of creating calibration sites that could be surveyed by all gear types. (Dr. Trembanis)***

Further consideration of potential factors resulting in biases to detection probability were considered and discussed including how to operate the vehicle to minimize disturbances affecting behavior. As new technologies emerge to address potential sources of bias identified, they can be incorporated.

Use of ROV Methodologies to Improve Stock Assessment and Management Advice
The results can be used in stock assessments in the following ways:

1. **Density estimates as an index of relative abundance.** Our analysis primarily focused on identifying appropriate GLM distributions to address overdispersion in the fish count data and variables to include to normalize the index of relative abundance for use in modeling the density of various nearshore groundfish species since a time series of this information is not yet available. With continued sampling efforts, additional years of data will be available to provide a time series allowing for indices of relative abundance from this survey to be included in stock assessments.

2. **Estimates of abundance from habitat area expansions as an index of absolute abundance.** Abundance estimates can be used as an index of absolute abundance accounting for habitat area providing more representative indices than density estimates. As with indices of relative abundance from density estimates derived from a GLM, additional sampling will be required to represent changes in abundance over time from an index of absolute abundance.

3. **Estimates of abundance used to scale integrated assessments.** The design-based or model-based methods for estimating abundance can be used to inform the scale an assessment, which is often otherwise lacking and subject to considerable uncertainty when only catch and relative indices of relative abundance are available. Other data sources such as fishery dependent indices of relative abundance for which long time series are available to inform the trend in abundance. Estimates of abundance from the design-based and model-based methods presented here would help "peg" the scale of the abundance trend observed in the indices of relative abundance. If the design-based and model-based based abundance estimation methods are approved, they can be used to scale the assessment for gopher rockfish in 2019 as well as copper, brown and vermilion rockfish in 2021.

4. **Independent estimates of abundance multiplied by current F_{MSY} proxies to derive overfishing limits.** Application of the F_{MSY} proxies to the estimates of abundance from the design-based or model-based estimates of abundance can provide suitable category 2 or 3 estimates of OFL for use in management for stocks lacking OFLs. If surveys are conducted with sufficient frequency and have adequate spatial coverage to provide adequate stand-alone estimates of abundance they can form the basis for an estimate of OFLs for stocks with insufficient data to inform a full stock assessment or as an alternative method for estimating OFLs. With adequate sampling, finer scale stock assessments with these methods may also provide the potential for regional fishery management. This is otherwise prohibited by the decrease in data available for fully integrated stock assessments in any one area as the scale of the assessments decrease. If the design-based and model- based abundance estimation methods are approved, they could be used in combination with F_{MSY} proxies to provide alternative estimates of OFLs for gopher rockfish in 2019 as well as copper, brown and vermilion rockfish in 2021. If they are deemed category 3 assessment methods, then the merit of their use verses the current DB-SRA or DCAC based OFLs should be weighed in part of the basis of how recent the model-based or design-based estimates of abundance are and if they provide more representative estimates.

5. **Methods to inform allocation of nearshore rockfish annual catch limits.** The density mapping and estimates of abundance on a state-wide basis hold the potential to improve allocation of annual

catch limits across boundary lines and prevent localized depletion from disproportionate harvest relative to abundance. The estimates of abundance from properly stratified design-based methods or model based-estimates can be used to inform the proportion of the OFL within California that should be allocated north and south of 40°10' N. Lat or at other management boundaries.

Summary

Evaluation of correlations of density with variables explored with various GLM distributions indicate that the negative binomial distribution is the most appropriate model as indicated by the low AIC values in part due to its ability to address overdispersion. Depth, proportion hard bottom and latitude are consistently significantly correlated with density across species and model distributions, while correlations with take and terrain attributes were inconsistent though significant in some cases. The correlations identified may be useful in normalizing indices of relative abundance from density estimates if funding for continued sampling is available to produce a time series, and to inform poststratification in design-based estimation of abundance for other stocks.

The GAM model-based expansion method and design-based methods provided comparable results when hard bottom was used as the basis for habitat area, providing a validation of the results from model-based methods. Comparison of the areas encountered to the density projected by the model in density maps confirmed that the model-based method is providing results consistent with field observations. Analogous methods were used to assess cowcod in the Southern California Bight (Dick and MacCall 2013), lingcod and shelf rockfish stocks in Alaska (NPFMC 2013) and groundfish stocks in Puget Sound (Pacunski et al. 2016). Expansion of ROV surveys to the Rockfish Conservation Area would provide a non-lethal means of assessing stocks to provide much needed data on rocky reef dwelling species.

Given input from the methodology review panel, further development of the methods presented here has the potential to enhance stock assessment and management of nearshore and shelf groundfish that are well represented in the survey. Next steps in application to application of expansion methods would involve estimation of abundance for gopher rockfish for the remainder of the state for use in the 2019 assessment to provide information on the scale of the abundance state-wide. Further development of flexible spline-based GAM methods may be advantageous to more closely model non-linear relationships between density and correlated variables that may arise for other species.

Continued ROV sampling to develop a time series and expansion of the survey to bolster the proportion of sampled habitat in nearshore waters and sample in deeper depths in the rockfish conservation areas would enhance stock assessments and improve our knowledge of the abundance of fish residing in areas where fishing has been prohibited. The density estimates based on model predictions may also prove useful in application to MPA monitoring and management for which these surveys were designed. We hope that these methods and the review will enhance opportunities for additional funding for supplemental sampling to address sampling design needs relevant to fisheries management and lend synergistic support for continued MPA monitoring.

Acknowledgements

Thanks to Jason Roberts of Duke University for his input on the draft methods and model development using MGET. Also, thanks to Scott Marion of the Oregon Department of Fish and Wildlife for discussions regarding development of our respective methodologies. Our gratitude to Kathryn Meyer of the CDFW State Managed Species project for her assistance with R code used to split transects into 20 m segments. Thanks to Dr. Nick Perkins post doctorate researcher with the CDFW MPA project for his review of the draft document. Lastly thanks to Dayv Lowry of the Washington Department of Fish and

Wildlife for his input on our methods and considerations regarding expansion of estimates of density to provide abundance estimates.

References

Chakraborty A., A.E. Gelfand, A.M. Wilson, A.M. Latimer and J. A. Silander. 2011. Point pattern modelling for degraded presence-only data over large regions. Journal of the Royal Statistical Society. Appl. Statist. 60, Part 5, pp. 757-776.

Dick E.J. and A. MacCall. 2014. Status and Productivity of Cowcod, *Sebastes levis*, in the Southern California Bight. Pacific Fishery Management Council. Portland, OR.

Dormann C.F., McPherson J.M., Arau´jo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W. and W.D. Kissling, Ingolf Ku¨hn, Ralf Ohlemu¨ller, Pedro R. Peres-Neto, Bjo¨rn Reineking, Boris Schro¨der, Frank M. Schurr and Robert Wilson. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30: 609-628.

Dunn, K. P., and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 1-10.

Goldfinger C., Henkel S.K., Romsos C., Havron A., Black B. 2014. Benthic Habitat Characterization Offshore the Pacific Northwest Volume 1: Evaluation of Continental Shelf Geology, BOEM Award M10AC20002 (CFDA No.) 15.423, Corvallis.

Green K., J. Stahl, and M. Kallenberger 2013. 2013 Demersal shelf rockfish remotely operated vehicle survey. Alaska Department of Fish and Game, Regional Operational Plan ROP.CF.1J.2013.09, Anchorage.

Hedley S.L. and S.T. Buckland. 2004. Spatial models for line transect sampling. Journal of Agricultural, Biological and Environmental Statistics, 9, 181–199.

Karpov, K.A., A. Lauermann, M. Bergen, and M. Prall. 2006. Accuracy and precision of measurements of transect length and width made with a remotely operated vehicle. Mar. Tech. Soc. J. 40(3):79-85.

Karpov KA, Bergen M, Geibel JJ, Valle CF, Fox D. 2010. Prospective (a priori) power analysis for detecting changes in density when sampling with strip transects. Calif Fish Game 96: 69−81

Kline, Donna, Shrestha, June, and Starr, Rick. (2014) Comparison of lengths of North Coast MPA fishes estimated from an ROV using parallel lasers and stereo-video analyses. Unpublished report.

Laidig T.E. and M.M. Yoklavich. 2013. A comparison of density and length of Pacific groundfishes observed from 2 survey vehicles: a manned submersible and a remotely operated vehicle. Fishery Bulletin 114(4).

Laidig, Thomas E. and Krigsman, Lisa M. and Yoklavich, Mary M. (2013) Reactions of fishes to two underwater survey tools, a manned submersible and a remotely operated vehicle. Fishery Bulletin, 111(1), pp. 54-67.

Olson A., J. Stahl, K. Van Kirk, M. Jenicke and S. Meyer. 2013. Assessment of the Demersal Shelf Rockfish Stock Complexes in the Southeast Outside District of the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources for the Gulf of Alaska, North Pacific Fishery Management Council, Anchorage, Alaska pp. 555-608.

Pacunski R., D. Lowry, L. Hillier and J. Blaine. 2016. A comparison of groundfish species composition, abundance and density estimates derived from a scientific bottom-trawl and a small remotely-operated vehicle for trawl habitats. Washing Department of Fish and Wildlife Fish Program Science Division FPT 16-03. http://wdfw.wa.gov/publications/01824/wdfw01824.pdf.

Phillips S.J. and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161 – 175.

Roberts J.J., Best B.D., Dunn D.C., Treml E.A., Halpin P.N. 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. Environmental Modelling & Software 25: 1197-1207. doi: 10.1016/j.envsoft.2010.03.029

Saarman E., Liebowitz D., Serpa P., Raimondi P, Carr M. 2015. North Central Coast MPA Baseline Program Integration: Filling in the nearshore "White Zone". Report to California Ocean Science Trust. http://www.oceansciencetrust.org/wp-content/uploads/2018/12/North_Central_Coast_MPA_Baseline_Program_Integration_Filling_in_the_nearshore-%E2%80%9CWhite_Zone%E2%80%9D_UCSC_2015_NCCSotR.pdf

Stoner, A. W., Ryer, C. H., Parker, S. J., Auster, P. J., and Wakefield, W. W. (2008). Evaluating the role of fish behavior in surveys conducted with underwater vehicles. Canadian Journal of Fisheries and Aquatic Sciences 65, 1230-1243.

Sward, D., Monk, J. and Barrett, N. (2019). A Systematic Review of Remotely operated Vehicle Surveys for Visually Assessing Fish Assemblages. Frontiers in Marine Science, 6:134. doi: 10.3389/fmars.2019.00134

Young M and M.H. Carr. 2015. Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef species. Diversity and Distributions 1-13.

Appendix: Transect Summary and Site Locations

Table 1. Summaries from each cruise for number of fish transects, total fish count, and number of observed taxa. *Total number of taxa is a cumulative summary of each taxa observed from all cruises.

	Total no. of survey lines completed	No. of Transect cts 100m (Fish)	Fish counted	No. of Fish Taxa (Approx.)	Fish per km	% of total fish
Cruise A (South Coast)	99	141	18,812	41	300	2
Cruise B (South Coast)	155	384	403,459	51	4,768	51
Cruise C (North Coast)	115	552	34,203	39	472	4
Cruise D (North Central)	146	810	20,717	42	270	3
Cruise E (Central)	183	1,023	320,152	44	1,749	40
Totals:	**698**	**2,910**	**797,343**	**101***		

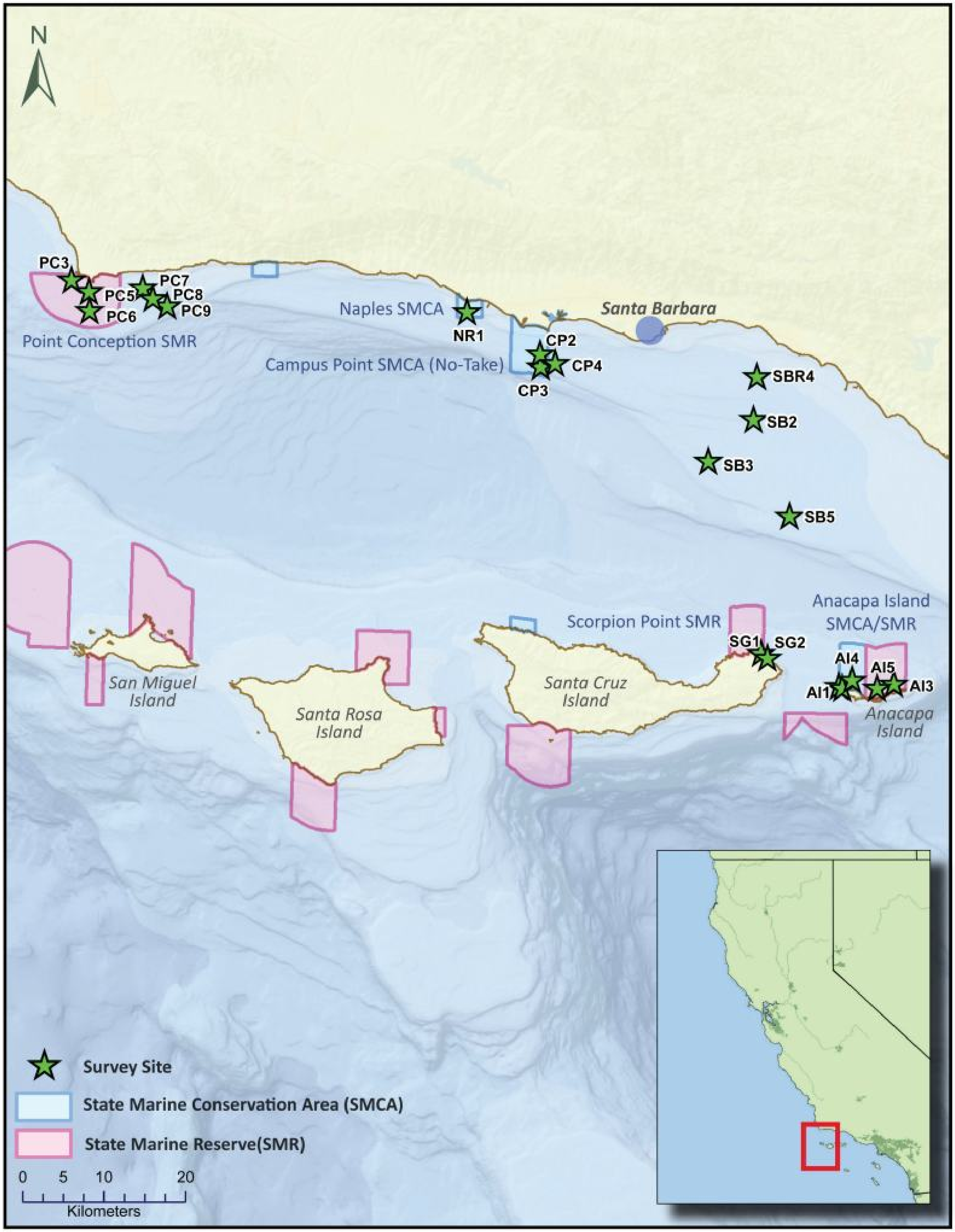

Figure 1. Site locations for CIAP cruise A in southern California, January 2014. Note that ten additional sites were sampled in the vicinity of the Channel Islands by MARE that are not represented here.

Figure 2. Site locations for CIAP cruise B in southern California, July 2014.

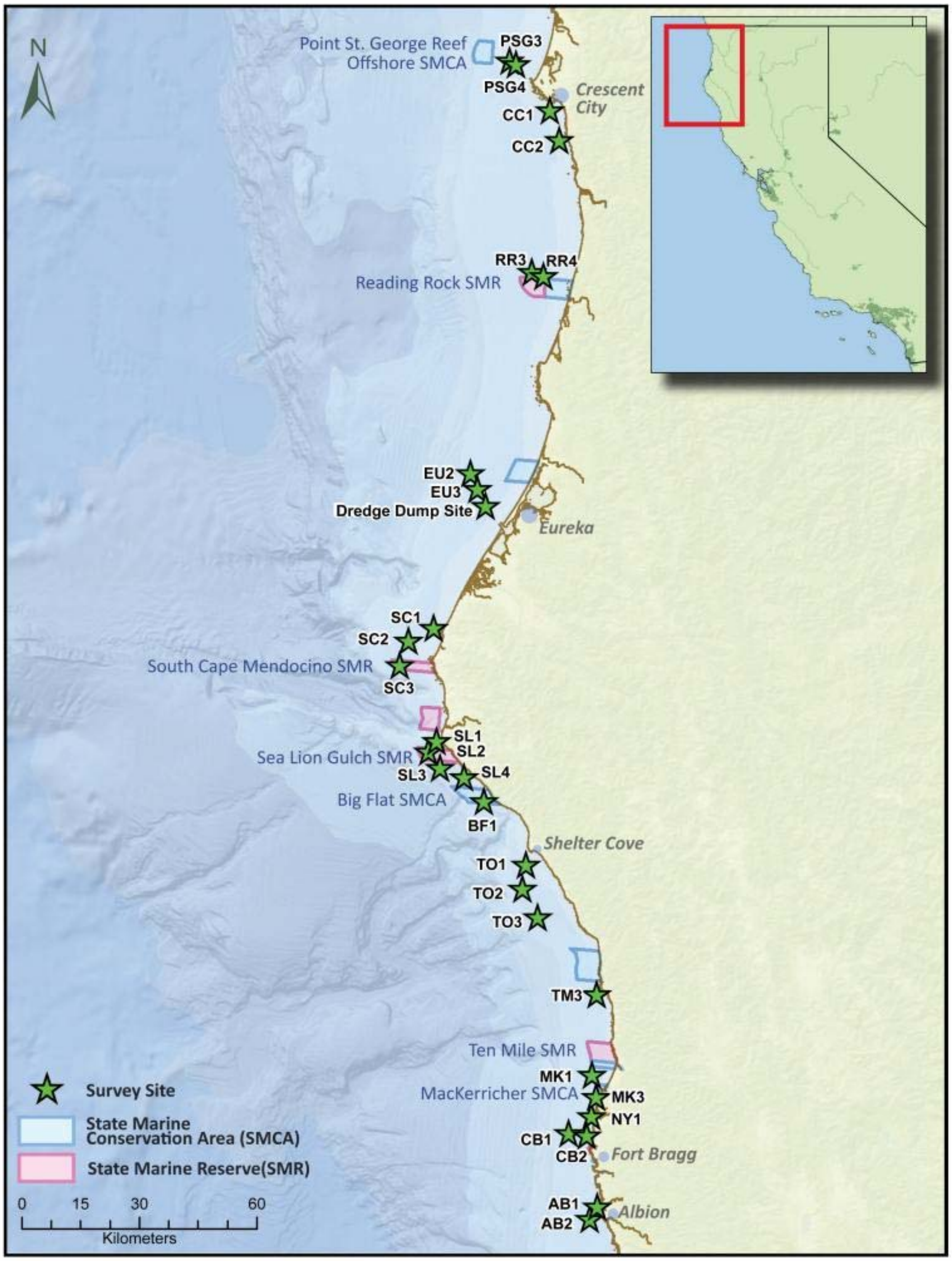

Figure 3. Site locations for CIAP cruise C in northern California, September-October 2014.

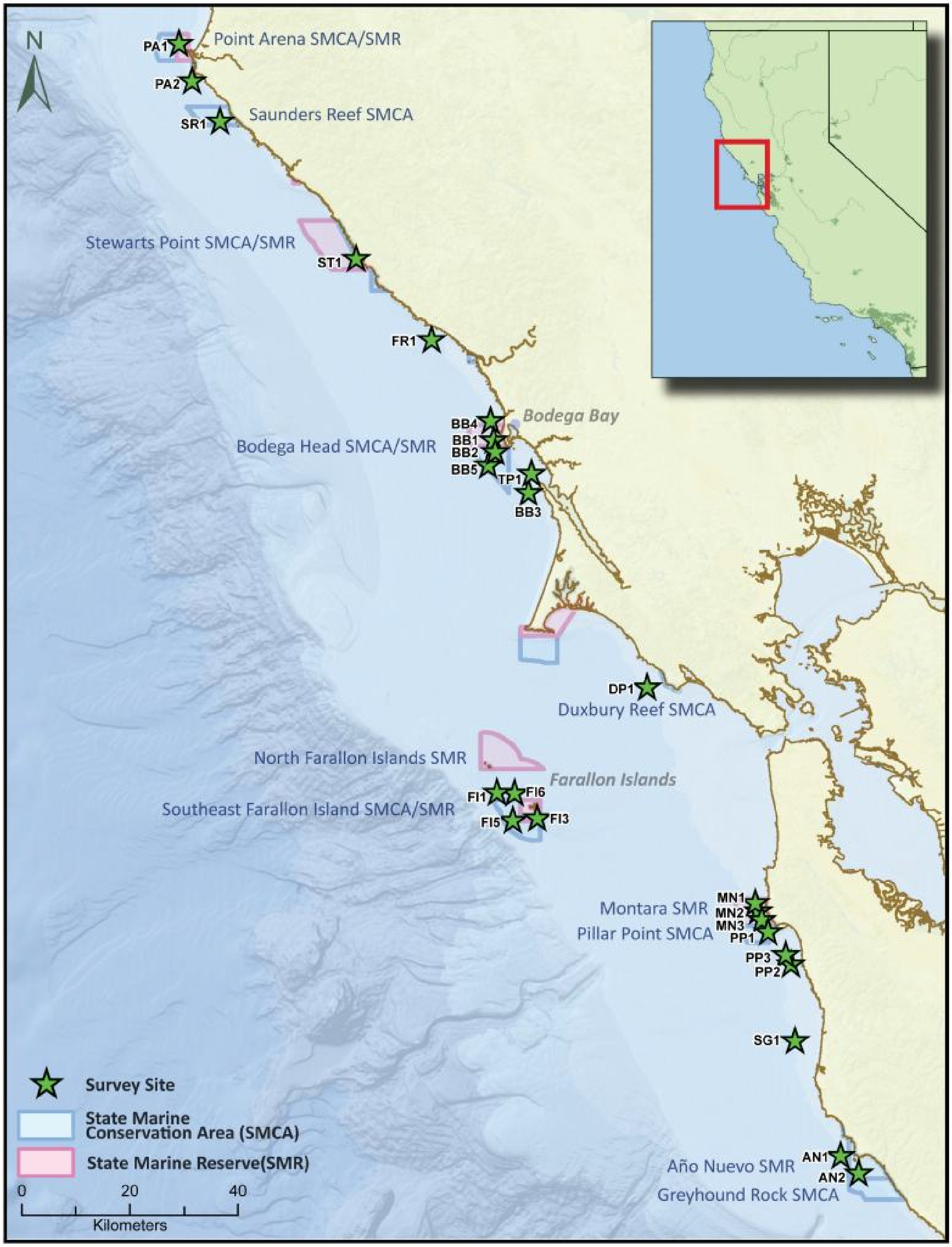

Figure 4. Site locations for CIAP cruise D in north central California, September-October 2015.

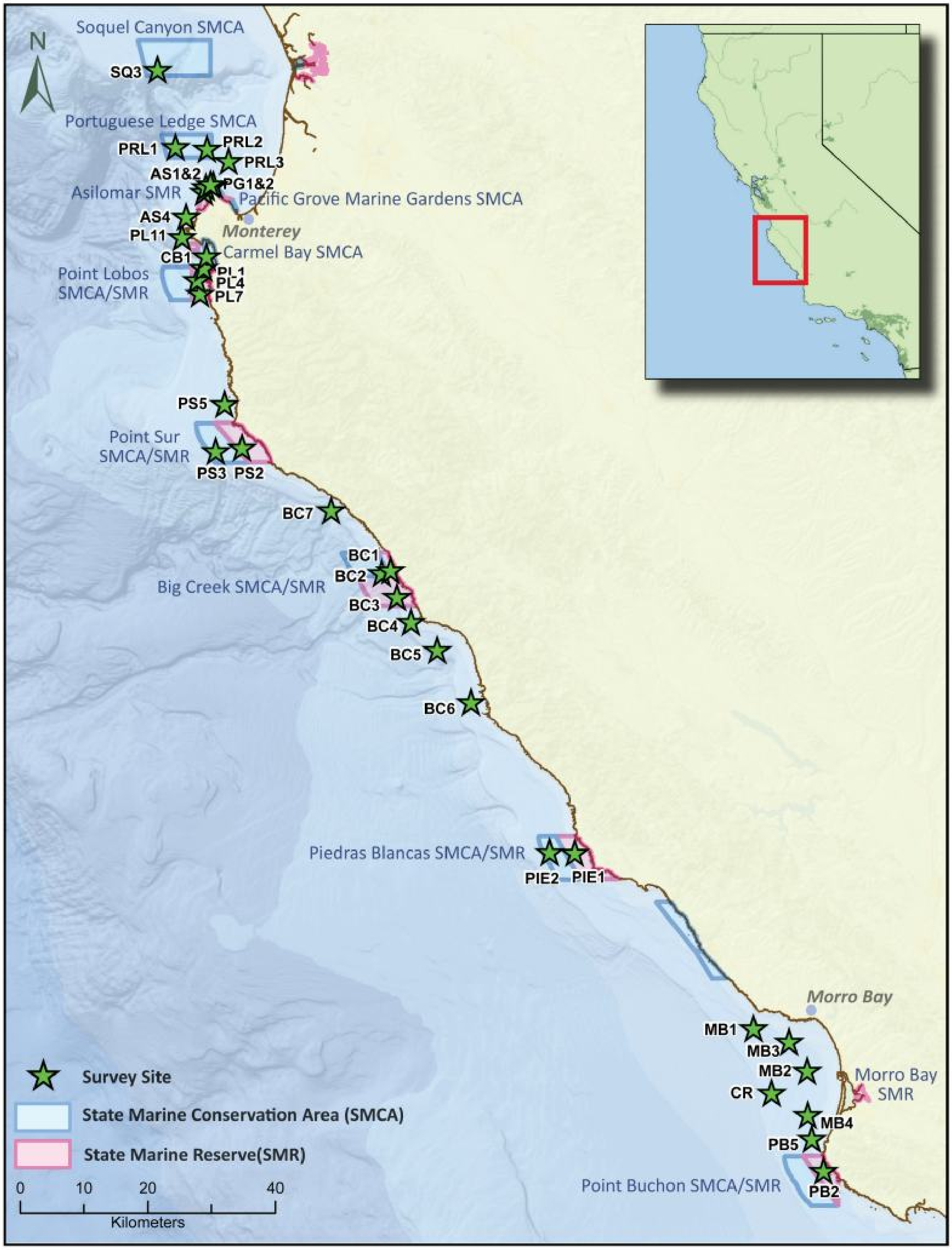

Figure 5. Site locations for CIAP cruise E in central California, September-October 2016.