

California Halibut 2020 Stock Assessment Review Panel Report

September 30, 2020

Review conducted via webinar in four sessions, held
June 1, June 9, June 15, and July 31, 2020

Participants

Review Panel Members

E.J. Dick, National Marine Fisheries Service, Southwest Fisheries Science Center (Chair)
Chantel Wetzel, National Marine Fisheries Service, Northwest Fisheries Science Center
David Sampson, Oregon State University
Kiva Oken, University of California, Davis

Stock Assessment Team (STAT)

Kathryn Meyer, State Managed Fisheries Program, Marine Region, California Department of
Fish and Wildlife; currently with Washington Department of Fish and Wildlife

Ocean Science Trust Representative

Anthony Rogers, California Ocean Science Trust

CDFW Advisor

Kirsten Ramey, State Managed Fisheries Program, Marine Region, California Department of
Fish and Wildlife

Overview

The stock assessment review panel (“Panel”) met via webinar on June 1, June 9, June 15, and July 31, 2020, to review a draft stock assessment of California halibut (*Paralichthys californicus*) in waters off California, U.S. This assessment was led and presented by Kathryn Meyer, formerly with the California Department of Fish and Wildlife (CDFW). Anthony Rogers with Ocean Science Trust (OST) and E.J. Dick with the National Marine Fisheries Service (NMFS; Review Panel Chair) welcomed participants, briefly reviewed the *Scientific and Technical Review Instructions* prepared by OST, and discussed logistics for the meeting. Kirsten Ramey with CDFW summarized the Department’s goals and purpose for conducting this assessment. Rapporteur duties during each webinar were shared among members of the Panel and OST representatives.

The draft (pre-review) assessment document and background material (previous assessments, previous Review Panel report, etc.) were provided electronically to the Panel two weeks in advance of the Panel meeting. CDFW, OST, and the Panel chair agreed in advance that the Pacific Fishery Management Council’s (PFMC) “Terms of Reference for the Groundfish and Coastal Pelagic Species Stock Assessment Review Process for 2019-2020” would be used as a general guide for content of the assessment document, acknowledging that management targets, harvest control rules, and other requirements specific to the PFMC process would not apply. Google Drive was used for common access to agendas, presentation materials and model runs that were conducted over the course of the review.

The California halibut (“halibut”) stock assessment was conducted using Stock Synthesis (SS; version 3.30.14.08). The population was modeled as two, independent stocks. A southern stock was defined from the U.S.-Mexico border to Point Conception (Santa Barbara County). A northern stock was defined as the region between Point Conception and Point Arena (Mendocino County). The STAT recognizes that this assumption is likely violated due to some degree of connectivity between the southern and northern stocks, as well as between the southern stock and halibut south of the U.S.-Mexico border. However, the Panel considered the use of two models to be a reasonable compromise given regional differences in exploitation history, management, growth, maturation, and data availability.

The population dynamics in the southern region were modeled from 1971-2019. The northern model began in 1980 due to a lack of composition or survey data prior to that time. The halibut fishery began operating as early as the late 19th century in some areas, so both models attempted to estimate initial equilibrium fishing mortality rates conditioned on assumed equilibrium catch levels. This approach was similar to the previous (2011) assessment and was motivated in part by uncertainty in fleet-specific exploitation patterns prior to the modeled periods. However, attempts to estimate the initial fishing mortality parameters were ultimately only successful for the southern model, and the parameters remained fixed in the northern model.

The difficulties with estimation of initial conditions motivated many Panel discussions about the benefits of exploring alternative model structures in future assessments. Possibilities include 1) reconstructing historical catches back to an unfished or nearly unfished condition, and 2) starting the model from a depleted level by using a stock-recruitment regime parameter (an offset from equilibrium recruitment), without attempting to fit equilibrium historic catches. Recommendations specific to each area are given below and in the section titled “Recommendations for Future Research and Data Collection.”

Other topics covered at length during the review were the estimation of growth parameters and the treatment of discards. Some growth parameters were estimated in the northern model by adding age composition data (conditioned on length) from CDFW research cruises. Growth parameters in the southern model were estimated outside the model using the most recent available data and input as fixed values in the model. Discard data in both models were revised (fitted as rates) and retention curves were estimated when possible. Discard mortality rates were also updated based on re-examination of haul duration distributions in the trawl fleet.

The Panel does not consider the northern area base model for halibut to be adequate for use in management, as it was presented during the final review webinar. This conclusion is not intended to reflect poorly on the STAT, who were extremely responsive to the Panel’s questions and requests throughout the review. Rather, it is based primarily on four issues identified with the northern base model. First, the parameters defining initial conditions (initial equilibrium fishing mortality rates, or “initial Fs”) could not be estimated and attempts to diagnose the problem produced contradictory results. When the initial Fs were estimated, they consistently hit upper boundaries defined in the model. However, likelihood profiles over the same parameters showed that the negative log likelihood was minimized at the smallest mortality rates. Fixing initial Fs in the northern model, combined with the assumed equilibrium catches, effectively pre-determines the size of the population in the starting year of the model. Second, the choice of whether or not to estimate recruitment deviations prior to the start year greatly affected initial stock status (spawning output relative to unfished spawning output). Third, stock status in the terminal year was extremely sensitive to the estimation of additive variance parameters for abundance indices, suggesting a conflict between the indices and other data types. Lastly, data weights were not applied to age composition data and the method used to weight the length compositions was not consistent with current accepted practices. In summary, the northern base model made strong assumptions about the scale of the population and could not resolve the relative stock status in either the initial or terminal years.

The Panel recommends that future efforts to model the stock north of Point Conception focus first on catch reconstruction (e.g., back to the early 1900s) or estimation of initial conditions. As described by the STAT during the first webinar, catches in the north increased more gradually than in the south where the stock was believed to be in a depleted state when catch recording began (ca. 1916). As a result, the northern model may benefit more, in terms of stability, from a catch reconstruction effort. Estimated selectivity curves for gear types in the early fishery were

very similar, so obtaining reasonable estimates of total annual catch (across fleets) is more important than describing allocation among fleets. A model with reconstructed catches would not require estimation of initial conditions and may improve advice for the northern stock.

Alternatively, as noted above, estimation of an offset parameter for recruitment in the starting year is another model structure that is worth exploring for the northern region. A benefit of this approach is that it does not require historical catch reconstruction. However, it remains unclear whether estimation of an offset parameter for recruitment in the start year will be possible given the available data.

Many of the issues identified in the northern model were not problems in the southern assessment. The STAT was able to estimate initial conditions (initial F_s) for three of the four fishing fleets in the south. Initial F was fixed for the commercial hook-and-line fleet, but this was a minor component of the historical catch and had little influence on the assessment results. Also, the choice of whether or not to estimate early recruitment deviations had little effect on southern model results. Similarly, additive variance parameters did not influence terminal stock status. In these ways, the southern model showed greater stability across alternative model structures.

However, the Panel identified some technical issues in the current southern base model, as of the final review webinar held on July 31, 2020, and recommends further investigation into these topics prior to using the model to inform management. The three highest priority topics were data weighting, treatment of the CalCOFI index, and calculation of initial equilibrium catches.

Data weights in the southern model were not estimated for age composition data, i.e., all weights were fixed at a value of 1.0. Weights for length composition data were estimated using the Francis (2011) method, but were only adjusted in the base model when the confidence interval around the point estimate did not contain a value of 1.0. Francis weights affect the relative influence of data types in the model, and are estimated to prevent the model from over-fitting to composition data as a result of the large numbers of composition observations tending to dominate other information sources. The base model should be updated to include data weights for age composition data, and to use the point estimates for all composition data – lengths and ages -- rather than the confidence interval approach.

The southern base model uses data from the CalCOFI ichthyoplankton survey to create a relative index of spawning output. Since a large fraction of tows did not catch larval halibut, the data were aggregated into 3-year “super years,” and standardized to account for seasonal and spatial (station) effects. Uncertainty in the index (log scale standard error for each super year) is included as an input to the assessment model. For this index, the Panel noted that estimates of uncertainty were extremely large (i.e., an average of 2.4), and this is likely an error. Large values for the input standard errors cause the model to effectively ignore observed trends in the index. CalCOFI ichthyoplankton data are available since 1951 in southern California. The Panel suggests that the southern model start in 1951 (or 1952 if 1951 is included in a 3-year “super year”). This would allow the CalCOFI index to inform changes in spawning output.

It was unclear whether or not estimates of equilibrium catch included both landed (retained) and discarded catch, as they should for use in Stock Synthesis. Also, equilibrium catches for the commercial hook-and-line fleet were identical to the two recreational fleets, despite being in different units (mt for commercial versus 1000s of fish for recreational). A sensitivity analysis to alternative assumptions about equilibrium catch (halving and doubling the preferred values) dramatically changed estimates of relative abundance at the beginning of the modeled time period in the southern model. Relative abundance in recent years was stable. However, the input equilibrium catch was the largest component in likelihood profiles over R0, suggesting that estimates of population scale were driven by the choice of equilibrium catches in the southern model. The panel was not shown the effect of the equilibrium catch sensitivity analysis on absolute abundance. Given the importance of equilibrium catches in determining population scale and early year depletion, the Panel requested that the final assessment document include a description of the data sources and methods used to calculate the initial equilibrium catches by fleet.

The southern model runs presented to the Panel during the review webinars were consistent in terms of relative spawning output in recent years. The Panel cannot anticipate whether the above-mentioned revisions will change that result. If estimates of stock status are unaffected by the Panel's suggestions, that would suggest the stock has generally been fluctuating below B_{MSY} (the stock size that produces maximum yield) since the late 1970s. However, there is no evidence – based on the current southern base model – of persistent declines in stock status as a result of recent catches. Uncertainty in stock status and other model outputs is larger than the reported intervals because important parameters are fixed (steepness, natural mortality, and growth).

The Panel recognized other aspects of the southern model that could benefit from further investigation. These analyses are recommended for future assessment efforts, but are not as pressing as the three issues described above. Similar to the northern model, estimation of an offset parameter for recruitment in the starting year is an alternative way to model initial conditions for the southern area. This approach does not require reconstruction of catches or fitting to initial equilibrium catch estimates, and the Panel suggests exploring this option in future assessments. Alternatively, reconstructed historical catches could be used to model the population from an unfished or nearly unfished condition. Catch reconstruction may be more challenging for the southern region due to the extended history and high exploitation rates that pre-date catch monitoring efforts in the state.

Although considerable age and length data exist for the southern region, attempts to estimate growth in the southern model were not successful. Parameters describing mean length at age and uncertainty in length at age affect how the model estimates productivity of the stock (future yields) and historical exploitation rates (stock status). Estimation of growth within the model is generally preferred because internal estimates account for the effects of size (or age) selection on observed length at age data, and uncertainty in growth is propagated into model outputs.

Neither assessment model (north or south) attempted to forecast future population dynamics or yield. Estimated annual recruitments in the current southern base model were below average from roughly 2008-2015. This result was consistent across a range of steepness values from 0.6 to 1.0. Recruitment deviations since 2016 were not estimated in either model, and therefore the models assume average recruitment in recent years. If recruitments in recent years are also below average, then the estimated increases in recent stock status may not be realized and near-term stock productivity may be reduced.

The Panel greatly appreciated the documentation, presentations and analyses prepared by Kathryn Meyer, who was effectively a “STAT of one.” Given the complexity of the halibut assessment, which required two assessment models, future review panels would benefit from having more than one analyst assigned to the STAT. For example, if managers of the primary data sources could assist with data-related questions, then requests that focus on modeling could be the responsibility of the lead stock assessor. Few assessments of this magnitude are completed by a single individual. As this was Ms. Meyer’s first full stock assessment, the Panel commends her ability to generate informative responses to a large number of requests, clearly communicate the results, and do so in a timely manner.

The Panel suggests that the stock assessment documentation be updated to reflect all changes in the base models over the course of the review. The document should be archived with all data sources and Stock Synthesis files to assist authors of future stock assessments for halibut.

Summary of Data and Assessment Models

The following descriptions are based on models presented during the fourth and final review panel webinar, held on July 31, 2020. Some of the results from these post-review models were quite different from the pre-review models.

Southern Model

The post-review model for waters south of Point Conception included three commercial (trawl, gillnet, and hook-and-line) and two recreational (CPFV and ‘other recreational’) fishing fleets. Research surveys included an ichthyoplankton index of spawning output (CalCOFI) and a survey “fleet” that allowed a separate selectivity function to be applied to age data collected during CDFW research cruises.

Commercial and recreational landings from 1971-2018 were provided by CDFW. Initial equilibrium catches were set equal to roughly 133 mt per year for the trawl and gillnet fisheries (each) and 14 mt for the commercial hook-and-line fishery. Equilibrium annual catch for the recreational fleets was in numbers of fish (rather than metric tons) and set equal to roughly 14,000 fish for both the CPFV and ‘other recreational’ fleets. The Panel noted that the equilibrium catch value for the commercial hook-and-line fleet (in units of mt) exactly matched the equilibrium catch for the recreational fleets (in numbers of fish). Discarded catches were modeled by fitting to discard rates from the NMFS West Coast Groundfish Observer Program

(WCGOP), two gillnet observer programs (CDFW and NMFS/SWFSC), and CPFV logbook records. Standard error for all rates was set equal to 0.6. Discard length compositional data, to inform the length-based retention curves, were available only for the trawl, gillnet and CPFV fleets. Discarded catch was modeled in fleets without discard rate data by mirroring parameter values in similar gear types or with values estimated external to the model.

Indices of relative abundance included two fishery-dependent sources: a trawl logbook index (1998-2019) and a CPFV logbook index based on inshore statistical blocks (1981-2019). The fishery-independent CalCOFI survey data, which was used in the model as an index of spawning output, contained a large proportion of zeros (no halibut caught), so the standardized index grouped observations into 3-year blocks over the period 1975-2017. Units of biomass were specified for trawl logbook, numbers for CPFV, and eggs for CalCOFI. Additive variance parameters were estimated for all indices.

Length composition data were included for trawl gears (combined), gillnet, commercial hook-and-line, CPFV, and ‘other recreational’ fleets. Ages were included as conditional age-at-length data although growth parameters were not estimated in the southern model. Age data were available from the trawl and gillnet fleets, as well as from CDFW research cruises (combined into a single ‘fleet’). A single ageing error matrix was applied to all fleets with a constant CV of 20%.

The southern model was a 2-sex model with dimorphic growth estimated external to the model. Available data suggest that females grow larger and more slowly than males. The natural mortality rate was assumed to be different for females ($M=0.18$) and males ($M=0.235$) and fixed at the median of the prior distribution using $M=5.4/A_{\max}$ (Hamel 2015), with a log-scale standard deviation of 0.438. Observed maximum ages were 30 and 23 years for females and males, respectively, based on statewide data. Steepness of the Beverton-Holt spawner-recruit relationship was fixed at 0.9 and recruitment deviations were estimated starting in 1951 (prior to the first year with data) through 2018. Variability in log-scale recruitment was defined as a normal distribution with a standard deviation (“sigma-r”) of 0.6. The recruitment deviations for the main period had a standard deviation of 0.675.

Initial equilibrium fishing mortality parameters were estimated for four of the five fishing fleets, but fixed for the commercial hook-and-line fleet at $F_{\text{init}}=0.0029$. This small value reflects the fact that landings by this gear type were small prior to the modeled time period and model results were not sensitive to a range of plausible fixed values.

Selectivity in Stock Synthesis can be defined as a function of length, age, or both. In the southern model, it was primarily length-based and asymptotic for all fishing fleets, except the gillnet fishery which had a domed length-based selectivity. Many selectivity parameters were fixed, and retention and discard mortality options were enabled. Selectivity for the gillnet fleet was time blocked, allowed to change in 1996 to reflect management actions. Selectivity was assumed constant over time for all other fleets. Age-based selectivity was set equal to 0 for age-0 fish, with exception of the combined age data from CDFW research cruises, which had age-0

selectivity equal to 1. Age-based selectivity was set equal to 1 for all other ages. The CalCOFI ichthyoplankton survey had selectivity set equal to egg production.

Data weights were applied following Francis (2011) for length compositions but only changed from a value of 1.0 for data from a given fleet when confidence intervals for the weight estimate for the fleet did not include 1.0. This resulted in Francis weights being applied to data from some fleets and not others, and was not consistent with current practice in other assessments of demersal species on the U.S. West Coast. The panel recommended using the point estimates from the weighting procedure for all fleets instead, which would likely reduce the weights applied to some of the composition data. Weights for age composition data were fixed at 1.0, and the panel recommended applying the same weighting procedure to ages as was recommended for lengths.

Northern Model

Note: The northern stock was sometimes referred to as the “central” stock during the review and in some of the documentation. These two names referred to the same model for waters off California between Point Conception and Point Arena.

The post-review model for waters between Point Conception and Point Arena included three commercial (trawl, gillnet, and hook-and-line) and two recreational (CPFV and ‘other recreational’) fishing fleets. Research surveys included a trawl survey of age 0-1 fish in San Francisco Bay and a survey “fleet” that allowed a separate selectivity function to be applied to age data collected during CDFW research cruises.

Commercial and recreational landings from 1980-2018 were provided by CDFW. Initial equilibrium catches were set equal to roughly 25 mt per year for the trawl fleet, 31 mt for the gillnet fleet, and 3 mt for the commercial hook-and-line fishery. Equilibrium annual catch for the recreational fleets was in numbers of fish (rather than metric tons) and set equal to roughly 2,100 fish per year for the CPFV fleet and 2,600 fish per year for the ‘other Rec’ fleet. Discarded catches were modeled by fitting to discard rates from the NMFS WCGOP, data from two gillnet observer programs conducted in Southern California (CDFW and NMFS/SWFSC), and CPFV logbook records. Standard error for all rates was set equal to 0.6. Discard length compositional data, to inform the length-based retention curves, were available only for the trawl fleet.

Discarded catch was modeled in fleets without discard rate data by either borrowing data from the Southern California model (gillnet only, 1983-1994) or mirroring parameter values in similar gear types or with values estimated external to the model.

Indices of relative abundance included two fishery-dependent sources: a trawl logbook index (1998-2019) and a CPFV logbook index based on inshore statistical blocks (1981-2019). The fishery-independent San Francisco Bay Study conducted benthic trawl surveys to document ecological impacts associated with changes to the Sacramento/San Joaquin Delta outflow. Data from this survey were included for the years 1981-2014. Data from 2015 were not used due to mechanical issues with the survey vessel, and the survey was terminated that same year. Units of

biomass were specified for the trawl logbook index, and numbers for the CPFV and San Francisco Bay Study indices. Additive variance parameters were estimated for all indices.

Length composition data were included for trawl gears (combined), commercial hook-and-line, CPFV and ‘other recreational’ fleets, and the San Francisco Bay Study trawl survey. Age data from the trawl and gillnet fleets, as well as from CDFW research cruises (combined into a single ‘fleet’) were included using a conditional age-at length format. A single ageing error matrix was applied to all fleets with a constant CV of 20%.

The northern model was a 2-sex model with dimorphic growth. Unlike the southern model, which fixed all growth parameters, female size at age 15 and the von Bertalanffy growth rate parameter (‘k’) were estimated by the model. Other growth parameters remained fixed, including size at age 0, CVs of length-at-age, and all male growth parameters (specified as exponential offsets to the female parameter values). These fixed parameters were derived from an external analysis of length and age observations. Available data suggest that females grow larger and more slowly than males, and this was reflected in the fixed male offset parameters. The natural mortality rate was assumed to be different for females ($M=0.18$) and males ($M=0.235$) and fixed at the median of the prior distribution using $M=5.4/A_{\max}$ (Hamel 2015), with a log-scale standard deviation of 0.438. Observed maximum ages were 30 and 23 years for females and males, respectively, based on statewide data. Steepness of the Beverton-Holt spawner-recruit relationship was fixed at 0.9 and recruitment deviations were estimated starting in 1960 (prior to the first year with data) through 2018. Variability in log-scale recruitment was defined as a normal distribution with a standard deviation (“sigma-r”) of 0.6. The recruitment deviations for the main period had a standard deviation of 0.887.

Initial equilibrium fishing mortality rate parameters were fixed for all five fishing fleets in the northern model. Fixing these parameters and specifying the initial equilibrium catches places a very strong constraint on stock size in the start year (1980). Several of the Panel’s requests were related to allowing free estimation of these parameters, in both models, as described in detail below.

Selectivity in the northern model was length-based and dome-shaped for the gillnet and both recreational fishing fleets (although many parameters were fixed). The combined trawl and hook-and-line fleets had asymptotic length-based selectivity, and retention and discard mortality options were enabled for all fleets. Selectivity was assumed constant over time for all fleets. Length-based selectivity for the San Francisco Bay Study trawl survey was assumed to follow a descending logistic pattern (with fixed parameters) due to the small size range of fish caught by the survey. Age-based selectivity was set equal to 0 for age-0 fish, with exception of the San Francisco Bay Study trawl survey and the combined age data from CDFW research cruises, which set age-0 selectivity equal to 1.0. Age-based selectivity for all other ages was set equal to 1.

Data weighting followed the method of Francis (2011) for length compositions but only changed from a value of 1.0 when confidence intervals for the weight estimate did not include 1.0,

resulting in weights being updated for some fleets but not others. This was not consistent with current practice in other assessments of demersal species on the U.S. West Coast. The panel recommended using the point estimates from the weighting procedure instead, which would likely reduce the weights applied to some of the composition data. Weights for age composition data were fixed at 1.0, and the panel recommended applying the same weighting procedure to ages as was recommended for lengths.

Requests by the Review Panel and Responses by the STAT

During the first webinar, the Panel noted that the pre-review base models used fixed parameters to describe the initial equilibrium fishing mortality rates (F) and treated annual estimates of discard by fleet as data. Fixing initial F s were seen as problematic for two reasons. First, fixed initial F s strongly inform estimates of population scale when initial equilibrium catches are also fixed (as they were in both models). Second, the models' estimates of equilibrium catches, given the fixed F values, were not consistent with the input equilibrium catches.

The Panel also noted that the specification of discards in the model was incorrect. External estimates of annual discarded catch by fleet were input as data, as opposed to fitting the models to the discard rate data, which were only available for a limited number of years. In the Panel's first set of requests, changing the treatment of discards was a recommended first step (Request 1, below). Requests 2-12 were intended to help diagnose why initial fishing mortality rates (F) were not estimable in either model and why estimated equilibrium catches differed so significantly from the input values. Requests 13-21 were designed to help the Panel better understand choices related to the treatment of various data inputs. Requests 1-21, which were developed during the first webinar, were all with regard to the versions of the model that were provided prior to the first webinar.

During the second webinar the Panel was shown that considerable progress had been made on the topics of growth estimation and treatment of discards. However, some additional requests related to growth and discards remained. Estimation of initial conditions for both models continued to be the Panel's largest concern. It was noted during the meeting that models with initial equilibrium catches are constrained by the fact that total equilibrium catch must be less than estimated Maximum Sustainable Yield (MSY). It was unclear whether this constraint affected the models' ability to estimate initial conditions.

A second set of requests (#22-30, below) continued to explore estimation of initial conditions, specifically, estimation of initial equilibrium fishing mortality rates. Estimation of growth parameters, within the model, was also explored after including age-at-length data from CDFW research cruises. The Panel clarified that responses to each request should include changes in likelihood by data type, as well as time series of spawning output, relative spawning output, and

recruitment deviations. Any large changes in parameter estimates and/or lack of convergence should also be noted.

During the third webinar, Dr. Wetzel noticed that the likelihood component for initial equilibrium catch had a multiplier (“lambda”) that was inadvertently set equal to zero. As a result, the models were not attempting to fit to the input values of initial equilibrium catch. It was agreed that a fourth webinar would be added to the review, and the Panel requested that the STAT prepare the following materials for the fourth and final webinar:

1. Complete sets of r4ss output and SS input/output files, including fits to all data sources.
2. A list of fixed and estimated parameter values (with standard errors for the latter).
3. Likelihood profiles over R_0 , h , and M .
4. A table of data weights (iterated from a starting value of 1.0).
5. Convergence checks (no parameters on bounds, jitter starting values to ensure model convergence, hessian inversion, gradient information, parameter correlations, etc.).
6. A sensitivity analysis to alternative initial equilibrium catches (changing total magnitude, but with the same relative allocation among fleets).
7. All previous requests addressed in some sense (e.g., an explanation for why the request was not addressed).

In the event that either model did not estimate initial F 's, the Panel asked to see a likelihood profile over the initial F s, with $\lambda = 1$ so the equilibrium catch likelihood would contribute to the total likelihood.

Since a few topics (estimation of initial conditions, treatment of discards, and growth parameter estimation) were high priorities to the Panel, requests related to these topics were identified below with bracketed text after the request number. Notation used for the Schnute (1981) parameterization of von Bertalanffy growth was inconsistent during the Panel meetings. This report uses “L1” and “L2” to refer to mean length at the younger and older reference ages, respectively. Also, “CV1” and “CV2” refer to the coefficients of variation of length at the younger and older reference ages.

Meeting 1 Requests

Request 1: [Discards] Enter discard information as rates (option #2) into SS, i.e., discard / (discard + retained), and remove the external estimates of discard. When setting up retention curves, order the fleets such that fleets with discard size data have lower numbers than fleets without discard size data (to facilitate mirroring). Define retention curves for fleets without size data by mirroring the retention curve of the most similar fleet (e.g., commercial hook and line retention mirrors rec CPFV). Specifically, consider treating retention as follows:

Southern Model: Estimate retention for the trawl and CPFV fleets using available length-composition data. Mirror other fleets as needed to define retention curves.

Northern Model: Estimate retention for the trawl fleet using available length composition data; fix retention curves for the other fleets at values used in the southern model.

If the fits are still not well visualized given the r4ss plots, provide a summary of the fits to the discard data across time from the DISCARD_OUTPUT section of the Report file in order to evaluate whether the model is fitting these data on average.

Rationale: The draft base models used discarded catches that were estimated externally and entered as data for the entire duration of each model (Figure 1). The proposed approach avoids entering “data” for years in which there are no observations. Report changes in relative magnitude of discard likelihood using this approach compared to the draft base model.

Response:

Southern Model: Changing from estimated discards to discard rates allowed estimation of retention curves where data were available. The change did not resolve the discrepancy between estimated and ‘observed’ values. Model outputs prior to 1980 were very sensitive to changes in discard specification. The likelihood component for discards decreased relative to the previous base model. Observed and expected values were only different for the recreational fleets, but this was due to the fact that recreational catch is presented in numbers, and is a default setting in the post-processing software (r4ss) that needs to be corrected. The fit to length composition data degraded (based on likelihood) relative to the previous base. It was necessary to fix the peak selectivity parameter for CPFV because the discard change caused it to drift toward unrealistic values. Retention curves were estimated for bottom trawl (WCGOP lengths), gillnet (SWFSC survey lengths), and CPFV (which informs all hook-and-line fleets).

Northern Model: This model run added discard rates from the southern model into the northern model as data. As presented, there were no major changes to time series, and the discard likelihood component decreased relative to previous base model.

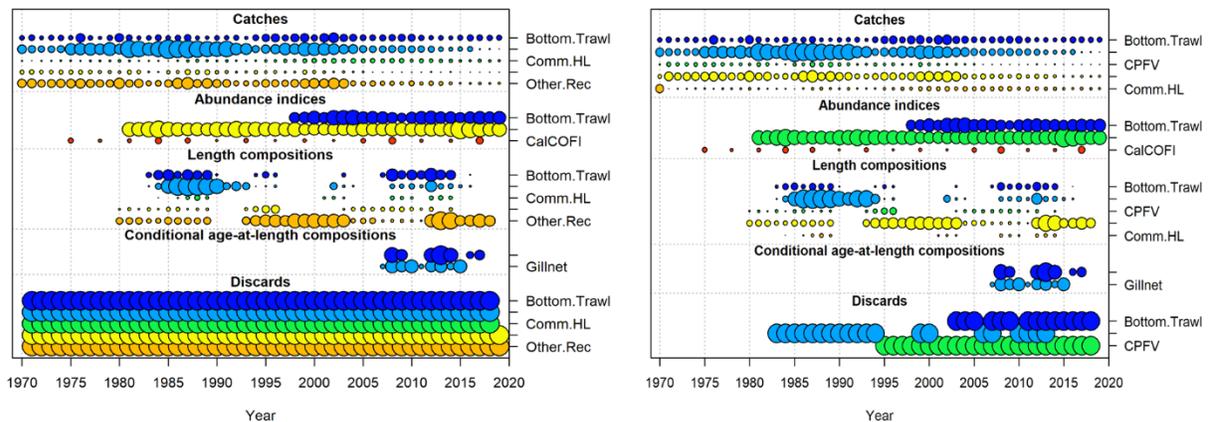


Figure 1. Data summary plots illustrating the change in discard data configuration between the pre-review base model (left) and the revised configuration using discard rates. Example shown is for the southern model.

Request 2: Develop northern and southern models with the following features. Fix natural mortality (M) for both sexes at the median of the prior distribution. Fix growth parameters at the externally estimated values for each region. Fix initial fishing mortality rates (F) at the same values used in the draft base models. Force the selection curve for the commercial hook & line fleet (fleet 3) to be asymptotic.

Rationale: The current models have fixed growth and the estimated M values are fairly far from the median of the prior. The requested runs will use the externally estimated growth parameters from the draft base models with fixed M, and serve as baselines for comparison with subsequent runs in which growth is internally estimated and/or initial conditions are estimated. Also, the current central area model uses all dome-shaped selectivity curves, which could make estimation of natural mortality unreliable.

Response: Results for request #2 built upon request #1. Fixing M, growth, and initial F's is similar to the previous base model, apart from having fixed M and one asymptotic selectivity curve.

Southern Model: Several parameters were flagged as having high gradients. The stock was less depleted in the terminal year, but the relative biomass time series was more variable. General pattern in biomass time series is similar, but the period prior to 1980 is sensitive to changes in the fixed parameters.

Northern Model: Lower stock status, likely due to fixing M at a lower value than what was estimated in the previous base model.

Request 3: [Growth] Using the fixed-M models from Request #2, parameterize male growth as offsets from female parameter values (e.g., M, k, L1, L2, CV1, CV2, example parameterization will be $M_{\text{male}} = M_{\text{female}} * e^{M_{\text{male_offset}}}$ and estimate all growth parameters internally for both models [L1, L2, k, CV1, and CV2].

Rationale: Parameterizing male growth as offsets from female parameters may improve the model's ability to estimate sex-specific growth curves. Estimation of growth should, in principle, be more reliable than estimation of natural mortality.

Response:

Southern Model: Produced highly unrealistic estimates of growth parameters for both sexes (e.g., $k = 0.032$, $L2 = 133$). External estimates use age/length data that are not included in the model as conditional-age-at-length (CAAL) data (because they were not associated with any of the fishing or survey fleets). Investigate ways to include samples that are not associated with the fishery to better inform growth.

Northern Model: Estimates for northern CA were much more reasonable, but CVs were very large.

Request 4: [Growth] Using the models from Request #3, fix a subset of the growth parameters to see if internal estimates of growth curves are stable and provide reasonable values. Start by fixing the CV of length at age 1 at 0.1 for both males and females (i.e., male offset should be zero). Then estimate the remaining parameters for both males and females (L1, L2, k, CV2).

Rationale: Estimation of all growth parameters in the draft base model caused estimates of CV1 to hit the parameter bounds.

Response:

Southern Model: The STAT completed a run using a combination of estimated and fixed growth parameters. All male parameters were fixed as offsets of female growth in order to produce models consistent with externally modeled growth. CV1 and CV2 were fixed at 0.1. Female L1, L2, and k were estimated. L1 decreased by 9 cm and L2 increased by roughly 6 cm when compared to the externally estimated values, causing the maximum length to be significantly larger than the external estimates (Figure 2). All estimated parameters were also flagged for steep gradients. This growth parameterization was used in the following request in an attempt to resolve these issues.

Northern Model: Some parameters were estimable, but the STAT needed to fix all CV's and set male L-min equal to female value.

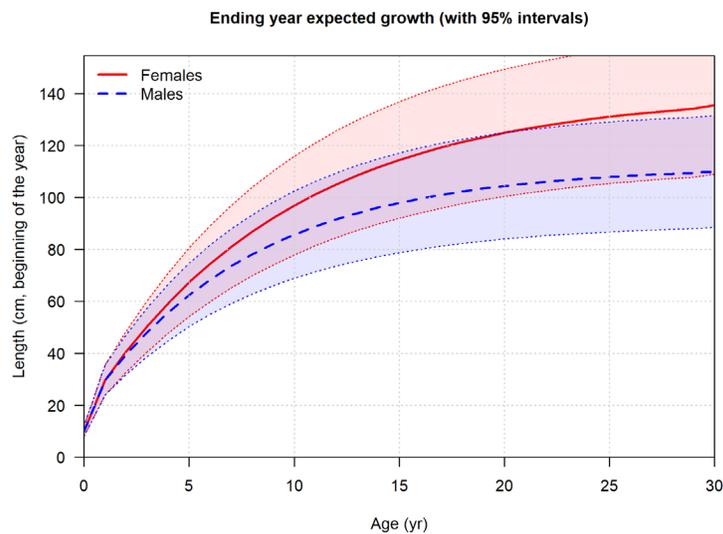


Figure 2. Estimated growth models for male and female California halibut in the southern stock. Estimated maximum length was significantly larger than the external estimates.

Request 5: [Growth] Using the models from Request #4, increase the input ageing error standard deviation (SD) from 0.05 to 0.1 and 0.2 as a function of age (two runs for each model) where the input ageing error for SD will be equal to $(1:\text{max age}) \times \text{SD}$, with age-0 fish having an SD equal to the SD value. Retain the assumption that there is no bias in the age value (e.g., -1). Example ageing error matrix input for SD = 0.2:

Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	...	Age 25
-1	-1	-1	-1	-1	-1	-1	...	-1
0.2	0.2	0.4	0.6	0.8	1.0	1.2	...	5.0

Rationale: Ages are assumed to be very precise in the draft base models, which may cause the model to have difficulty estimating growth given the limited amount of CAAL data in the model.

Response:

Southern Model: Increasing ageing error did not help with estimation of growth parameters. All growth parameters were fixed to the externally estimated values after exploring multiple approaches to estimating parameters within the assessment. Fixed growth in the southern model was used for all subsequent model runs.

Northern Model: Increasing ageing error appeared to help stabilize the model. The STAT was able to estimate some growth parameters (e.g., CV2 for both male and female fish). The STAT recommended estimating the majority of growth parameters, while fixing CV1 at 0.1 and L1 values (both sexes) equal to external estimates. Growth parameter estimation was possible once an age-increasing SD of 0.2 per year was added (i.e., a constant CV of 20%). Parameters were not estimable with a constant ageing error CV of 10%.

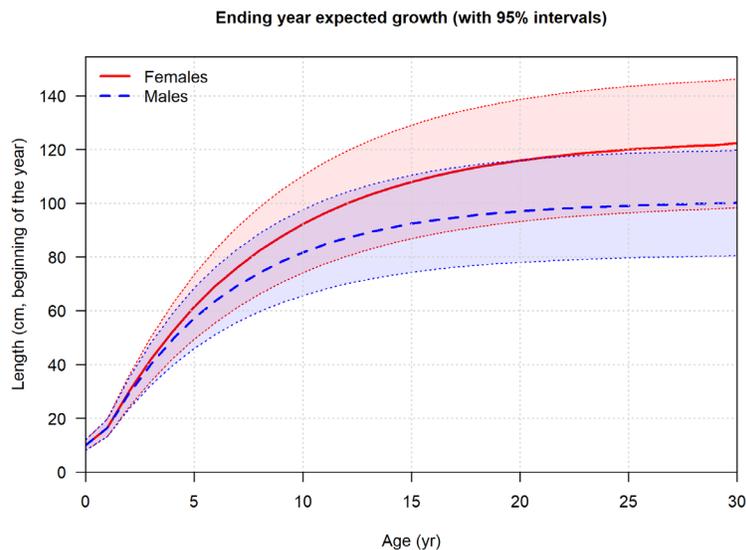


Figure 3. Fixed growth models for the southern stock.

Request 6: Using the models from Request #5, choose fecundity option #1, $\text{eggs} = W_t \cdot (a + b \cdot W_t)$, then set the intercept parameter (a) equal to 1 and the slope (b) equal to zero in both models. This assumes that total egg production is proportional to the biomass of mature females.

Rationale: This step is a simplifying assumption to help identify factors that might prevent estimation of initial F parameters in both models. Fecundity in the pre-review base model and request #5 uses option #1 with both intercept and slope set equal to 1. This is neither proportional to mature female biomass or equivalent to the reported allometric fecundity-length relationship with exponent ~ 5.9 .

Response: Switching from a model with intercept and slope equal to 1 (parameters a and b, as defined in the request) to a model with fecundity proportional to mature female biomass affected estimates of population scale and trend (Figure 4). The model with fecundity proportional to biomass (as in the base model) estimated a smaller, less depleted stock, relative to the model which assumed weight-specific fecundity increased with size.

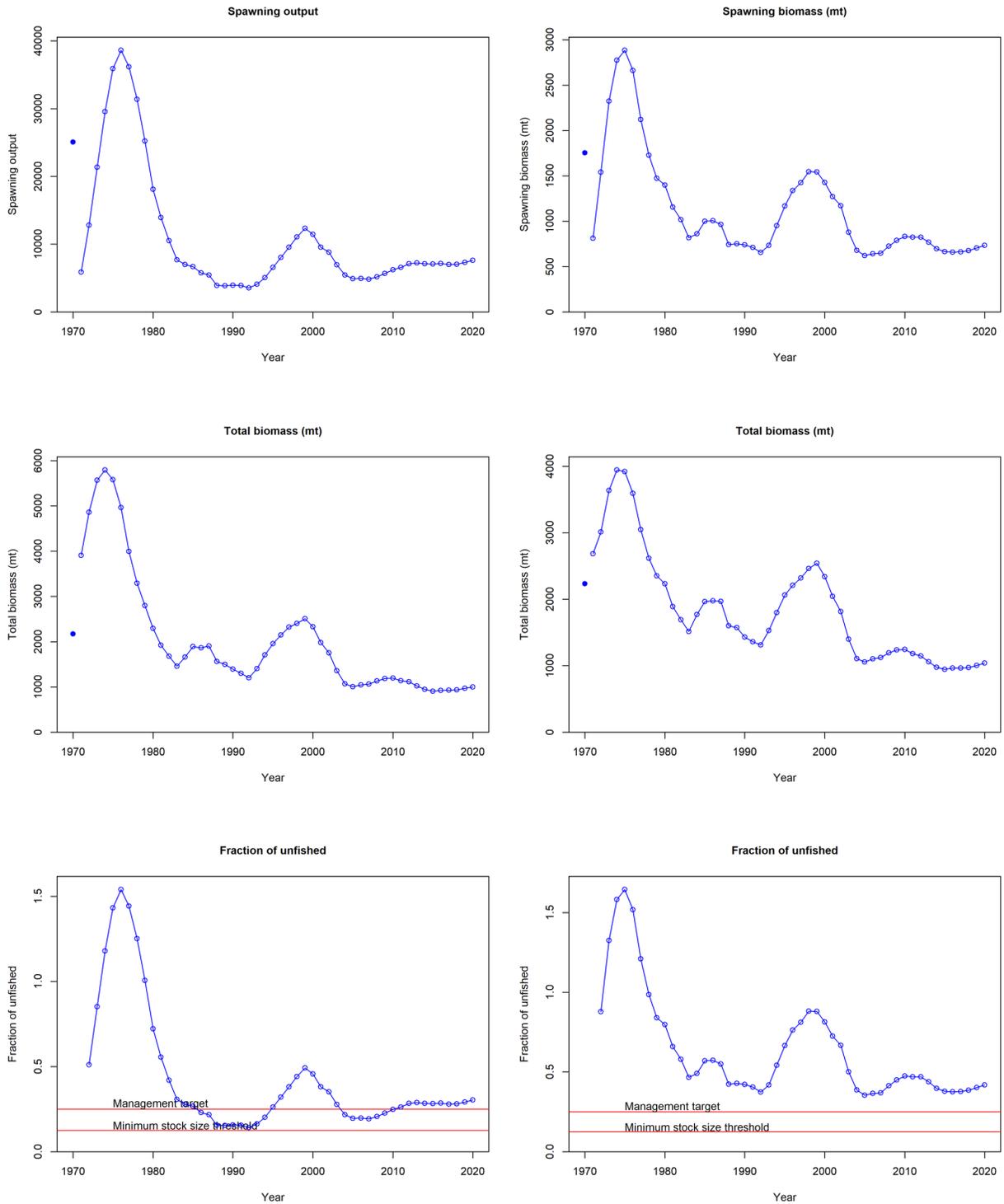


Figure 4. Changing the fecundity specification from Request 5 (left column) to an assumption of egg production proportional to the biomass of mature females (Request 6, right column) in the southern model. This changes units of spawning output, as expected (eggs vs. metric tons; top row), but also decreased total biomass in the early years (middle row). In terms of spawning output relative to unfished conditions, the model with fecundity proportional to biomass (Request 6) predicts a less depleted stock (bottom row).

Request 7: Using the model from Request #6. Use a descending logistic curve to model selectivity for the SF Bay fleet.

Rationale: The double-normal parameterization for the SF Bay survey fleet was hitting bounds. The logistic selectivity parameters (inflection and width) may be easier to estimate in this case.

Response: Initially, the STAT found no change to residuals for the SF Bay survey lengths, but the logistic parameterization resolved the ‘no move’ parameter flag and reduced the number of parameters overall. Later in the review, it was discovered that age-based selectivity for the SF Bay survey was equal to 0 for age zero fish, which were a significant component of the survey catch (based on observed lengths). A combination of descending logistic length-based selectivity and age-based selectivity set equal to 1 for all ages (0+) was implemented in later runs and helped reduce residual patterns observed for small fish in the survey, particularly in recent years (Figure 7).

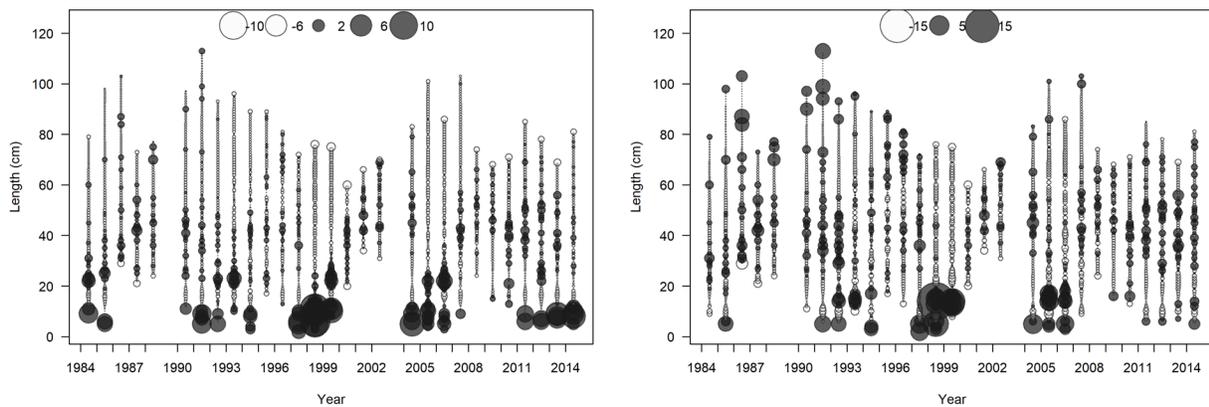


Figure 7. Residual patterns for SF Bay Study length composition data with descending length-based selectivity. The left panel (Request 7) shows positive residuals for small fish when age-based selectivity is zero for age-0 fish. The right panel assumes selectivity for age-0 fish is equal to 1 (final northern base model).

Request 8: Use the model from Request #6 or Request #7, depending on the outcome of the selectivity change for the SF Bay fleet. Force the remaining fleets (apart from the SF Bay fleet) in both models to have asymptotic selectivity.

Rationale: This step is an attempt to standardize the model structures between regions prior to estimating the initial F values (see Request 10). Once initial Fs are estimable, selectivity assumptions can be relaxed (e.g., allow for dome-shaped).

Response: Detailed results were not shown as this was a transitional step. The STAT indicated that the requested changes were applied to the southern model from Request #6 and the northern model from Request #7.

Request 9: [Initial Conditions] Using the models from Request #8, estimate initial F_s given the equilibrium catches in the draft base models (average of earliest 5 years in the model). Add priors if necessary?

Rationale: These runs use fixed M , fix some growth parameters, use stable and simplified selectivity assumptions, set fecundity proportional to female mature biomass, and will hopefully produce reasonable estimates of initial F_s in each region.

Response:

Note: the response to Requests 9 was based on models that did not include initial equilibrium catch in the likelihood function. See text at the beginning of this section for details. This error was fixed in the final base models.

Northern Model: The STAT indicated that this request produced initial F estimates that produced unrealistic estimates of initial equilibrium catch. The STAT explored using lognormal priors to approximate F 's needed to produce the first year of observed catches. Initial F for the commercial hook-and-line fleet was fixed to prevent it from hitting bounds.

Southern Model: Results for the southern model were similar to the northern model. The STAT was still in the process of exploring priors to achieve levels of catch in year-1 which were similar by fleet to the start year.

Request 10: [Initial Conditions] Using the models from Request #8, estimate initial F_s with equilibrium catches that better reflect the magnitude of catches prior to the initial year if possible (e.g., average annual total catch over a 30-year period (based on max. observed age) prior to the first model year). Allocate equilibrium catches to fleets based on best available information about the historical fishery. The review panel understands that the nature of the data may make this challenging given the assumptions that may need to be made.

Rationale: The equilibrium catches in the draft base model are small relative to historical catch. Equilibrium catches in this run better reflect the magnitude of recent historical catch.

Response: The STAT noted that reconstructing the catch by fleet, and particularly for the recreational fisheries is a challenge for California halibut. During the fourth review panel webinar, the STAT revised the input equilibrium catch levels for the southern model to approximate a 30-year average catch in the years preceding the model's start year (Table 1). The Panel requested further documentation of the sources and methods used to derive the estimates. The Panel also noted that the revised equilibrium catches for the CPFV, "Other Rec," and commercial hook-and-line were identical, despite the fact that recreational catch is in numbers of fish and commercial catch is in metric tons.

Table 1. Alternative assumptions about initial equilibrium catch by fleet in the southern model and associated estimates of the initial fishing mortality rate (F). Commercial catch units are metric tons (mt) and recreational catch units are in 1000s of fish.

	30 yr average Catch	Estimated F's	Previous Equil Catch	Estimated F's
BT	132.9071456	0.0336853	25.37052701	0.0078384
GN	132.9071456	0.0456313	63.2108397	0.026003
CPFV	13.99022585	0.020421	17.13974476	0.0285603
Rec	13.99022585	0.0149134	70.3003169	0.0830535
HL	13.99022585	0.0034379	2.731677543	0.0008136

The STAT reported that there was little change in relative biomass for the southern model when using the revised equilibrium catches, compared to previous runs. The Panel noted that comparisons of relative biomass alone do not provide information about the scale of the population, which has important implications for potential yield. The Panel recommends that trends in spawning output and summary biomass be included along with relative biomass trends in the final assessment document.

Attempts to estimate initial Fs in the northern model were unsuccessful. The F parameters hit the upper bounds even when the bounds were increased to unrealistically large values (e.g., 10), with significant changes in population scale.

Request 11: [Initial Conditions] If initial Fs are still not estimable in Requests 9 and 10, change the model start date to 1900 and assume the population starts from an unfisher state. Assume annual catches from 1900 to the start year of the draft base models equal the equilibrium catches. Compare the estimated initial F's from the models in Request 8 to the annual F's in the same year of these runs (e.g., 1980 for the central region and 1971 for the southern region).

Rationale: If initial F parameters are not estimable, then assuming constant catch over an extended time period should force the model to equilibrate to the roughly same state in the start years (1971 and 1980 for south and central, respectively). Fishing mortality rates in 1971 and 1980 in this configuration should be good approximations of the initial Fs when estimated freely.

Response: The STAT's response to this request came during the final review panel meeting, and focused on the northern model because estimates of initial F were possible in the southern model by that time. Catches were set to the 30-year average catch levels back to 1900, and estimates of F by fleet in 1979 were input as fixed parameters into the northern model. Results from the model that began in 1900 were not shown. Since the initial F parameters in the northern model were not estimable, the STAT ran a likelihood profile across a range of F values ranging from half to double the 1979 values in the previous run (starting in 1900). The relative magnitude of F among fleets was kept constant. The likelihood profile (Figure 8) showed a minimum at the smallest values of F (half the values derived from the model starting in 1900).

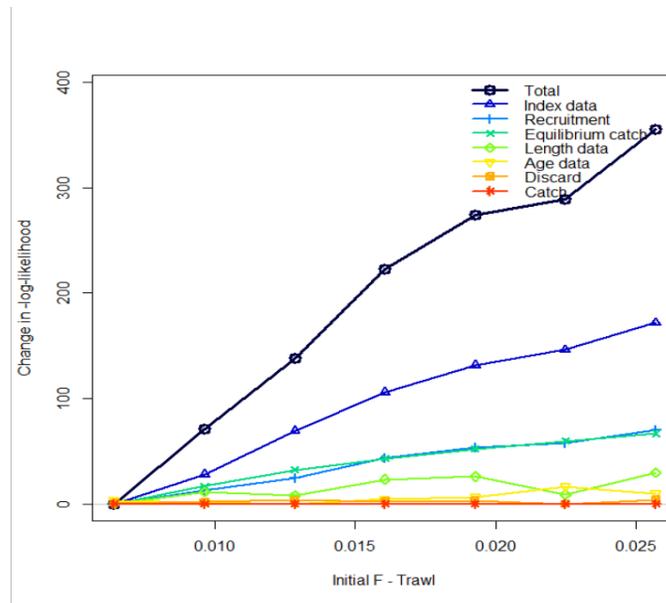


Figure 8. Likelihood profile over initial F values in the northern model. Results for non-trawl fleets show the same pattern at different scales for F.

This pattern in the likelihood profile (suggesting a better fit with smaller F values) contradicted the results for Request 10, which showed initial F values going to their upper bound when estimated. The STAT and panel were not able to arrive at an explanation for the differences in the time available.

The Panel suggested that future assessment efforts attempt to reconstruct catches farther back in time, as this may stabilize the models (particularly the northern model) and make them less sensitive to the initial F values. Historical catch data for some of the commercial and recreational fleets are published in Fish Bulletins 32, 49, and 174 (see References section). However, the reported fleet-specific catches are gross under-estimates and no catch information is available for roughly the first 50 years of the southern fishery, among other considerations for future reconstruction efforts. Correctly estimating the general magnitude of catches will be more important than defining allocations among gear types, as size at first selection does not appear to be dramatically different among fleets. Another possible solution to this problem of estimating initial conditions in the stock is to include an estimated SR_Regime parameter to allow the stock to be depleted at the start of the modeled period.

Request 12: [Discards] Plot histograms of tow duration by year and region (south & central).

Rationale: There is considerable uncertainty in discard mortality rates for the trawl fleet. A better understanding of the distribution of tow durations over space and time may help inform assumptions about the proportion of fish that survive after being discarded.

Response: The STAT provided histograms of trawl tow durations by region (Figure 9). The assumed discard mortality rates were larger in the south, although average tow duration was

longer in the north, which was the opposite of what was expected. The STAT agreed to revise the analysis and results are reported in Request 30 (below).

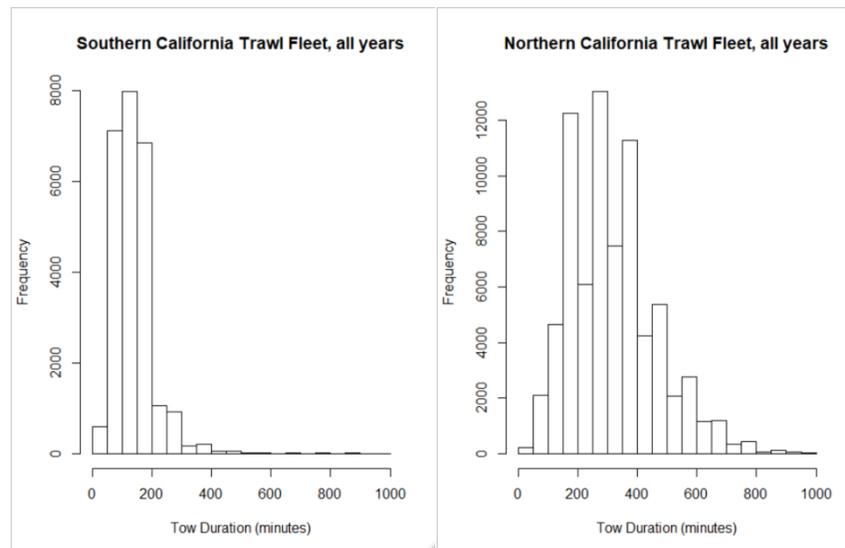


Figure 9. Distributions of trawl tow duration by area.

Request 13: Plot histograms of recreational length compositions by CRFS district (all years combined).

Rationale: Length composition data for the recreational fleet is not weighted by catch. If size compositions are similar across areas, then this is a reasonable assumption. If not, then length compositions should be catch-weighted.

Response: It was not possible to complete this request during the review. The Panel recommends that this analysis be completed prior to the next assessment to ensure that recreational length composition data are appropriately weighted and representative of the total catch in each model.

Request 14: Plot the distribution of bag sizes in the CPFV catch data.

Rationale: It is not clear whether bag limits affect CPUE estimation for the CPFV fleet. If anglers consistently catch their limit, then CPUE may not reflect changes in abundance. It is thought that this may be more of an issue in the central region than the southern region.

Response: The STAT reported bag size distributions from 2004-2018. The distribution of bag sizes in the north, where the bag limit is 3 fish, suggests that bag limits may have an effect on angler CPUE, particularly in recent years (2017-2019). The STAT noted that catch rates may still be accurate if effort (angler hours) drops when bags limits are reached. Also, CPUE since 2017 was at the highest level in the time series, and catch rates without bag limits would be expected to increase if effort remained constant. As such, recent increases in CPUE may be a conservative estimate. Bag limit distributions for the south show that anglers rarely catch their limit of halibut

and therefore bag size is unlikely to affect trends in angler CPUE. No adjustments to CPUE were made as a result of the analysis. The Panel recommended extending the analysis back to the beginning of the CPUE time series, if possible.

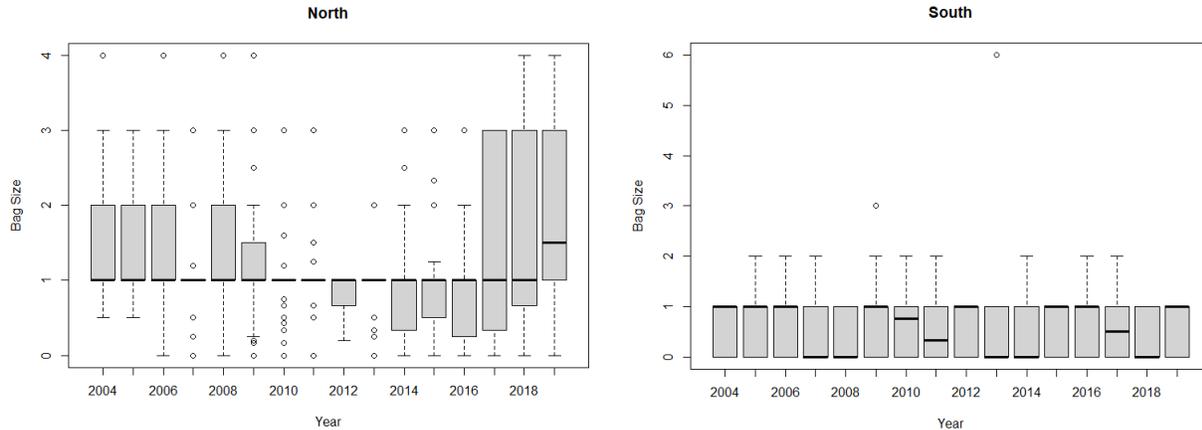


Figure 10. Bag size distributions for California halibut caught by the CPFV fleet in the northern region (left) and southern region (right).

Request 15: Compare trends in mean annual CPUE based on four seasonal subsets of the data (Jan-Mar, Apr-Jun, Jul-Sep, and Oct-Dec).

Rationale: The main effects structure of the CPFV index standardization model may cause the year effects to be sensitive to seasonal changes such as shifts in fishing effort and/or onshore/offshore migration of adults. Seasonal changes in vulnerability of adults may also influence estimation of selectivity curves. Further, the number of sample observations by month are not well balanced across years.

Response: The STAT plotted time series of CPUE using the requested subsets. Differences in CPUE trends among seasons were more apparent in the north than in the south (Figure 11). The Panel was concerned that all the index standardization models used only main effects (no interaction terms), which would not account for seasonal shifts in fishing effort and/or stock distribution. The STAT noted that in the south, the seasonal trends were generally consistent with each other although at different scales, with highest catch rates in the spring and lowest in the winter. The northern CPUE trends were less consistent among seasons, but also had gaps in sampling coverage.

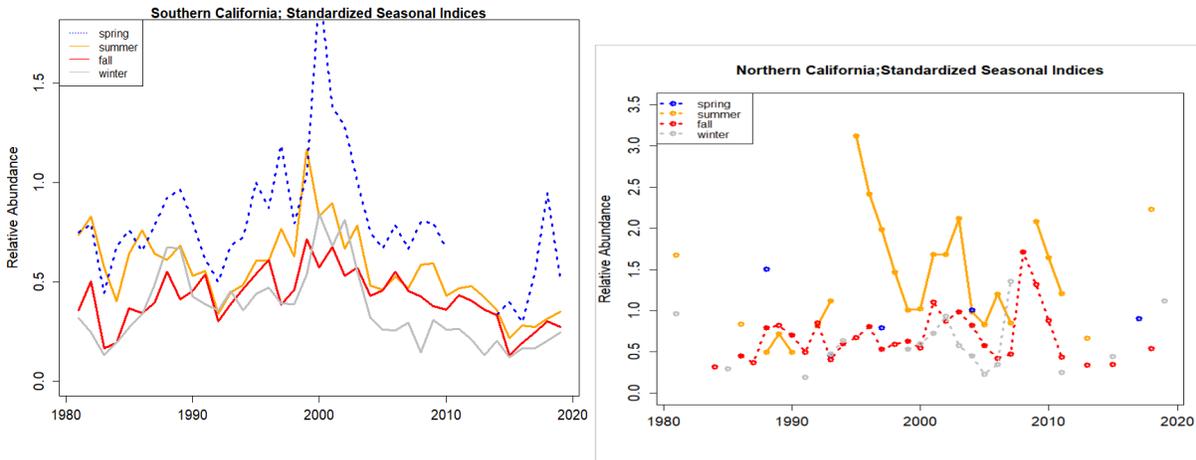


Figure 11. Mean CPUE time series by season in the south (left) and north (right). See request 15 for season definitions.

Request 16: Plot striped bass as a proportion of total CPFV catch by port or county for the northern model.

Rationale: Stephens-MacCall filtering in the north identifies striped bass as a strong predictor of halibut occurrence in the catch. This may be driven by trips in the SF Bay area, and may not reflect species compositions far away from the Bay (e.g., Morro Bay and Point Arena).

Response: The STAT obtained catch estimates from RecFIN, which are reported at the district level, and reported the proportion of striped bass in the total CPFV catch (Table 2). The STAT also noted that the logbook data separate Morro Bay from Monterey and Santa Cruz, so differences in species composition of the catch may be easier to detect in the logbook data than in district-level RecFIN data.

Table 2. Proportion of striped bass in the total recreational CPFV catch by CRFS district, 2005-2019.

Time Range	District Name	Retained All Species (MT)	Retained Striped Bass (MT)	Proportion Striped Bass
2005 to 2019	Bay Area	4189.08	224.75	5.37%
2005 to 2019	Central	3349.26	0.06	0.00%

Request 17: Do a sensitivity run using the alternative maturity ogives reported in the draft assessment. Report sample sizes for each study, and range of lengths observed, where available.

Rationale: The ogive based on histological examination is nearly knife-edged, which is unusual for most species.

Response: The STAT verified that histological methods are considered to be more accurate. However, the strong “knife-edge” pattern in maturity at length could be due to the fact that the majority of samples were either well above or below the minimum legal size (MLS). Data were sparse where size at maturity was most variable, i.e., around the MLS (Figure 12). The STAT compared trends in relative biomass for the northern model when using different published maturity ogives (Figure 13). Trends in relative status were generally consistent. The Panel recommended adding plots of spawning output (or summary biomass for mature ages) to be able to detect changes in population scale.

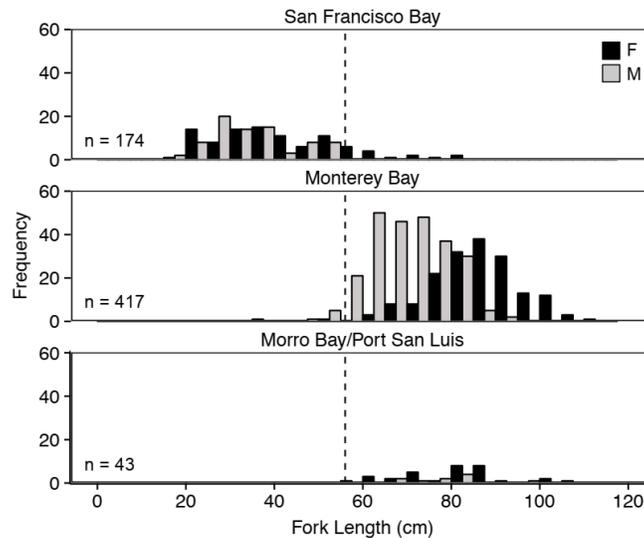


Figure 12. Size distributions of California halibut by sex and region used to develop maturity ogives in the assessment (figure from Lesyna and Barnes 2016).

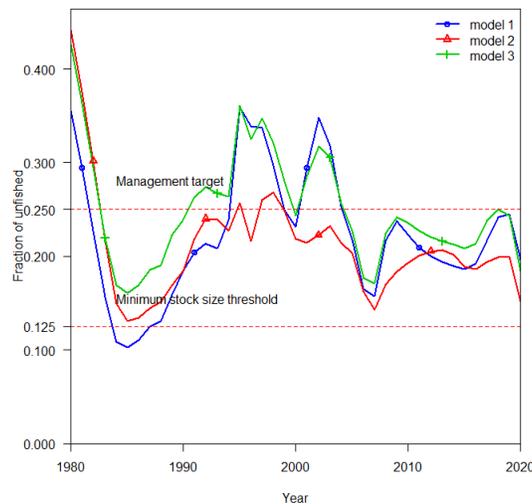


Figure 13. Effects of alternative maturity ogives on relative biomass in the northern model. Model 1 = base model; model 2 = Lesyna and Barnes 2016; model 3 = Love and Brooks 1990.

Request 18: The uncertainty for some data points from the indices is very low, resulting in the model not fitting some years. Estimate additive variance parameters for the indices in each model (can be done in the Q set-up section in the control file).

Rationale: Allows the model to increase the variance (reduce the influence) of indices that are poorly fit.

Response: Trends in relative abundance (especially terminal stock status) for the northern model were very sensitive to the estimation of additive variance parameters (Figure 14, left panel). Two indices in the northern region show an increase, and appear to be driving the recent spike in relative biomass. The estimation of additive variance parameters reduces the influence of the indices (Figure 14, right panel), and suggests a potential conflict between the indices and other data sources in the model.

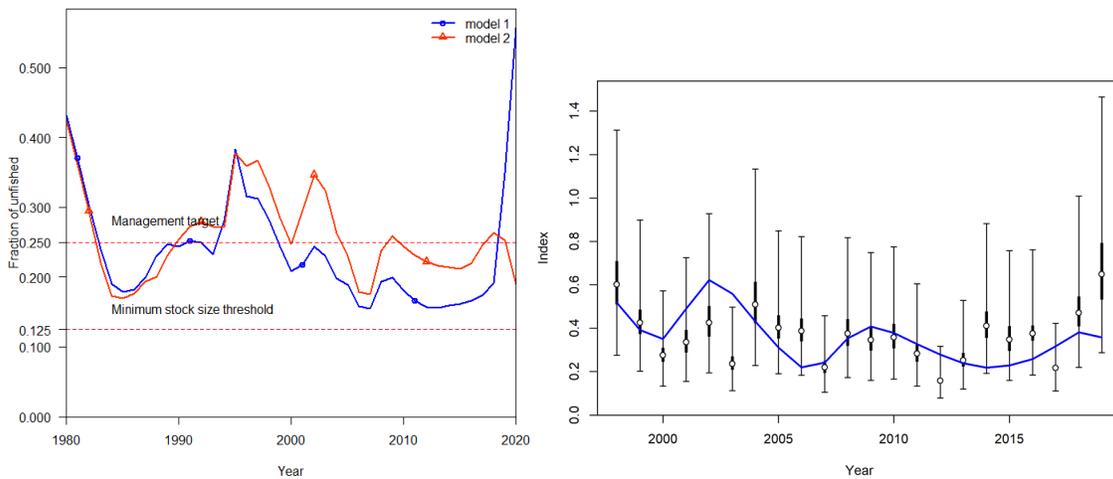


Figure 14. Influence of additive variance parameters on trends in relative abundance for the northern model (left panel: model 1 no added variance, model 2 with estimated additive variance) and fits to the bottom trawl index (right panel: model 2 with estimated additive variance). Lengths of the thick vertical bars in the right panel are the input variances, and thin bars illustrate the additional variance estimated by the model.

Request 19: Specify fecundity using the allometric fecundity-length relationship reported in the draft assessment (exponent roughly equal to 5.9). Make sure parameter estimates from the study are in the same units as SS (length in cm).

Rationale: Although this only represents batch fecundity, it reflects increases in weight-specific batch fecundity.

Response: The STAT developed a fecundity-length relationship and estimates of annual fecundity based in part on the work of Barnes and Starr (2018). There was not sufficient time to complete the analysis. The Panel recommends that the next assessment convert the units of the (batch) fecundity-length parameters reported by Barnes and Starr to units consistent with other data in the assessment (e.g., eggs or 100,000 eggs and fork length in cm), and to use the

converted fecundity-length parameters in the model. This assumes that the number of batches does not change with size or age. If there is evidence that the number of batches also changes with size or age, then additional calculations are needed to estimate annual fecundity. In addition, there may be important regional differences in batch frequency which should be considered in future models.

Request 20: Present tables showing the number of True Positives, False Positives, True Negatives, and False Negatives, based on the Stephens-MacCall model and threshold, for each application.

Rationale: There was some confusion about which category was removed. The panel defined “True Negatives” as trips that were predicted to catch no halibut, and caught no halibut (these should be excluded from the data set). The assessment said that “False Positives” were excluded, but this may be due to a different definition of the term.

Response: A generalized linear model was fit to presence-absence data of all species that occurred in the CPFV catch in an effort to objectively remove trips from the dataset that were unlikely to have caught a halibut. In this model, halibut (presence/absence) was the response and all other species (presence/absence) were the predictor variables for each individual CPFV trip. This approach follows the method described in Stephens and MacCall 2004.

The binomial model was used to predict the probability of observing a halibut, given the composition of associated species, for each individual CPFV trip. A critical probability threshold was defined (0.2517) in order to parse those results into something more meaningful for the purposes of determining whether or not a trip should be included in future analyses. Per the recommendations in Stephens and MacCall 2004, that probability threshold was set at the value that produces an equal number of false positive observations (i.e., halibut is predicted to occur on a trip when they do not), and false negatives (i.e., halibut is not predicted to occur on a trip when it does). In the southern model for example, this probability threshold resulted in 452,001 trips that neither produced a halibut, nor were predicted to produce a halibut. These ‘true negative’ trips were excluded from the model (Table 3).

Table 3. Number of CPFV trips per predictive category using a binomial GLM, with California halibut presence as the response variable, and all other species (presence/absence) as the predictor variable. This table represents the southern California CPFV logbook dataset.

True Negative	True Positive	False Negative	False Positive
452,001	31,630	43,863	43,863

Request 21: Develop the CalCOFI index for the core stations since 1951. Use a binomial GLM to predict annual probabilities of observing halibut in a standardized tow (rather than a delta-GLM).

Rationale: The extended time series may show evidence of the declines predicted by the southern model. Pre-model years can also provide additional information on seasonal and spatial patterns. Because incidence of halibut in the CalCOFI survey is quite low (around 5%), a simple binomial GLM is preferable for developing an abundance index.

Response: The STAT developed a binomial index for the full time series, but did not extend the modeled time period to include the extended index. The Panel noted that the log-scale standard errors in the data file for the southern base model are extremely large (average value of 2.4). There is likely an error in their estimation. The Panel recommended extending the modeled time period in the southern model to accommodate the entire CalCOFI time series, estimation using a binomial GLM with 3-year ‘super-years,’ and verification of the calculations used to compute the input standard errors.

Meeting 2 Requests

Request 22: Provide relevant model output (e.g., r4ss plots and tabular summaries of key parameter values [e.g., R_0 , M, growth] and likelihood components [e.g., for equilibrium catch, recruitment deviations, and parameter priors, and for length and age by fleet) for responses to June 1st requests. Also, include the data.ss, control.ss, and Report.sso files for the model developed in response to Request 1.

Rationale: Panel does not have access to all results presented during the webinar and would like to more closely examine the detailed results.

Response: The STAT provided all the requested materials via the shared Google drive prior to the start of meeting #3.

Request 23: [Discard] The revised model for Northern California uses discard rates from Southern California as data. To evaluate the potential impact of these borrowed data, conduct a sensitivity run for the northern model from Request 1 with a small lambda (e.g., 0.1) applied to the borrowed discard fraction data.

Rationale: Estimates from Southern California are used as data in the Northern California model. Need to confirm that these borrowed data are not overly influential.

Response: Borrowing the discard data from the southern model for use in the northern model, had little to no impact on the estimated spawning output, depletion, or annual recruitment deviations. The approach of borrowing the discard data was retained for the northern model.

Request 24: [Growth] Investigate literature values for length of young of the year halibut. Set the lower age in the growth model to zero years (effectively 6-month-old fish) and fix length at age zero to the literature value for both males and females.

Rationale: If unable to estimate size at age 1 (L1), having a fixed length for age-0 fish may produce more reasonable estimates of growth.

Response: The STAT team fixed the size of L1 at 10 cm for both males and females based on literature values (Allen 1988) and changed the age of L1 in SS to an age = 0. This eliminated the visual “kink” in the growth curve resulting in visually reasonable length at age for both males and females. The value of L1 and CV1 was fixed at the same values by sex and not estimated internally.

Request 25: [Growth] Include age data from collections that were not random samples from the fishery and/or surveys and attempt to estimate growth parameters within the model. A “dummy” fleet would only have the CAAL data in the model data file with the selectivity of this fleet in the control file set to selectivity pattern = 0 (equal to 1 across all ages) or logistic selectivity form. If selectivity pattern option 0 is selected this should be applied to both length and age selectivity, however, if the logistic selectivity form is chosen then it should be either length or age (not both). The decision of selectivity form is up to the analyst. Include the new age data in both southern and northern models.

Rationale: Some age data from non-random sampling was used to estimate growth externally, but not included in the model. Adding a “dummy” fleet of CAAL data from non-randomly sampled data may improve the model’s ability to estimate growth.

Response:

Northern Model Results: Added all non-randomly sampled ages in the northern model using a dummy fleet (fleet 7) with selectivity set equal to 1 across all ages and lengths. The likelihood contribution from the length data when using the dummy fleet data could improve growth estimates. The internally estimated values for the growth rate (‘k’) parameter differed the most from external estimates, relative to changes in length at the older reference age (L2) for each sex. Additionally, when the dummy fleet was added the L1 estimate seemed to be better informed and estimation of this parameter was turned back on for subsequent requests.

Southern Model Results: Added a total of 380 non-randomly sampled ages in the southern model using a dummy fleet (fleet 7) with selectivity set equal to 1 across all ages and lengths. Models which included the dummy fleet data were run estimating female growth parameters ‘k’, L1, and L2 individually and in combination. The additional age data did not appear to inform estimation of any of the parameters, and consequently these parameters remained fixed in all subsequent requests.

Request 26: [Growth] Produce tabular comparisons of the length-at-age data by fleet (e.g., Number of fish, mean[L(age)] and SD[L(age)]).

Rationale: To better understand how the dummy fleet data may differ from the "regular" data.

Response: A table of sample sizes by age for each dummy fleet source was provided. The Panel clarified that the purpose of the request was to compare length distributions (means, variances) to understand how the research data (dummy fleet CAAL data) compares to other fleets. This response was not completed, and is recommended for future assessments using the research data to estimate growth.

Request 27: [Initial conditions] Evaluate whether to estimate pre-model recruitment deviations or to assume an equilibrium age structure in the first year (no pre-model recruitment deviations).

Rationale: To determine whether estimates of initial Fs are influenced by estimation of pre-model recruitment deviations. Since age structure in year 1 is affected by both initial equilibrium F and pre-model recruitment deviations, attempting to estimate both may cause the model to be unstable.

Response:

Note: the first set of responses for requests 27 & 28 were based on models that placed a weight ('lambda') of zero on the likelihood component for equilibrium catch. Results reported here are from the fourth review panel (July 31, 2020), after this error had been identified and corrected. Initial Fs were estimable in the revised southern model, but not the northern model.

Southern Model: Turning off early recruitment deviations had little effect on the southern model. Trends in relative abundance were qualitatively consistent, and population scale was not significantly affected. A slight change in estimated recruitment deviations was noted, with slightly lower estimates in the early years and slightly larger estimates in later years, but overall very similar (Figure 15).

Northern Model: Treatment of early recruitment deviations had a large impact on trends in relative abundance for the northern model (Figure 16, left panel). When estimated, early recruitments are all below average (negative deviations), although it was unclear as to which data source was driving this pattern.

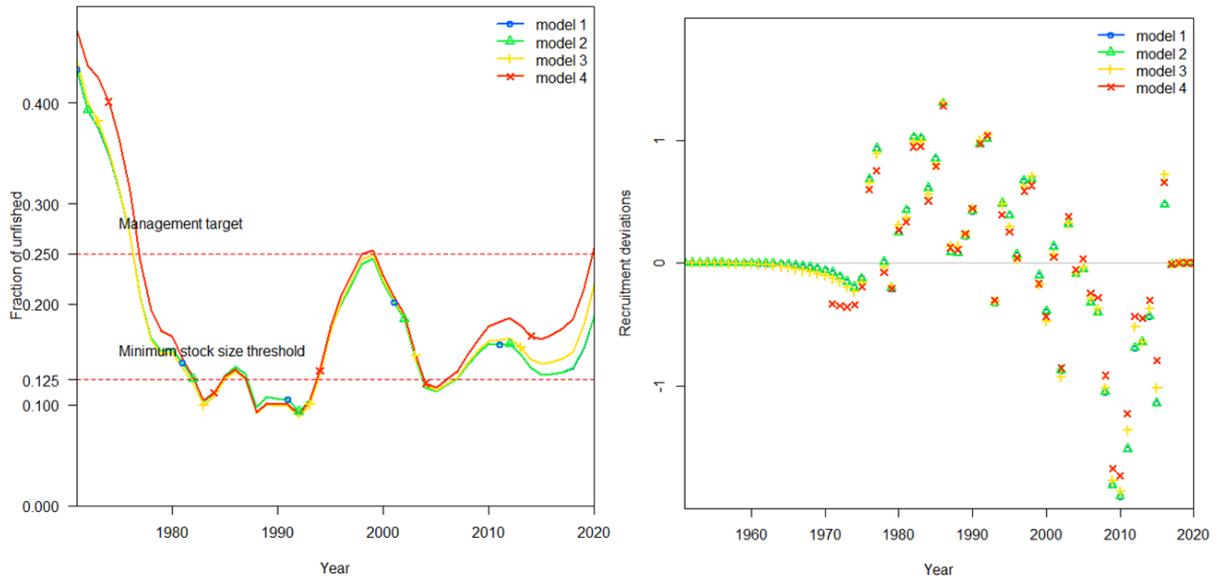


Figure 15. Relative spawning biomass (left) and deviations from mean recruitment (right) in the southern model changed very little when early recruitment deviations were turned off (model 4) relative to previous runs that estimated early deviations (models 1-3).

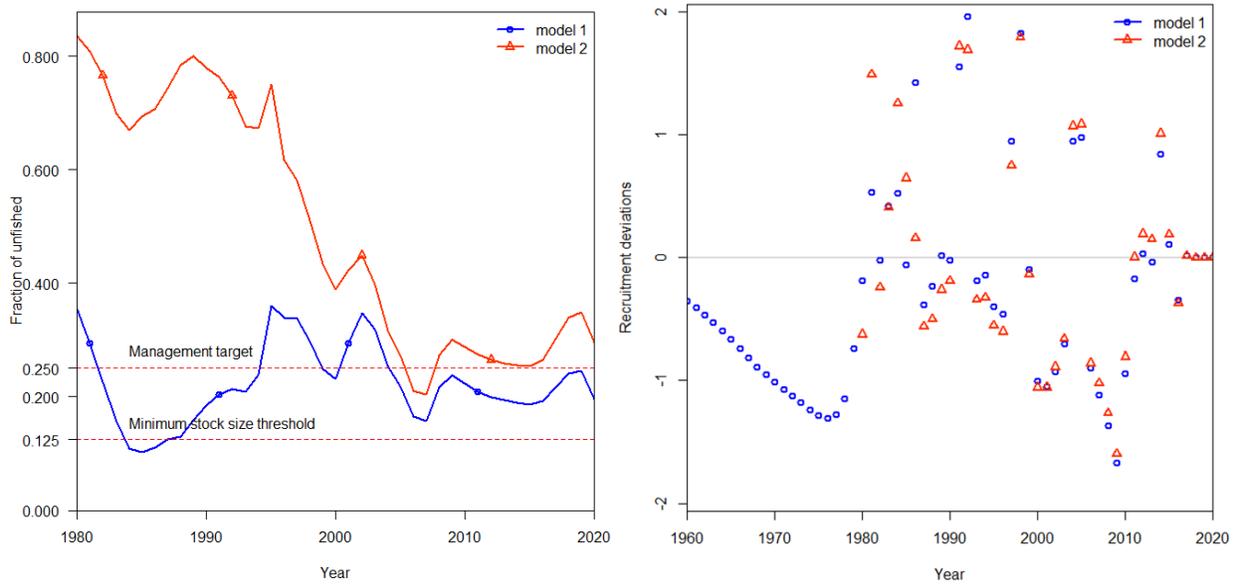


Figure 16. Differences in relative abundance trends for the northern model. Model 1 estimates early recruitment deviations back to 1960, whereas model 2 starts from equilibrium recruitment in the initial year (1980).

Request 28: [Initial conditions] Beginning from the fixed F_s used in the current base model (as determined by the assessor), scale fleet-specific F_s by factors of 0.5, 0.75, 1, 1.25, 1.5, 1.75, and 2 (a range that halves and doubles the initial F_s). Produce (negative log) likelihood profiles for the total likelihood as well as by data type (indices, lengths, ages, discard, equilibrium catch,

etc.), as well as time series plots comparing spawning output, relative spawning output, and recruitment deviations. The initial equilibrium catch likelihood values can be extracted from the CATCH section of the Report.sso file (Year = INIT for each fleet).

Rationale: Likelihood profiles will help identify which level of initial equilibrium fishing mortality is most consistent with the data, given the assumed allocation among fleets.

Response: Since initial F parameters could be estimated in the southern model, the response to this request focused on the northern model. The likelihood profile over multiples of the initial F parameters in the northern model (Table 4) displayed a strange pattern in the length component which was the major component of the total likelihood (Figure 17). The Panel felt that this was an indication of convergence (or other) problems in the northern model during the profile runs. There was not time to examine this profile further during the review panel.

Table 4. Initial equilibrium F parameters in the base model (“Init F’s”) and multiples of the initial F values used in the likelihood profile

	0.5 x F's	.75 x F's	Init F's	1.25 x F's	1.5 x F's	1.75 x F's	2 x F's
BT	0.241	0.361	0.482	0.602	0.722	0.843	0.963
GN	0.207	0.31	0.414	0.517	0.621	0.724	0.827
CPFV	0.007	0.01	0.013	0.016	0.02	0.023	0.026
Rec	0.027	0.04	0.053	0.066	0.08	0.093	0.106
HL	0.005	0.029	0.039	0.048	0.058	0.01925	0.077

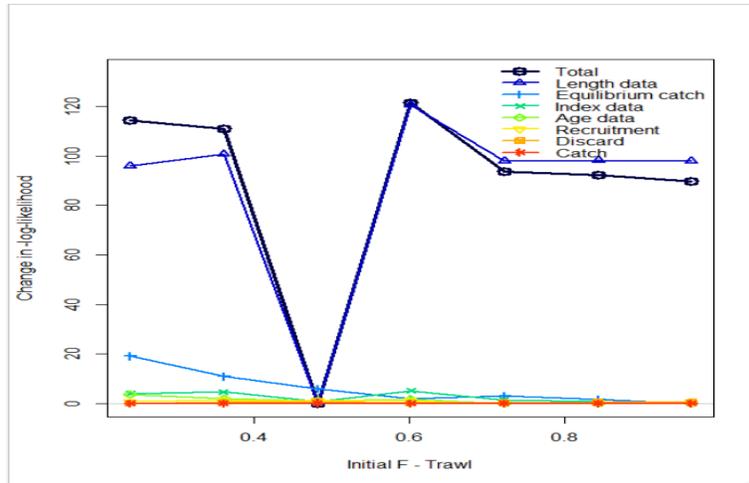


Figure 17. Likelihood profiles across multiples of initial equilibrium fishing rates by likelihood component and total likelihood in the northern model.

Request 29: [Initial conditions] Explore alternative levels of equilibrium catch. Keeping the same relative allocation among fleets, halve and double the equilibrium catch in each model and estimate initial Fs (4 model runs total).

Rationale: Equilibrium catch in the base model is set equal to the average catch over the earliest 5 years in the time series. As noted in other requests, differences in allocation are unlikely to have a major effect due to similarities in selectivity curves among fleets. However, the assumed magnitude of total equilibrium catch may affect estimates of initial conditions.

Response: The STAT found that changes in the magnitude of equilibrium catch (Table 5) had little effect on the initial relative stock size, but a large effect on terminal depletion (Figure 18) for the northern model. There was not time during the review to identify what was driving the change in stock status or to understand how the changes affected fits to the data.

Table 5. Multiples of equilibrium catch by fleet used for request 29 for the northern model.

	Equil Catch/2	Equil Catch	Equil Catch *2
BT	64.37543	128.750866	257.501732
GN	55.31072	110.62143	221.24286
CPFV	0.9585	7.8156	15.6312
Rec	3.9078	70.3003169	140.6006338
HL	5.179715	2.731677543	5.463355086

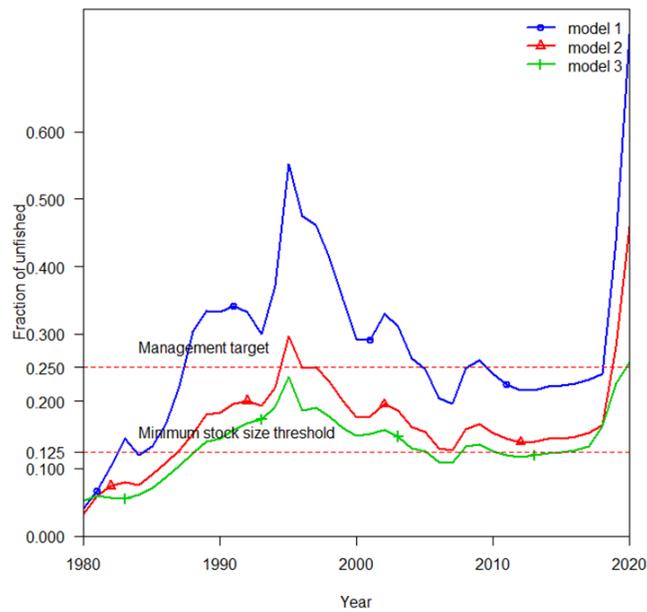


Figure 18. Effect of alternative equilibrium catch levels on trends in relative biomass for the northern model.

Request 30: [Discard mortality] Create a bar plot with the number of tows less than and greater than 60 minutes in each region. And, calculate the proportion of tows greater/less than 60 minutes duration and similarly for 90 minutes duration in each region.

Rationale: There was some confusion regarding the tow duration histograms, since the North one seemed to have a greater number of long (perhaps some unrealistically long) tows, but a greater fraction of tows were longer than 60 minutes in the South. A bar plot with a single break at 60 minutes may make the patterns more clear.

Response: Histogram figures for both the northern and southern model were provided (Figure 19). There was a calculation error identified in the previous request (Request 12). In the Northern area 99.6% of tows were greater than 60 minutes and 97% in the southern area. Based on this analysis the discard mortality rates were increased from 0.5 to 0.91 & 0.99 (South and

North, respectively). The Panel requested that the final assessment document be updated to include details of the revised calculations, data sources, and assumptions.

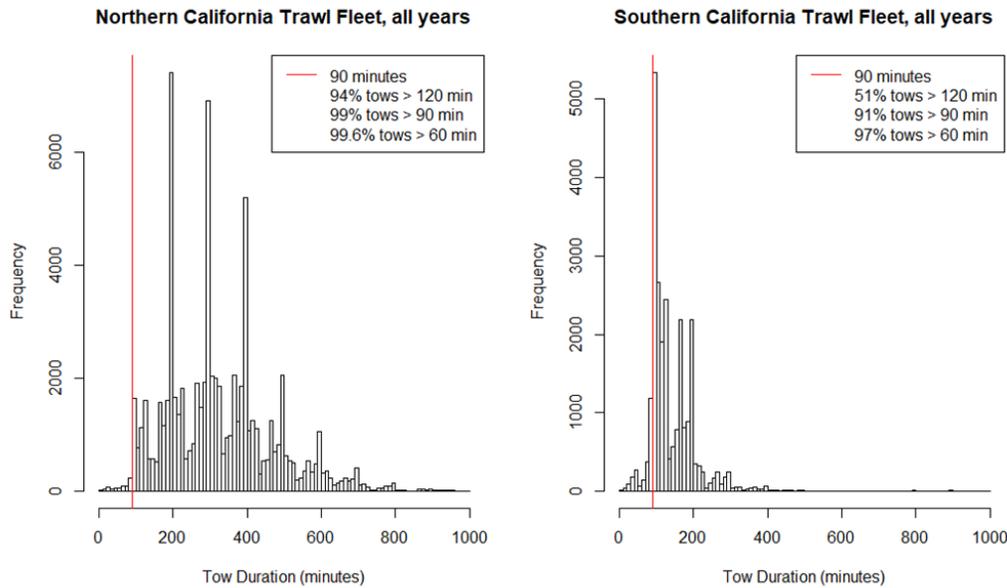


Figure 19. Frequency of trawl tow duration in minutes by region according to self-reported logbook data. Figure legends include percentage of tows that were greater than three time thresholds, which were selected to represent a plausible tow duration which could lead to significant mortality in the catch. The red vertical line represents the 90 minute threshold, which was selected to inform the discard mortality rate for each region. For example, 91% of tows were greater than 90 minutes in southern California so the discard mortality rate was set to 0.91 for the trawl fleet in that region.

Technical Merits of the Assessment

The assessment makes use of Stock Synthesis (SS v.3.30.14). This modelling framework can make use of a variety of disparate data and is particularly useful when time series data are discontinuous or where there are intermittent observations on length or age. It is therefore an appropriate choice for the assessment.

The STAT responded to many recommendations from the 2011 assessment, e.g., use of a Stephens-MacCall filter for recreational CPUE indices, inclusion of gillnet observer data, a time block for selectivity of the southern gillnet fleet, and combining trawl gears into single fleet. The STAT also completed an exceptionally large number of sensitivity analyses at the request of the panel. The current base models incorporate all available age data (fishery-dependent and fishery-independent) and estimate discard mortality using the best available data on discard rates by fleet.

Technical Deficiencies of the Assessment

Model

- The northern and southern models should both use point estimates for data weighting of length compositions based on the Francis (2011) method. The method used in the current base model incorrectly assumed weights were equal to 1.0 when the estimated confidence intervals contained 1.0.
- Age composition data were assumed to have a weight of 1.0 in the northern and southern models. Estimate Francis weights for age data in both models using point estimates, as noted above.
- The equilibrium catches entered into Stock Synthesis should include both landed fish and discarded fish that are assumed dead. It was not clear during the review whether the initial equilibrium catch levels were based on landings (retained catch) alone, or if they included dead discards.
- The log-scale standard errors for the CalCOFI index in the southern model are much larger than expected, and this is likely an error. This artificially reduces the influence of the index in the model.
- The CalCOFI ichthyoplankton time series extends back to 1951, but the index of spawning output was truncated in the southern model to exclude years prior to 1971. Extend the modeled time period for the southern stock to include catches from 1951 to the present and include the entire CalCOFI time series in the model.
- Length composition data for the recreational fleets were calculated from raw sample data. If size distributions differ by CRFS district, then length compositions should be catch-weighted to make them representative of the total catch. See Request 13.

Document

- In addition to the tables that indicate which selectivity and retention parameters are fixed and which are estimated, the assessment document should indicate which estimated values were considered unreasonable (provide the estimates or a qualitative description), and describe how fixed parameter values were chosen.
- Describe the data and methods used to estimate the input equilibrium catches for each fleet.
- Sensitivity analyses during the review often focused on changes in relative stock status. Each sensitivity should report changes in likelihood by data type and/or visual fits to data (goodness of fit), time series of spawning output (scale of the population), relative spawning output (status), as well as estimated growth parameters and recruitment deviations (productivity). If estimated, changes to other major parameters affecting stock productivity (e.g., steepness and natural mortality) should also be reported for each sensitivity run.

Areas of Disagreement Regarding Review Panel Recommendations

Among Review Panel members:

None

Between the STAR Panel and the STAT Team:

None

Management, Data, or Fishery Issues Raised by other Representatives during the Review Panel Meeting

None

Unresolved Problems and Major Uncertainties

The following issues were not resolved as of the final review webinar and/or contributed substantially to uncertainty in the assessment results. The next assessment should attempt to address each topic.

- The northern model was not able to estimate initial equilibrium fishing mortality rates. Fixing these rates while specifying the initial equilibrium catches (as in the base model) largely pre-determines stock abundance at the beginning of the modeled period. See previous recommendations regarding catch reconstruction and alternative treatments of initial conditions, e.g., starting from a (nearly) unfished state or estimating an offset parameter for recruitment in the first year.
- The northern model was very sensitive to modeling assumptions resulting in large changes in the estimated relative stock status across the modeled period (e.g., estimating added variance to the indices, initial F_s , estimating early recruitment deviations). This may be resolved by alternative treatment of initial conditions.
- Despite the inclusion of all available age data in the southern model, growth parameters could not be estimated internally (female or male). All growth parameters were fixed in the model at externally estimated values. The Panel and STAT were unable to determine why the parameters were not estimable, which seemed unusual given the amount of age data in the model. This issue occurred in the 2011 assessment, as well, so may take considerable effort to resolve.
- Only two female growth parameters were estimated in the northern model (von Bertalanffy “k” and mean length at age 15). All male growth parameters (offsets) were fixed at externally-derived estimates.
- Choices about the magnitude of input equilibrium catches in both models had a large influence on the total likelihood. This suggests that the assumed catches are not consistent with the available data.
- Both the northern and southern models were very sensitive to changes in steepness and natural mortality rates. A decision table analysis could provide a more credible measure of uncertainty in model results. This type of analysis considers the effects of alternative management actions under different assumptions about the state of nature (e.g., values of steepness and natural mortality).

Recommendations for Future Research and Data Collection

Analyses that rely primarily on existing data sources

1. Address topics listed under the “Technical Deficiencies” section
2. Reconstruct historical commercial and recreational landings using all available data (e.g., Fish Bulletins 32, 49, and 174) and reasonable assumptions about early, undocumented time periods. A time series of historical catches extending back to (or nearly to) the start of the fishery would eliminate the need to estimate initial fishing mortality rates. This is the most common approach to modeling groundfish populations off the U.S. West Coast. Estimates of initial F_s were sensitive to assumptions about initial equilibrium catch in the southern model, and were not estimable in the northern model. The fishery in the northern area developed later than in the south, which may make it possible to get a more complete reconstruction for the northern fishery; one that would not require starting the model from a highly depleted condition.
3. The CalCOFI ichthyoplankton time series extends back to 1951 in Southern California. Extending the modeled time period (at least back to 1951) would allow for the entire CalCOFI ichthyoplankton time series to be included in the southern model. Investigate why estimates of uncertainty (i.e., the annual log-scale standard errors) for this index were so large in the revised binomial index.
4. The current base models assume that all sizes and ages are equally vulnerable to the gears used in the research cruises. Compare size and age distributions of aged fish from research cruises and the fishing fleets. If they are similar, it may be more appropriate to assume that selectivity in the research cruises mimics one or more fisheries.
5. Likelihood profiles indicated that recruitment deviations were a large component of the total likelihood. The specified level of recruitment variability (σ_R of 0.6) was less than the standard deviation of the estimated recruitments (0.887) in the southern model. The influence of σ_R on estimates of recruitment and other model outputs should be examined further.
6. Given the size that halibut can attain, it may be beneficial to format the composition data using 2-cm length bins, rather than 1-cm bins as in the base models. This reduces model dimension, speeds estimation, effectively doubles the number of samples per data bin in conditional age-at-length data, and has been shown to produce unbiased results (Monnahan et al. 2016). The 2-cm bin width should apply *only to data bins*. Population length bins should remain 1-cm.
7. Evaluate the effects of CPFV boat limits on catch-per-unit effort indices. This analysis could include a description of the proportion of trips that reach the boat limit over space and time. Or, inclusion of an indicator variable for trips reaching the boat limit in the CPUE standardization model.
8. The southern model is sensitive to assumptions about the fecundity-size relationship (see Request #6). Develop size-dependent brood (batch) fecundity relationships (e.g., fecundity-length) that reflect increases in weight-specific fecundity (eggs/gram) with female size or age.
9. Long-term declines in the amount of estuarine habitat in California may have affected the productivity, status, and/or scale of the California halibut population relative to historical periods. This environmental factor is not explicitly accounted for in the current

assessment, but may be a topic worth exploring as part of a Management Strategy Evaluation.

Analyses that require additional data collection

10. Increase sampling of the full population's age structures to allow for internal estimation of all growth parameters in the models. This will also allow for comparison of regional growth patterns. The Panel recommends that CDFW increase efforts to collect age structures (otoliths) on an annual basis to assist with estimation of growth and recruitment parameters in the model. Sampling should account for both sexes in a way that is representative of the largest fisheries (e.g., bottom trawl and non-CPFV recreational boats). Roughly 50-100 otoliths per sex and major fishery (200-400 otoliths total, per year in each region) is a recommended minimum, with a target sampling rate of twice that amount.
11. Collect information on discard rates and the size distribution of discarded fish in major commercial and recreational fleets in both the Northern and Southern areas. This could include enhanced sampling of halibut trips by existing CDFW programs (e.g. onboard CPFV observer programs) and coordination with other agencies (e.g. NMFS' West-Coast Groundfish Observer Program) to ensure that sampling rates meet halibut assessment needs. Collection of discard information from the recreational skiff fleet remains a challenge (no observers). However, collection of representative samples of discarded catch from the skiff fleet would eliminate the need to make strong assumptions about the nature of discarded catch.
12. Barnes and Starr (2018) describe batch fecundity as a function of size. Investigate whether the number of broods produced in a year varies as a function of female size or age.
13. Collect additional gonads for maturity studies, being sure to adequately sample the 50-70 cm size range to better estimate the slope of the maturity ogive; do this for both the northern and southern areas.
14. Information on the densities and size/age compositions of California halibut, in particular in areas directly south of the U.S. California-Mexico border, would improve our understanding of ranges, dynamics and status of stocks which extend into Mexico.
15. Begin data collection in the expanding fishery north of Point Arena.
16. Investigate the influence of ocean warming on the distribution and life history characteristics of the two stocks.

Recommendations for Future Assessment Review Panels

1. Create a Terms of Reference for the stock assessment document that includes a list of required elements; ensure that these are all included in the draft disseminated prior to the first review panel and updated with any changes in the final document.
2. As noted in the terms of reference for the review process, "A primary goal of fishery management under the Marine Life Management Act (MLMA) is to ensure that fishing levels are sustainable and do not result in an overfished stock." A clearly defined list of

management goals or targets (e.g., a clearly defined harvest control rule, target biomass, relative stock status, or fishing mortality rate) will assist future review panels in determining whether the assessment is adequate for use in management.

3. Retain original data, analyses, and model input and configuration files (e.g., Stock Synthesis control, data, starter, forecast files, and executable) for reference during future assessment efforts. The Stock Synthesis files from the 2011 California halibut assessment were not available to the STAT, preventing direct comparisons with the previous assessment.
4. Consider adding one or more industry representatives as advisors to the review panel. Advisors familiar with the details of the fishery can provide details that can help inform decisions about model assumptions and data treatment.
5. Increase the size of the STAT (more than one person) during the review. This allows the STAT to divide up the work and provide a more thorough examination of the models. For comparison, assessments conducted by the PFMC are now limited to one model per week-long review panel, and STAT teams consist of multiple analysts. Ms. Meyer was the only analyst and simultaneously prepared results for two models.

Acknowledgements

The Panel thanks the STAT for their hard work, openness and responsiveness during the review. This assessment presented many unique challenges, and Ms. Meyer rose to the occasion every time. The Panel also thanks the Ocean Science Trust (Anthony Rogers, Dominique Kone, Kiya Bibby, and Jessica Kauzer) for inviting us to participate as reviewers and for managing the review process on behalf of CDFW.

References

- Allen, L. G. 1988. Recruitment, distribution, and feeding habits of young-of-the-year California halibut (*Paralichthys californicus*) in the vicinity of Alamitos Bay–Long Beach Harbor, California, 1983–1985. Southern California Academy of Sciences Bulletin 87(1): 19-30.
- Barnes, C. and R. Starr. 2018. Reproductive tactics of California Halibut (*Paralichthys californicus*): combining spawning season, interspawning interval, and batch fecundity to estimate annual reproductive output for a multiple-batch spawning fish. CalCOFI Reports 59: 102-114.
- Clark, G. 1930. Fish Bulletin 32. The California Halibut (*Paralichthys californicus*) and an Analysis of the Boat Catches. UC San Diego: Library – Scripps Digital Collection. <https://escholarship.org/uc/item/4390m8gc>
- Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68(6): 1124-1138.
- Hamel, O.S. 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES Journal of Marine Science 72: 62-69.

- Haugen, C. W. 1990. Fish Bulletin 174. The California Halibut, *Paralichthys californicus*, Resource and Fisheries. UC San Diego: Library – Scripps Digital Collection. Retrieved from <https://escholarship.org/uc/item/4867s6f4>
- Lesyna, K. and C. Barnes. 2016. Assessment of length and age at maturity for California halibut (*Paralichthys californicus*), including a histologically-based description of the reproductive cycle. California Fish and Game 102(3): 79-99.
- Monnahan C, K. Ono, S. Anderson, M. Rudd, A. Hicks, F. Hurtado-Ferro, K. Johnson, P. Kuriyama, R. Licandeo, C. Stawitz, I. Taylor, and J. Valero. 2016. The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fisheries Research 180: 103-112.
- Schnute, J. 1981. A versatile growth model with statistically stable parameters. Canadian Journal of Fisheries and Aquatic Sciences 38(9): 1128-1140.
- Staff of the Bureau of Commercial Fisheries. 1936. Fish Bulletin No. 49. The Commercial Fish Catch of California for the Year 1935. UC San Diego: Library – Scripps Digital Collection. <https://escholarship.org/content/qt24r046qt/qt24r046qt.pdf>