Widow Rockfish
Stock Assessment Review (STAR) Panel Report

NMFS, Northwest Fisheries Science Center
Newport Research Station, Bld. 955
2032 SE OSU Drive
Newport, OR 97365

July 27-31, 2015

STAR Panel Members:

Dr. David Sampson, Oregon State University, SSC (Chair)
Dr. Neil Klaer, Center for Independent Experts
Dr. Paul Medley, Center for Independent Experts
Dr. Ian Stewart, International Pacific Halibut Commission

Stock Assessment Team (STAT) Members:

Dr. Allan Hicks, National Marine Fisheries Service Northwest Fisheries Science Center

STAR Panel Advisors:

Ms. Heather Mann, Midwater Trawlers Cooperative, GAP
Ms. Lynn Mattes, Oregon Department of Fish and Wildlife, GMT
Mr. John DeVore, Pacific Fishery Management Council

Overview

The STAR panel reviewed a new full assessment of widow rockfish (*Sebastes entomelas*) off the west coast of the United States during a five-day meeting in Newport, OR. The stock assessment team (STAT), with remote assistance from STAT member Ms Chantel Wetzel (NWFSC) prepared a new benchmark assessment using the most recent Stock Synthesis 3 software, with a thorough reconsideration of the data and model structure, and including investigations of historical and recent landings and discards, length and age data and fleet structures. The last assessment was conducted in 2011.

The STAR Panel recommends that the new coastwide assessment for widow rockfish constitutes the best available scientific information on the current status of the stock and that it provides a suitable basis for management decisions.

Summary of Data and Assessment Models

Data

Stock Definition

Widow rockfish occur over hard bottoms along the continental shelf, forming dense, irregular midwater and semi-demersal schools at depths of greater than 100m at night and disperse during the day. They are medium-lived (rarely living longer than 20 for females and 15 for males), and bear live larvae. They range from Albatross Bank off Kodiak Island, Alaska to Todos Santos Bay, Baja California, Mexico, but are most abundant from British Columbia to Northern California.

There is little evidence that widow rockfish off the US West Coast are not a single coastwide stock, or that there would be much improvement to the assessment if it was split regionally. There is some evidence of differences in sampling programs among states, and that some population characteristics, such as mean size, vary with latitude and depth, but this could also be accounted for by migration of different components belonging to the same stock.

Although catches north of the U.S.-Canada border or south of the U.S.-Mexico border were not included in this assessment, it is possible that these populations form part of the U.S. west coast widow rockfish stock through adult and juvenile migration and/or larval dispersion.

The effect of a potential wider stock making a contribution to the US West Coast spawning stock biomass, particularly across the US/Canadian border, is unknown and is a considerable uncertainty. Under the criteria above, the selection of the stock boundary at the US border was done for data aggregation and management convenience. A Canadian catch history beyond 1998 was not available, which would otherwise have been useful for a sensitivity analysis.

Prior Probability for the Steepness Parameter

The prior derived for steepness and used in this assessment was based on the meta-analysis used for other 2015 rockfish assessments. For the pre-STAR base case model, the mean of this prior was used as a fixed value. However, the prior was incorrect because it contained the results of the 2011 widow stock assessment: described as ‘type B’ in Minte-Vera et al. 2005. The problem is that the type B prior is conditioned via the 2011 assessment likelihood profile on data also used in the 2015 assessment. Therefore data to 2011 are being used twice to inform the current assessment model parameter estimates. The ‘type C’ prior, leaving widow out of the prior distribution calculation, is the correct prior to use in this case, even if the value for steepness is fixed at some central tendency from that prior.

Historically, the likelihood profile for widow steepness suggested relatively low values, so its removal had a significant effect on the prior overall. It should be noted that this is also an issue for bocaccio, canary, darkblotched, black, and yellowtail rockfish assessments. The fixed steepness value in the final base case model followed SSC convention (use the mean of the prior), but used the Type C prior (omitting widow rockfish).

It appeared that SSC guidance for the 2015 assessments may have been inconsistent in recommending that STATs use the mean of the prior for steepness while also recommending that they use the median of the prior for natural mortality or other parameters. The STAR and STAT noted that the use of percentiles to develop low and high states of nature may be more consistent with the use of a median as a point estimate for a base model rather than the mean.

Catches

Most of the catches of widow rockfish have been taken by commercial trawl and hook-and-line fisheries since the early 20th century. They are desirable and not likely to be discarded for market reasons, although discards of smaller fish closer inshore are uncertain. Catches by recreational, commercial pot and commercial shrimp fisheries are low and were not considered in this assessment.

The shoreside and at-sea Pacific hake fishery catches widow rockfish as a bycatch, and catches from that fishery were estimated separately.

Catch for all rockfish species is uncertain, particularly for the historical period where rockfish catch was not reported by species and therefore needs to be apportioned using assumptions about species compositions. Further work can be done to evaluate catch uncertainty and to provide alternative plausible catch and discard series for sensitivity testing using the assessment model. Formal rockfish catch reconstructions have been done for Oregon and California, but not Washington. A Washington catch reconstruction (for commercial bottom trawl, midwater trawl, longline and net) was done for widow rockfish specifically for this assessment, and without reference to Washington catch reconstructions done for canary rockfish and black rockfish during the current assessment cycle.

The assessment appears to be less sensitive to historical catches (prior to the mid-1970s) than for many other rockfish, due to the very strong depletion signal during the peak catches over 1980-1982.

Length, age, sex and maturity data

There are good numbers of length samples by state for the non-hake fisheries (105-362 sampled trips per year overall 1981-2000, but lower recently at 13 to 84 per year 2001-2014). Good length sample numbers are available from the at-sea hake fishery (161 to 1491 hauls 1992-2014) and a small and irregular number of samples by state from the shore-side hake fishery (0 to 54 landings), mostly from Oregon. Good numbers of age-readings are available (80 to 272 fish ages per year during 1980-2000, 21 to 167 ages per year during 2001-2014) generally matching the availability of the length-only samples.

The method for expanding the compositional data from sampled fish to sampled catch by haul, and then from hauls to strata-level estimates was a topic of substantial discussion. The primary concerns were that the fixed number of samples targeted (for both fishery and survey data) requires expansion of the sampled fish, and not summary of raw counts, as was done for the conditional age at length data in the pre-STAR base case. Further, expansion to strata level estimates for the trawl survey data was performed using the biomass estimates by strata produced from the GLMM analysis, whereas numerical abundance is the appropriate metric. The use of the GLMM estimates is an important improvement in consistency over using the design-based expansions for the compositional data and the GLMM for the index; however, the use of biomass is not appropriate for compositional data, which are treated as observations of numbers of fish in the stock assessment model. Therefore, the STAR panel and STAT agreed that strata level weighting should be adjusted to use the numbers of fish consistent with the GLMM results (biomass estimates divided by the average individual fish weight) when combining across strata. The same approach of weighting by numerical abundance rather than by biomass also applies when combining age- or length-composition samples from different states or other geographic strata.

The STAT presented two differing estimates of maturity representing different sampling approaches, different time-periods and different spatial areas. It is currently unclear how to reconcile these two estimates, so an equal weighting was used. This was a source of uncertainty that could be reconciled in the future.

Abundance Indices

The relationship between fishery dependent indices and stock abundance are not necessarily well understood due to a variety of factors. The primary issue in this case is that the fishery independent indices are based on bottom-trawling, whereas widow rockfish spends time schooling in midwater. The use of the "exceptional catch event" (ECE) terms in the GLMM standardization is likely to make the index more robust, but it does not substitute for a sampling method with a higher encounter rate and more efficient sampling where fish abundance is greatest. An acoustic survey, conducted at night when the fish form dense midwater schools, might be more appropriate. However, efficient design of such as survey is likely to be challenging.

The treatment of discarding inside the assessment model represented an important exploration and improvement over just adding discard estimates to the landings. However, in some cases, where there was little evidence of size-based discarding or the magnitude of discards was very small, they could be added to the catches in future assessments. This would avoid the very large likelihood contributions from some sources (e.g., recent WCGOP for the trawl fishery) and any potential effect on the rest of the estimation, as well as simplify future analyses.

Unlike many previous assessments, the triennial survey was not modelled with a change in catchability between the 1992 and 1995 surveys, corresponding to a change in seasonal timing of the survey. This was justified through a sensitivity run (not documented) suggesting little estimated difference when the time-periods were separated.

Assessment model

The assessment used SS3 (ver. 3.24u) and was configured as a single area, single growth-morph, two-sex model, with recruitment deviations starting in 1900, catch starting in 1916, and steepness fixed at 0.798 (the mean of steepness prior developed during the STAR Panel based on a revised meta-analysis conducted by Jim Thorson that excluded widow rockfish information) and sigma R at 0.6. The maximum age bin was 40 years with a plus group. Five fishing fleets were defined: bottom trawl, midwater trawl, hake fisheries, hook & line and net. Natural mortality was estimated, using a lognormal prior based on a maximum age of 54. Abundance indices were weighted by input variance estimates and extra variance parameters were estimated for four of the six indices. Length and age composition data were weighted using three alternative methods, and conditional age-at-length was weighted, based on discussions during the review, to reflect appropriate sampling expansions and estimates of numbers in each strata. A standard iterative re-weighting procedure was applied to ensure weights on the different data sources within the model were internally consistent.

All fishery selectivity curves were specified as double-normal, although the only one fitted by the model as dome-shaped was midwater trawl (all others were forced to be asymptotic). Two selectivity time blocks were used with the bottom trawl fleet to account for implementation of the RCAs, four blocks were used with the midwater trawl fleet to account for management changes, and two blocks with the hook and line fleet to account for implementation of the RCAs. A cubic spline asymptotic selectivity curve was used for the NWFSC survey. Discards by the bottom trawl fleet were modeled via a length retention curve with five time blocks, for the midwater trawl fleet via a length-independent retention with four time blocks, hake fisheries assuming 100% retention and hook and line modeled via a retention curve with two time blocks.

A bridging analysis was carried out that separately examined the influence of the updated SS software version, updated catch, updated fishery-independent abundance indices and updated length and age compositions on the previous 2011 assessment. None of these changes caused a substantial change in the overall biomass trend, particularly in terms of relative depletion.

"Squid plots" of retrospective patterns in residuals show that it takes about eight years of data for estimated recruitments to settle as flat reliable estimates, indicating that there is likely to be considerable uncertainty associated with recent recruitments back to about 2008. This period includes the estimated high recruitment for 2010, suggesting that uncertainty in this estimate, particularly for projections, is an important uncertainty in this assessment. It appears that the most recent survey provides most of the signal that the 2010 year-class is relatively strong.

Data weighting was explored using three methods initially: harmonic mean, line fitted by eye through the scatterplot of effective N vs. input N, and Francis (2011) weighting. Appropriate weighting procedures are currently an active area of research and the Center for the Advancement of Population Assessment Methodology (CAPAM) has planned a workshop for October to examine alternative procedures and also hopefully to provide recommendations on standards. A known issue in SS is that reweighted input sample sizes less than 1 are rescaled back to 1 (due to an interaction with the requirements for bootstrapping) that potentially leads to a change in the relative weighting among years if small values are common in a data set (particularly a problem with highly partitioned conditional age-at-length samples). Strong arguments were made during the meeting to use the harmonic mean method until an alternative procedure is more broadly recommended, and this was agreed by the Panel and STAT for use in the base case for age and length compositions and also conditional age-at-length data. The problem of Synthesis re-scaling input-N values of 1 can be avoided by iterating to the required weighting value as usual, then setting all reweighting values back to 1 and applying the weight value directly to the appropriate likelihood component using a lambda. This method was also agreed by the Panel and STAT, while also recognizing that input and effective sample sizes in diagnostics are difficult to interpret when lambdas are used.

At the previous Black rockfish STAR Panel it was recommended that recruitment deviations be turned off for the early period where recruitments are not informed by subsequent composition data so as not to give the model freedom to introduce periods of above or below average recruitments in the early period to arbitrarily adjust biomass before the data rich period of the assessment. It was noted that turning off early recruitment deviations also turns off the associated error, therefore not properly accounting for model uncertainty during that early period. As the pre-STAR Widow rockfish base model available for examination did not show undesirable behavior of recruitment deviations in the earlier period (they were near flat), it was agreed to leave them turned on.

It is an open question whether early recruitment deviations for these assessments should be turned off by default for the base case. The current STAR Panel process requires the selection of alternative models that define axes of uncertainty rather than through the assessment of error within the base case model. Because individual model uncertainty is not used for management recommendations anyway, and to prevent undesirable behavior of early recruitment deviations in models chosen for axes of uncertainty because detailed diagnostics are not usually examined for those models, early recruitment deviations may not be useful in the base case. Advice is also required for how best to set bias adjustment if early recruitment deviations are switched on so that an average bias off the curve for the MLE model result is avoided.

The combined likelihood profile for natural mortality suggested an estimatedoffset value might be used to define the difference between the sexes, and that uncertainty in M can be expressed in one dimension. The difference in natural mortality between males and females might be explained by their sizes, the difference being consistent with a simple model of size-dependent mortality (Lorenzen 2000).

Presentation of MCMC results was useful and should be encouraged as a diagnostic, even though the uncertainty from MCMC is not generally used for management advice. MCMC provides a possible alternative source of information for, and check on, the axes of uncertainty.

The Panel, with concurrence from the STAT, requested a number of cumulative changes to the preliminary base model, as detailed in Request 4.1 below, that resulted in an agreed new base model. The final agreed base model has been well structured, thoroughly investigated by the STAT and is the best currently available for the formulation of management advice.

Treatment of uncertainty

Approximate parameter standard errors and correlations are estimated routinely using the local slope of the likelihood surface. This gives some indication of joint parameter uncertainty. Alternative and sometimes non-symmetric estimates can be obtained from MCMC simulations, although the simulations, which can take considerable tie to complete, had not fully converged for all parameters for the widow model presented during the STAR panel. Nevertheless, parameter uncertainty was estimated for all models.

In addition, likelihood profiles were conducted for R_0, steepness and natural mortality, which are key parameters in determining stock productivity and status. Likelihood profiles can be used as estimates for the marginal probability density of these parameters and indicate how much information is contained in the data to inform parameter estimates. They are created by fixing the parameter over a range of values with the prior removed and calculating the likelihood for each data source over the range. R_0 was well estimated, with the abundance indices favoring higher values than the age and length data. Natural mortality also appeared reasonably precisely estimated, with much likelihood strongly favoring estimates above 0.1. Most information on natural mortality came from the age compositions. The steepness profile suggested that there was very little information in the data on this parameter.

Retrospective analyses were conducted to see how well key parameters at the end of time series were estimated (Figure 1). No systematic problems were indicated, but estimates of recruitment strengths in particular were highly uncertain for 2009 onwards. This indicated that observations of roughly seven or more years are required on a cohort before estimates of its size are likely to be reliable. Therefore although the 2010 year class appears strong, how large it really is may not become clear for a number of years.

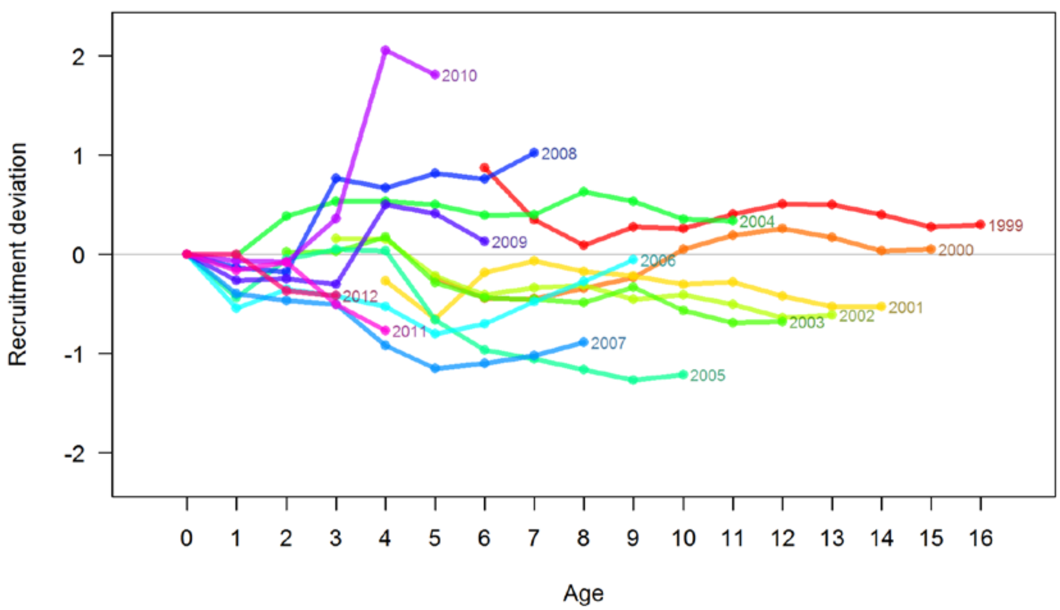


Figure 1. "Squid plot" illustrating retrospective patterns for 3-16 year olds recruited 2012-1999 respectively.

Sensitivity runs are probably the most important source of information on uncertainty as they can be used to map out the possible effects of structural errors and model assumptions. These are the greatest source of uncertainty, but are difficult to assess. Sensitivity runs covered low-high ranges for steepness, natural mortality, Francis (2011) weighting, asymptotic selectivity for the mid-water trawl (as opposed to the base case domed selectivity) and removal of the 2012-14 survey composition data. During the STAR panel, additional sensitivity were requested consisting of an alternative catch history (0.5 and 1.5 of the historical catch 1916 – 1982), alternative maturity curves, exclusion of the hook and line and net composition data, increased weight on trawl age compositions, fixing the male and female natural mortality to the same estimated value, and testing a low-high range for the 2010 recruitment deviate.

In most cases, the sensitivity changes had little or no impact on the assessment results. The greatest sources of uncertainty were associated with natural mortality, steepness and the size of the 2010 recruitment (Figure 2). For these sensitivities, the bounds for the parameter intervals were the 75% prediction interval around the model estimates using the estimated standard error for 2010 recruitment deviate and natural mortality, and the steepness prior 75th percentiles. The STAR Panel agreed with the STAT that the axis of uncertainty should cover not only uncertainty estimated by the model, but uncertainty external to the model. For this reason a steepness range was included as it represented critical uncertainty associated with stock recruitment which could not be estimated within the model. The size of the 2010 recruitment had little impact on the model except through the projections, which would be important in developing a decision table.

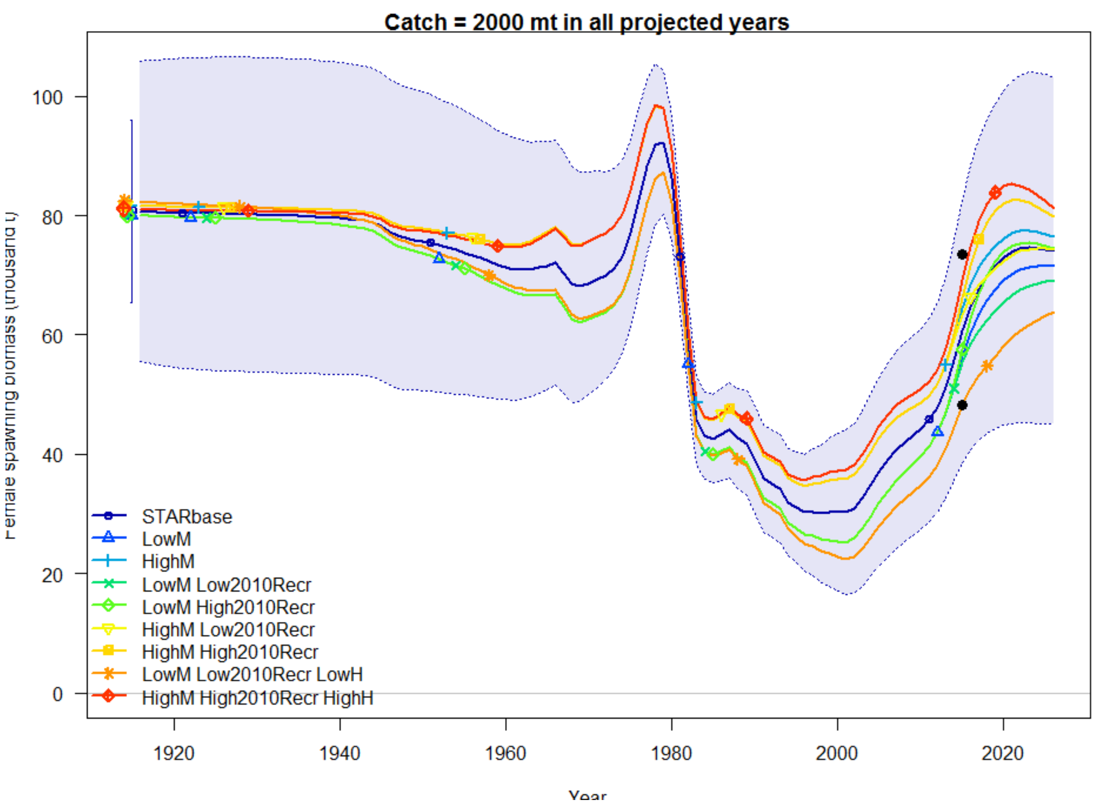


Figure 2 Sensitivity analyses used to define the axis of uncertainty showing the range of spawning stock biomass, including projections to 2030. The black dots for 2015 mark the biomass range recommended by the SSC as a suitable range for the uncertainty axis (biomass 25% - 75% around the base case MLE estimate).

Requests by the STAR Panel and Responses by the STAT

The pre-STAR draft document was very complete. This allowed for an efficient and effective review that could quickly identify the most important questions and allocate review time accordingly. The STAT team provided thorough responses to all requests.

Request 1.1: Characterize historical catch uncertainty (landings and discards).

Rationale: This will provide a basis for developing a sensitivity.

Response: The STAT did not have sufficient time to accomplish this request.

Request 1.2: Provide a summary of indices used in the assessment with rankings on expectation of how informative they may be as indices of abundance.

Rationale: Assist in gauging the *a priori* importance of the indices.

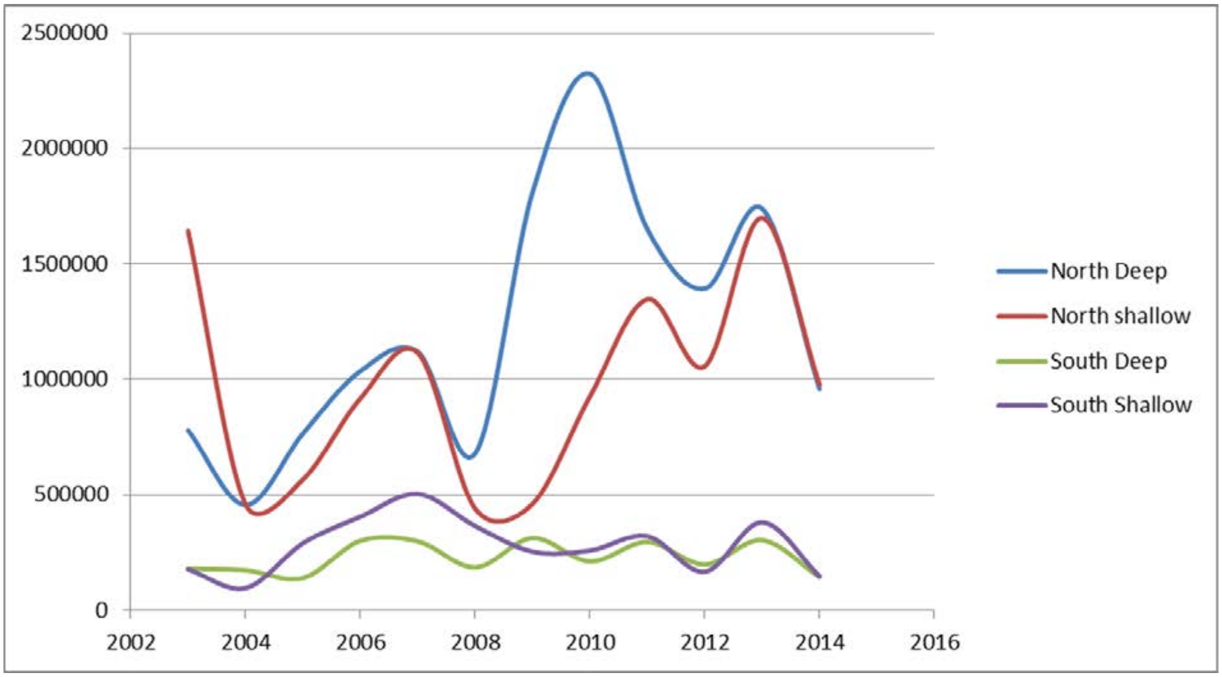
Response: The STAT provided the following table.

Region	Fleet	Years	Name	Fishery independent	Filtering	Method	Rank	Method endorsed	Units
Coastwide	8	2003–2014	NWFSC shelf/slope survey	No	South of 34.5 removed	GLMM, Gaussian, ECEs	1	SSC	Biomass
OR	1	1984–1999	Oregon Bottom Trawl	No	Jan–Mar 42.5–46.5 & 124.6-124.9 >1000 lbs	Delta-GLM	2	Past assessments	Pounds/hour
OR/WA	3	1991–1998	Domestic at-sea	No		Delta-GLM	3	Past assessments	
OR/WA	3	1983, 1985–1990	JV at-sea bycatch	No		Delta-GLM	4	Past assessments	
Coastwide	7	1980–2004 (triennially)	Triennial trawl survey	Yes	None	GLMM, Gaussian, ECEs	5	SSC	Biomass
Coastwide	9	1977-82, 1984-88	Foreign at-sea bycatch	No		Delta-GLM	6	Past assessments	
Coastwide	6	2004, 2005–09, 2011 2013-14	Juvenile Survey	No		ANOVA	7	Past assessments	Numbers

Request 1.3: Provide average length, growth, and average weight from all data sources by state. Provide GLM biomass estimates from the NWFSC survey by area.

Rationale: Explore the current weighting by biomass that was used to develop the compositional data.

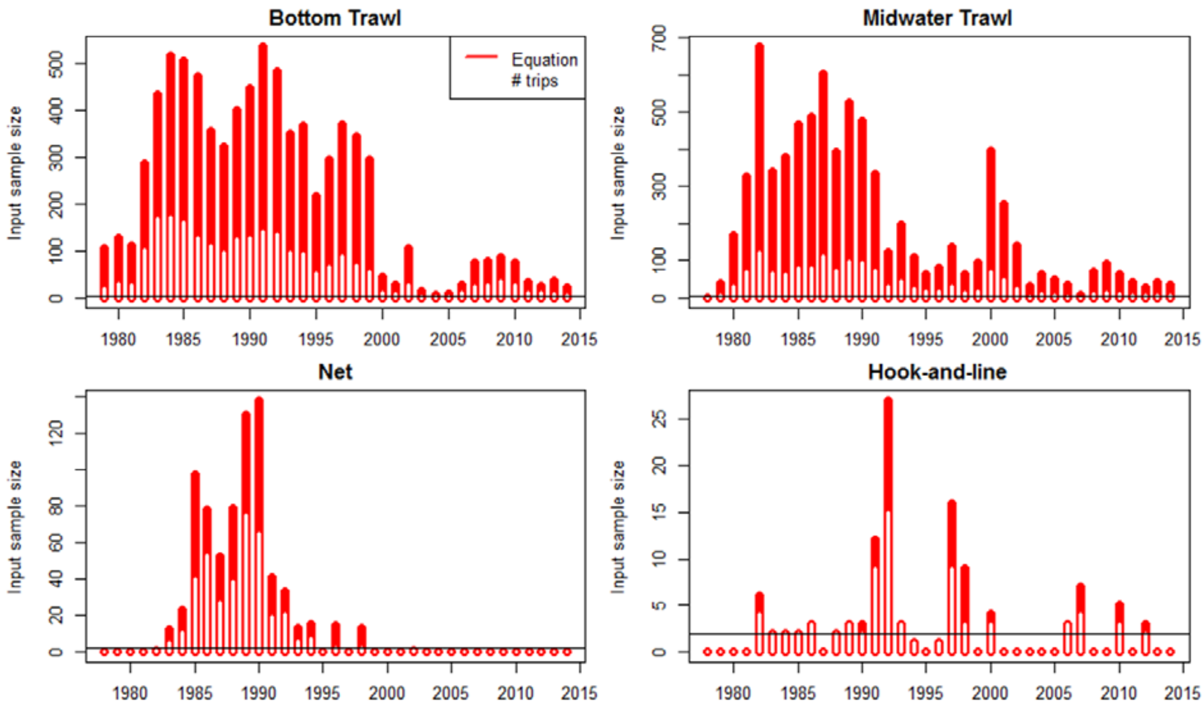
Response: The STAT provided a series of slides showing the weights of individual fish. The two slides based on survey data showed a tendency for there to be smaller fish in the southern portion of the assessment region (south of ~42ºN). The slides comparing the distributions of fish weights and lengths by state generally showed relatively minor differences between the states. The STAT also presented two slides comparing scatterplots of length against age, and fitted von Bertalanffy growth curves, by state and sex, for all data (fishery plus survey) and for survey data only. These plots supported the notion that there are no regional differences in growth. The final plot (below), which showed the strata coefficients from the GLMM analysis of the NWFSC trawl survey data, suggested that biomass trends over time have not been uniform across the four spatial strata.



Request 1.4: Provide a plot comparing input sample sizes by tow and compare to the current input sample sizes.

<u>Rationale</u>: Explore whether the current input weighting scheme would differ using number of tows.

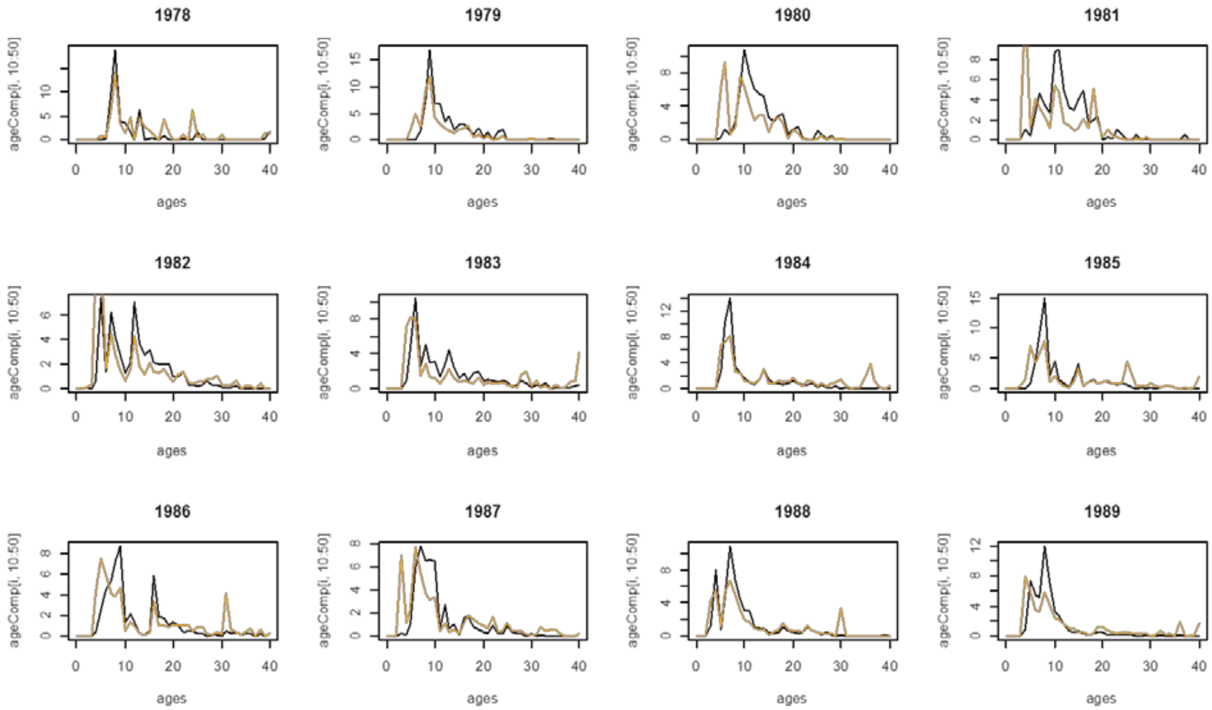
<u>Response</u>: The STAT provided a number of plots comparing (a) the input sample size calculated on the basis of the standard formula (from number of fish per sample and the number of trips sampled) and (b) the number of trips associated with the sampled fish. An example is provided below. The horizontal line indicates an input sample size of one, which apparently is a threshold below which Stock Synthesis will not let *Input-N* values drop. (This "feature" was implemented in Synthesis to prevent problems that otherwise arise when generating bootstrap sample data.). The evidence suggested that weighting by the number of samples might produce different results from those obtained using the standard calculation for *Input-N*.



Request 1.5: Collapse conditional age-at-length (CAAL) data (raw and expanded) to the marginal age and length comps.

Rationale: Explore consistency between marginal and conditional compositional data.

Response: The STAT produced two graphs comparing the marginal age-compositions obtained from collapsing the conditional age-at-length compositional data versus the marginal age-compositions obtained from the standard expansion process for two-stage samples (expand the sample to the trip, then expand the sampled trips to the entirety of the catch). The graph below is for the data from the bottom trawl fishery. The results suggest that the current treatment of the conditional age-at-length compositional data (no expansion for variable sampling fractions) is inconsistent with how marginal length-compositions are calculated (with expansion for variable sampling fractions).



Request 2.1: Sensitivity on the maturity curves from CA and OR samples. Explore greater weighting to OR samples by alternatively using all OR vs. all CA samples. If the results are sensitive to 100% weighting of samples from one area, use trawl survey index estimates in numbers of fish as a weighting mechanism (perhaps divide GLM biomass estimates by mean weight per stratum).

Rationale: The maturity curves derived from OR and CA samples are significantly different. However, the composite maturity curve used in the model gave equal weight to CA and OR samples. Since the apparent density of widow rockfish is higher in the north (OR is the center of population distribution), the effect of giving a higher weight to OR samples needs investigation.

Response: The STAT produced graphs and a table contrasting the influence of the CA versus OR samples on the model results. The table is reproduced below.

	Base	CA maturity	OR maturity
SB0	82828	86577	78011
SB2015	60771	68457	50047
RSB2015	0.73	0.79	0.64
EqYield_SPR	7707	7995	7318
M female	0.1544	0.1543	0.1545
M male	0.1679	0.1679	0.1680

Request 2.2: Provide runs with logistic selectivities for the surveys and the hake fishery.

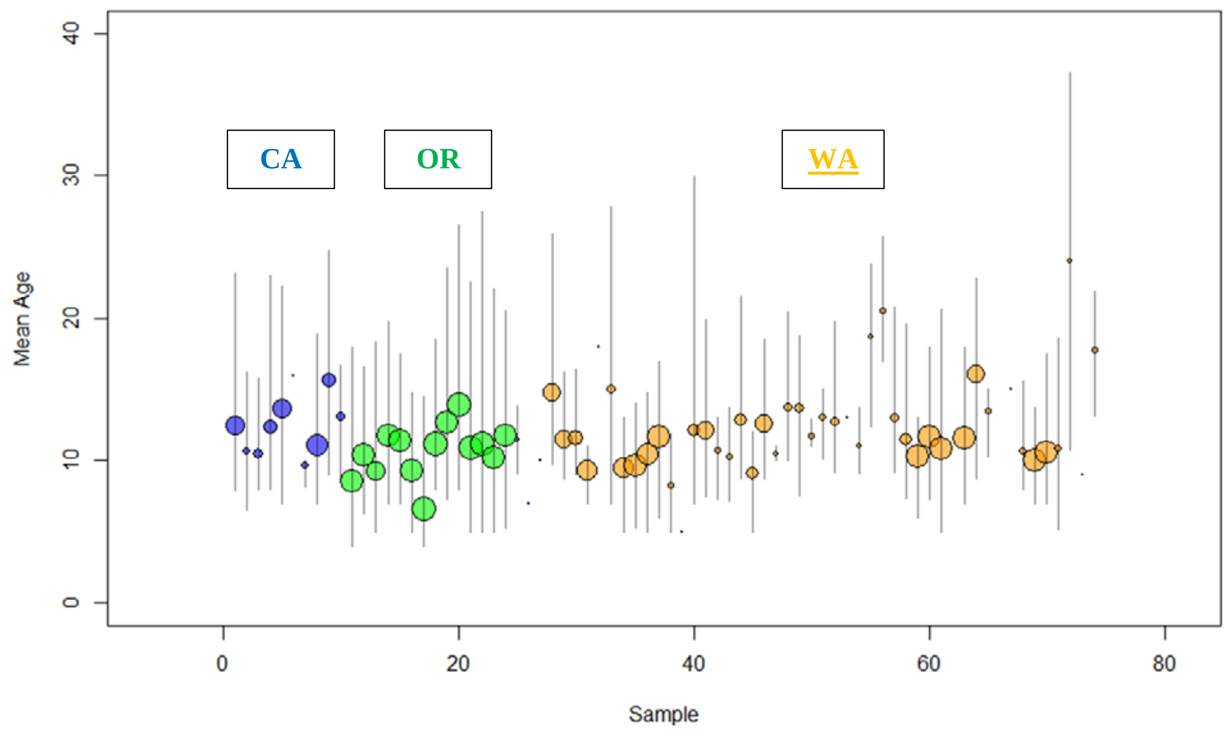
Rationale: Investigate whether this improves fits to the surveys and improves the residual patterns produced by the hake fishery data.

Response: The STAT showed that the model fits degraded with the requested change. The STAR agreed that this change should not be adopted.

Request 2.3: Provide box plots of lengths and possibly weights for individual tows in the NWFSC survey and trips for the bottom and midwater trawl fisheries of compositional data in the NWFSC survey by latitude.

Rationale: To better understand the biomass weighting of the length- and age-compositions.

Response: The STAT produced a series of box-and-whickers plots from the NWFSC survey data of the mean weights of fish per sampled tow, which illustrated patchiness in the sampling, with some tows catching mostly small fish and others catching mostly large fish (see example below of the weight per haul information.). The STAT also produced example plots based on the samples from the commercial fisheries (example below). The take-home message from these results were that when combining samples from different geographic strata, the expansions should be based on the estimated numbers of fish in each strata rather than on the estimated weight of the fish. The average weight is not uniform across the strata.



Request 2.4: Use the bottom trawl CAAL data and only use either marginal length- or age-compositional data for the other fleets. Put all lambdas at 1.

Rationale: Provide CAAL data for a longer period of time and remove double use of compositional data.

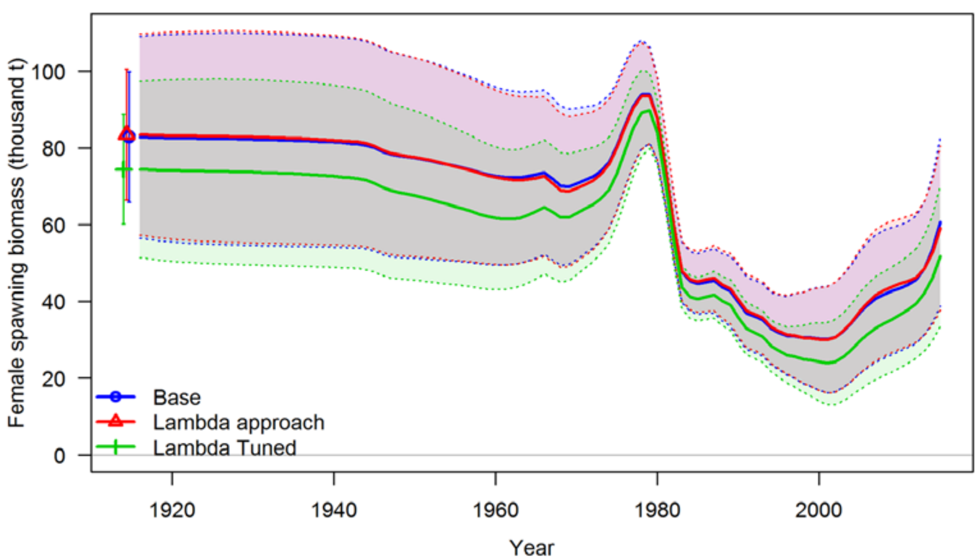
Response: The STAT attempted to satisfy this request by selecting a series of years for which there were good numbers of sampled fish with age-at-length observations and a reasonable balance of samples from different states and fleets. Unfortunately the model was not converging and the results were "insensible".

The goal of this request was to include conditional age-at-length observations for a sufficiently long time period that would allow possible exploration for evidence of temporal changes in growth. However, there were also concerns expressed that purposeful selection of samples was too subjective. An alternative approach would be to randomly choose the years for which the conditional age-at-length observations would be included, or to select one or more fleets and use the CAAL data, but not the lengths

Request 2.5: Compare a lambda weighting approach (as compared to the variance adjustments, which do not tune below a sample size of 1.0) to other approaches for compositional data informed by low sample sizes. Show weighting effects for CAAL data.

<u>Rationale</u>: This alternative weighting scheme will provide an understanding of how sensitive the model is to the base model weighting.

<u>Response</u>: The STAT took the following approach to satisfying the request. (1) The variance adjustment factors (for the age- and length- compositional data) from the (preliminary) base model were converted to lambda scalers. (2) The lambdas in the base model that were 0.5 (because the observations were used twice, in the marginal composition for both age and length) were multiplied times the variance adjustment factors. (3) The model was then tuned. The STAT provided plots of the resulting biomass trajectory (below) and recruitment trajectory (not shown) compared to the original base model results, which demonstrated almost exact correspondence between the two approaches for tuning (variance adjustments versus lambdas). The relevant lines in the graph below are the blue "Base" and the red "Lambda approach". (The green "Lambda Tuned" line used the variance adjustments directly as lambda values, without accounting to the original 0.5 lambda values that were used in the tuning process.) The close correspondence between the Base and Lambda approach trajectories suggests that the compositional data for this assessment were not affected by the Stock Synthesis constraint that small Input-N values be at least one.



Request 2.6: Provide a sensitivity of the net and hook-and-line fishery compositional data on model results by leaving out these data individually and together.

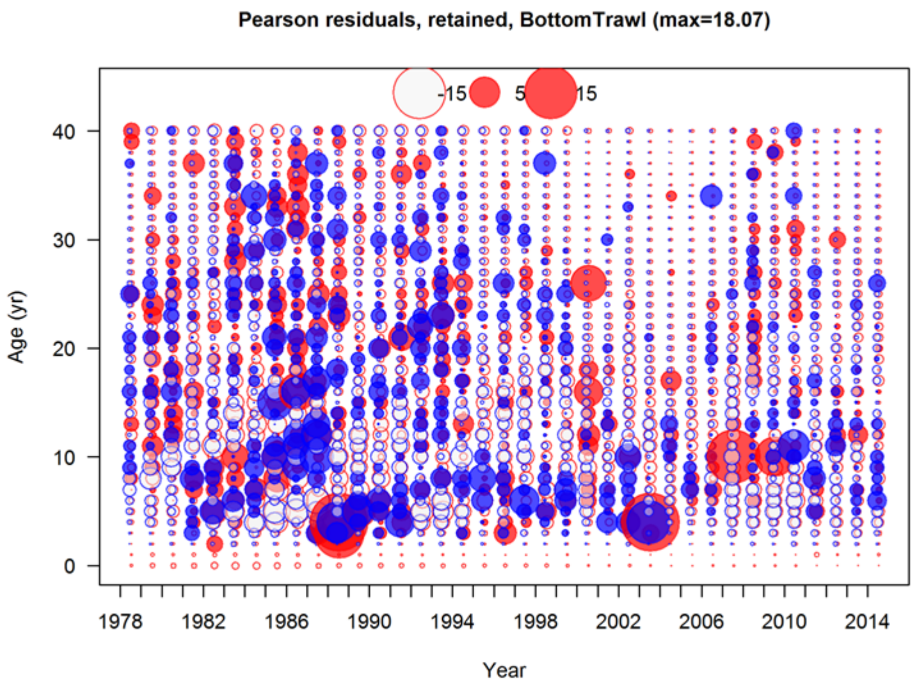
<u>Rationale</u>: Small sample sizes result in large residuals, which may have an inordinately large influence on model results.

Response: The STAT produced a series of plots clearly indicating that the model results were essentially unaffected by the net fishery compositional data, the hook-and-line fishery compositional data, or both together.

Request 2.7: Using the model from request 2.4, increase the lambda on age composition data for bottom trawl and midwater trawl individually and together as a sensitivity.

Rationale: This run was intended to evaluate whether the model could eliminate some of the observed residual patterns with greater weight on these data.

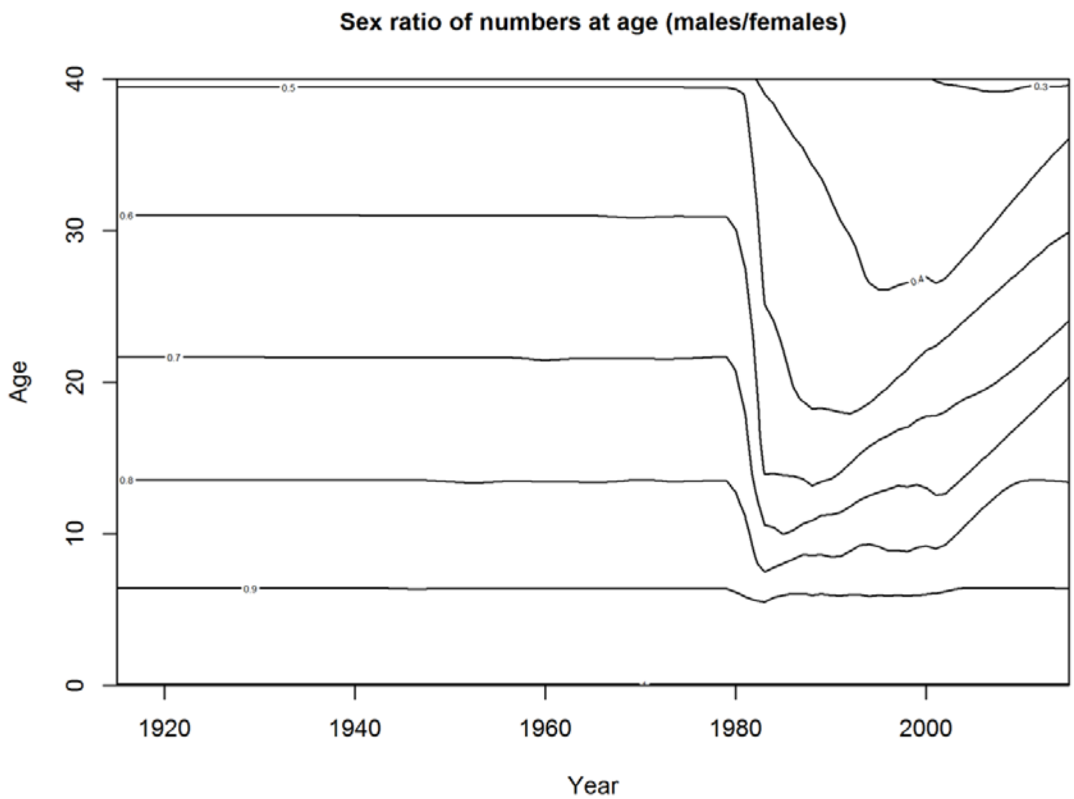
Response: The STAT used the model from request 2.5 -- The STAR Panel agree that this is a better starting place -- and produced a series of bubble plots of the age-composition residuals This procedure did not improve fits to the apparently strong 1970 and 1978 recruitments in this data set, as illustrated in the graph below, which shows residuals from the fit to the bottom trawl age-compositional data when the lambda values for the bottom trawl and midwater trawl fisheries are both increased 5x.



Request 2.8: Provide a comparison of the sex ratio in the early 1980s observations with the current model estimate.

Rationale: To determine if there is evidence in the data of sex ratio changes over time.

Response: The STAT provided histograms of age-composition by sex and plots of observed ratios by age of males to female that indicated a preponderance of females over males at older ages. An r4ss plot from the preliminary base model (below) shows that the model also predicts a tendency for a preponderance of older females (male:female ratio < 1) that becomes accentuated during the period of large removals that occurred during the late 1970s and early 1980s.



Request 2.9: Provide sensitivity runs with the same estimated M for both sexes (using the SS option to set male M to 0.0 without estimation (turn off male prior) – causes the model to use female M for males) and alternatively estimate sex-specific M with sex-specific selectivities for each run.

Rationale: To investigate mechanisms other than M to account for sex ratio differences.

Response: The STAT provided biomass trajectory plots (below) comparing the current base model results with one alternative that had the same estimated M coefficient for both sexes and a second alternative with potentially separate sex-specific M values and sex-specific fishery selection curves. The results are fairly sensitive to how the model is configured to account for sex differences.

Request 2.10: Examine evidence of unavailable spawning biomass to the fishery.

Rationale: To see if the response to request 2.9 results in a greater amount of unavailable spawning biomass in fisheries.

Response: The STAT used newly developed software to plot the amount of spawning output that is less available to exploitation due to the shape of the population-level fishery selection curve, which is comprised of two parts: small fish and large fish. The plot indicated that relatively small proportions of large fish biomass are less available to exploitation.

Because the trawl surveys are assumed to have asymptotic selection, there should be no concern about "cryptic biomass" seen by the model but not by any sampling process.

Request 3.1: Explore numbers-based weighting for survey compositional data for a possible new base model. Construct expanded CAAL compositional data for the NWFSC survey for use in a possible new base model. Weight these data using GLMM biomass by stratum divided by mean fish weight in each stratum.

Rationale: Unexpanded CAAL compositional data are inappropriate when linking these with length compositions.

Response: As requested, the STAT calculated new survey compositional data weighted by numbers rather than biomass. Although the STAT and STAR considered this new approach to be an improvement, it had little effect on the female spawning biomass trajectory.

Request 3.2: Explore CAAL for bottom trawl.

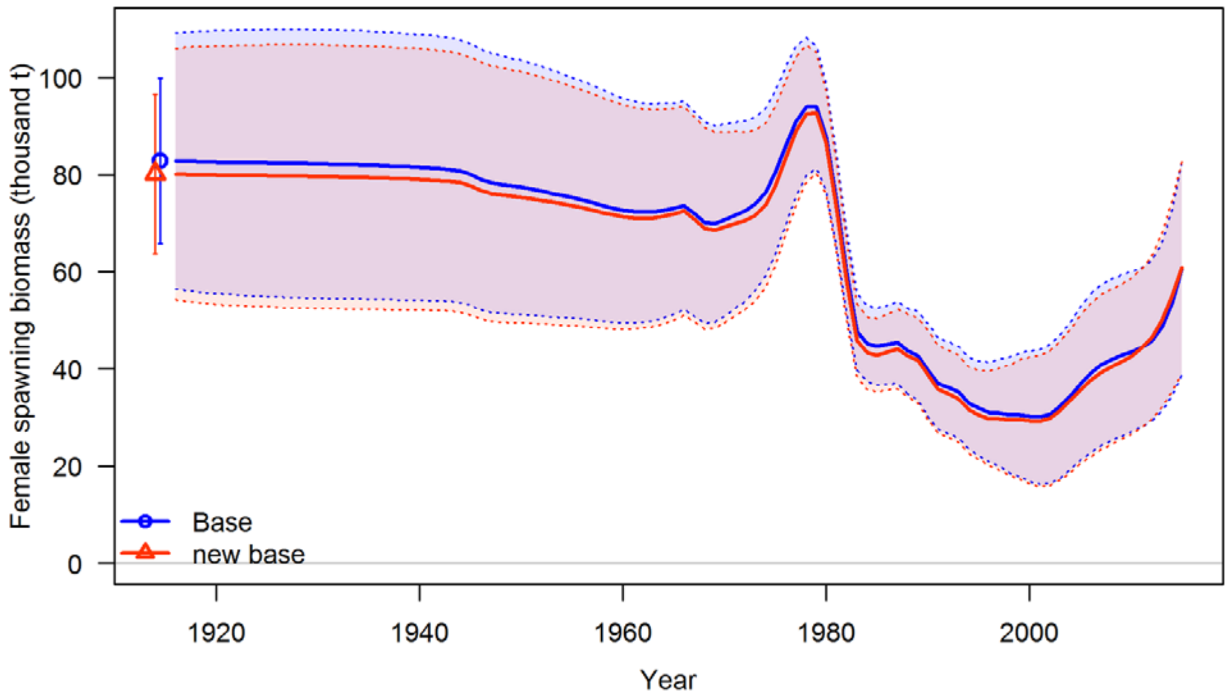
Rationale: Determine whether there are potential changes in growth.

Response: The STAT presented comparison plots of the growth curves (including variability) from the current base model versus one that included conditional age-at-length data from the bottom trawl fleet. The model with the additional CAAL data had reduced variability about the growth curve with almost no change in the average length-at-age. The spawning biomass trajectory was somewhat different from the revised base model (below) and there were changes in some estimates of recent recruitments. The STAR and STAT considered the model with the additional CAAL data to be superior to the current base model. The CAAL compositional data as treated here were based on simple tabulations of fish and did not include any expansions to account for variable sampling fractions.

Request 3.3: If requests #1 and 2 result in new base case, apply harmonic mean weighting and lambda (0.5) weighting sequentially.

Rationale: This is a more logically consistentmethod for weighting comps that allows iterative reweighting, but also retains the downweighting of length and ages collected from the same fish.

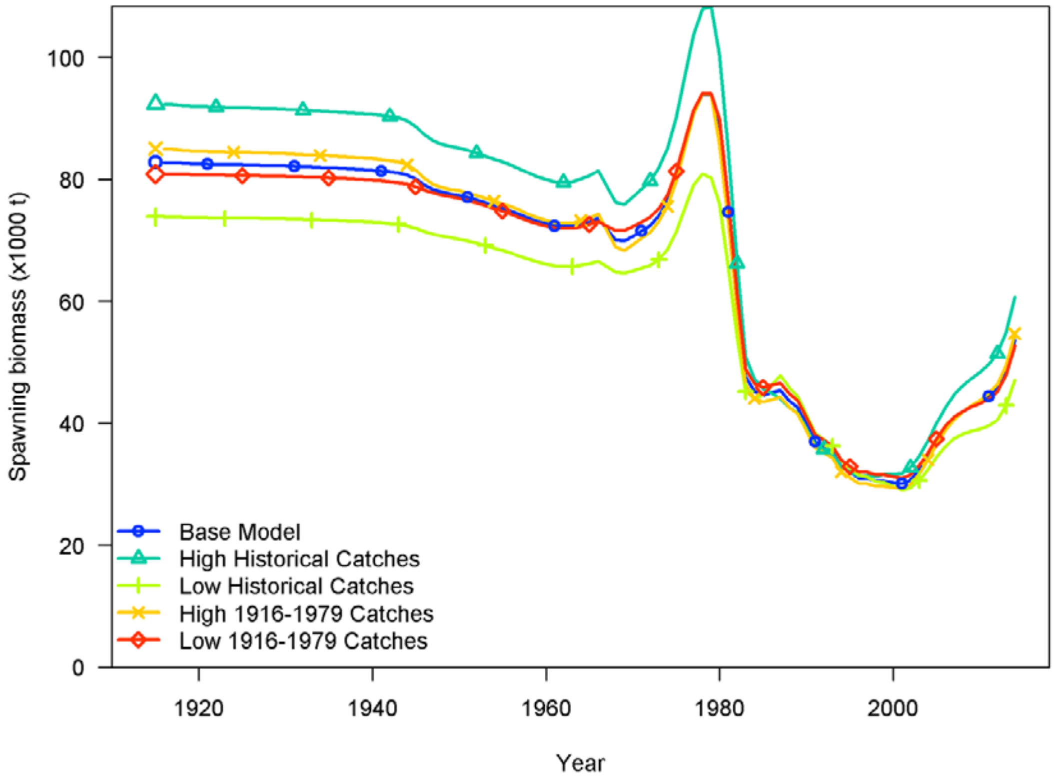
Response: The STAT was uncomfortable about moving forward as requested given the magnitude of the change to the model from adding large amounts of CAAL data from the bottom trawl fishery that had not been properly expanded. Instead the STAT proposed moving forward with a model that (a) used survey length compositional data (Triennial and NWFSC) and (b) CAAL compositional data (NWFSC), both weighted by numbers from the GLMM, and (c) marginal age- and length-compositional data for the fisheries. This new model configuration was tuned using the lambda approach and resulted slight changes from the previous base model (see below). The STAR concurred with the STAT's decision.



Request 3.4: Explore historical catch uncertainty by multiplying catches prior to 1980 by 0.5 and 1.5; multiply catches ±0.25 from 1980 to 1986.

<u>Rationale</u>: These are the periods when catch estimates were more uncertain and this exploration will indicate how sensitive the model is to catch uncertainty.

<u>Response</u>: The STAT produced a plot comparing the spawning biomass trajectories for the new base model and the requested alternatives (below) and tables showing the effects on parameter estimates. In the lines labelled "high/low 1916-1979 catches" the catch values were multiplied by 1.5 (high) and 0.5 (low). In the lines labelled "high/low historical catches" the catches during 1980 to 1986 were multiplied by 1.25 (high) and 0.75 (low), in addition to the high/low alterations of the 1916 to 1979 catches. The uncertainty in spawning biomass due to uncertainty in historical catches appeared to be driven by the catches by midwater and bottom trawl during 1980 to 1986, which was the period of steepest decline in spawning biomass, rather than by the earlier years. There was speculation that the large increases in bottom trawl catches of widow rockfish during the 1980s could be an artefact of limited species composition sampling rather than increased targeting of widow rockfish with bottom trawl gear. It seems that uncertainty in the large catches by midwater trawl are unlikely to be particularly large and that other axes of uncertainty will be of greater importance.



Request 3.5: Once a new steepness prior is available, range steepness about the prior using the prior distribution and run the models estimating M. Alternatively, think about bivariate axes of uncertainty by crossing higher steepness and lower M (and vice versa).

Rationale: This exploration may define an axis of uncertainty for a decision table.

Response: The STAT reported that the steepness prior without information from previous widow rockfish assessment had a mean value of 0.798, standard deviation of 0.132, median value of 0.821, 12.5th percentile of 0.632, and 87.5th percentile of 0.943. The STAT produced a plot of female spawning biomass based on runs with steepness (h) estimated and with it fixed at the 12.5 and 87.5 percentiles. The spawning biomass trajectory was relatively insensitive to different values of steepness. The STAT also explored using a bivariate axis of uncertainty based on h and M but the high state of nature was ill-behaved. The STAT proposed finding the combination of female-M and h that results in a high state of nature that matches the 87.5 percentile of 2015 female spawning biomass. The STAR Panel agreed that the results using the marginal distributions might be too broad and poorly behaved, but suggested exploring other ways to capture some of the uncertainty in M and steepness, as well as contrast in the projected trend via the strength of the 2010 year class.

Request 4.1: Clean-up for final base model: a) Remove the 8 cm fish from the length comps.; b) use the marginal ages and lengths (not CAAL) for the fisheries since no CAAL comps. are available that are expanded to account for differential sampling fractions; c) exclude the marginal survey age comps.; and d) tune sigma-r if it seems necessary.

Rationale: These are agreed to changes for the final base.

Response: The STAT accomplished the requested changes. The 8cm fish were not removed from the model because the STAT discovered they had been inadvertently displaced during compilation of the conditional age-at-length compositional data.

Request 4.2: Use the mean of the steepness prior without the 2011 widow steepness distribution influencing the estimate of the prior as the fixed base case steepness value.

Rationale: This is consistent with the concept of developing and using priors as well as past practice.

Response: The STAT accomplished the requested change and fixed the main period of recruitment deviations to begin in 1970. Also, the model was retuned.

Request 4.3: Explore two possible axes of uncertainty: a) ranging M, and b) ranging the strength of the 2010 year class for a candidate decision table. Do these runs sequentially starting with “a” to decide whether the decision table should be informed by a univariate or bivariate uncertainty.

Rationale: Natural mortality is an important axis of uncertainty but may not provide adequate contrast in projected trend. The strength of the 2010 year class is highly uncertain and important to projections. A bivariate decision table may provide adequate contrast.

Response: The STAT accomplished this final request during the STAR Panel meeting, and also added a high and low value for steepness to the cases. The STAT and STAR panel agreed that all three sources of uncertainty were important, and that this approach did not overstate the uncertainty in the assessment results (it only coincidentally approximated the 12.5^{th} and 87.5^{th} percentiles of the asymptotic distribution for spawning biomass). Further, the contrast in steepness, M and the 2010 year-class provided not only contrast in the current stock size, but also in the future trajectories contributing to the projections and decision table. There was insufficient time to produce the full decision table during the review, but the STAT will do so following the meeting.

Description of the Base Model and Alternative Models used to Bracket Uncertainty

The base model was a single area, single growth-morph, two-sex, with recruitment deviations starting in 1900, catch starting in 1916, and steepness fixed at 0.793 (the mean of the prior without widow rockfish) and sigma R at 0.6. Maximum age was 40 years with a plus group. Growth and natural mortality were estimated, with a prior on natural mortality based on a maximum age of 54. A combined maturity ogive was used based on combining two alternative maturity ogives built from California and Oregon data respectively. Five fishing fleets were defined: bottom trawl, midwater trawl, hake fisheries, hook & line and net. Three survey indices were used: a juvenile index, the triennial survey and the NWFSC shelf/slope survey, all standardized using GLMM. The conditional age-at-length compositional data were used only with the surveys, and marginal age- and length-compositional data were used with the fishery fleets. Seven abundance indices were developed from the survey and fishery data.

Indices were initially weighted with an input variance parameter. Length- and age-compositional data were tuned using the harmonic mean approach. A standard iterative re-weighting procedure was applied to ensure weights on the different data sources within the model were internally consistent. Samples used in both marginal length- and age-compositional data were further weighted with lambdas of 0.5 after the iterative re-weighting to account for non-independence of these data.

The models used to bracket uncertainty set the sex-specific natural mortality and steepness and the 2010 recruitment together at low (12.5^{th} percentile) and high (87.5^{th} percentile) values based on estimates of their probabilities. For natural mortality and the 2010 recruitment the estimated standard error was used assuming they are close to normally distributed. For steepness, the prior excluding the likelihood on widow rockfish was used to provide this range.

Technical Merits of the Assessment

The assessment applied the full abilities of SS3 and made good use of the available data. The STAT used an innovative approach for weighting and combining the trawl survey compositional data from different strata. The model data and alternative model structures were thoroughly explored and the base model well justified. Full sets of diagnostics were made available. The STAT was fully responsive to requests from the STAR.

Important improvements were made to this assessment, which could be considered in other West Coast assessments:

- The fleet structure (collapsed across states) seemed to be parsimonious while still capturing important gear differences.
- The weighting procedure for combining survey compositions was initially based on GLMM estimated biomass, but changed to GLMM estimated numbers, which seemed superior to other methods based on survey areas. It was agreed that future implementation should be based on numbers rather than biomass.

Technical Deficiencies of the Assessment

Overall, there were no serious technical deficiencies with the assessment. Although some patterns in the residuals for age, length and abundance indices were present, these may be due to inconsistent trends within and between different data sources.

In the absence of a clear default method for weighting different data sources, this remains a source of uncertainty. Procedures were agreed between the STAR and STAT teams which minimized the problems that had been identified, but further work outlined as recommendations below could refine these solutions further.

Areas of Disagreement Regarding STAR Panel Recommendations

Among STAR Panel members (including GAP, GMT, and PFMC representatives):

None.

Between the STAR Panel and the STAT Team:

None.

Management, Data, or Fishery Issues Raised by the GMT or GAP Representatives During the STAR Panel Meeting

None.

Unresolved Problems and Major Uncertainties

A number of problems remain unresolved:

- The available indices may not track widow abundance well. No direct account is taken of the semi-pelagic behavior of the species, which at best, greatly increases the variance of the indices. Development of an alternative acoustic survey may be worthwhile.

- There is no information in the data on steepness. The prior was used to provide a range for sensitivity runs. This prior is based on a small number of stocks and has changed substantially over recent assessment cycles. This prior and its use as a fixed input to stock assessments is a major source of uncertainty in the assessment.
- Some catches since 1982 remain uncertain, and potentially biased. Some may have been mis-specified, and Canadian catches are not available. However, unlike some other assessments, results are insensitive to errors in catches before 1980.
- The weighting for composition data within the stock assessment can be based on different methods, such as the effective sample size using the harmonic mean or the Francis (2011) methods. For this assessment, the harmonic mean was use. The "best" way to approach weighting compositional data remains unresolved for all stocks.
- Survey composition data are being raised and weighted in this assessment based on biomass estimated within the standardization GLMM which are then converted to numbers. While the approach was supported by the STAR panel, a consistent protocol to carry out this weighting procedure for raising samples to population composition estimates needs further development.
- A temporary solution was found to ensuring random age samples as marginal distribution could be used to identify cohorts in the population while limiting the double-counting that arises when including those same samples in the condition age-at-length data. However, it is likely that this approach could be improved.
- The different maturity ogives estimated affect the assessment results. Differences between the estimated maturity ogives could be due to genuine population differences or they could be sampling artefacts. Further sampling would be required to resolve this issue.

These are long term issues, which have also been raised in previous assessments. Although the unresolved problems increase uncertainty, they do not prevent the stock assessment being used for management advice.

The major uncertainties are:

- Natural mortality has been estimated from the data and is higher than expected, although consistent with the prior. While the estimated values for natural mortality are consistent with the observations, results remain sensitive to the natural mortality estimates.
- Recent recruitment, particularly the 2010 year class, has been estimated to be high. The most recent recruitments have been infrequently observed and are not well estimated.
- The stock recruitment relationship, and in particular steepness, is unknown.

Recommendations for Future Research and Data Collection

Specific recommendations for the next widow rockfish assessment

- The next iteration of this assessment should be an update assessment.
- Minor anomalies in the weight-length data from the PacFIN Biological Data System (BDS) should be excluded or reconciled.
- A reanalysis of the foreign at-sea index that best overlaps the period of largest stock decline could be conducted before the next assessment. In particular, an analysis should consider effort measures that include search as well as towing time, given the schooling nature of this species. Other fishery indices are unlikely to have an appreciable impact on the results and may not be worth reanalyzing.

- Widow rockfish should be considered in any future discussions about trans-national stocks. Although a joint assessment with Canada may be difficult to arrange, it should be explored. It is possible that lack of information from Canada affects estimates of productivity and, in particular, steepness. Until such time as a joint assessment can be conducted, evaluation of relative catches and trend information on abundance in Canadian waters would also be helpful. Potential exchange also clouds the clear interpretation of what represents steepness for this stock.
- Updated maturity data representing the current stock distribution should be collected and analyzed, preferably using histological methods.
- Since there was so little information in the data on steepness, the informative prior might be strong enough to allow for estimation in future assessments. This should be explored.
- Based on the variability estimated for the juvenile index, it should be removed from future analyses unless it can be improved and validated. Specifically, the estimated variance is greater than the RMSE of the recruitments, so it will add more noise than signal at the end of the time-series when there are no other data to inform recruitment. This decreases the predictive ability of the model.
- Although recreational removals are low in relation to other removals for this stock, these should at least be reported in a table for comparison in future documents.
- It may improve the model if the H&L and NET fisheries are combined with other fleets, as these represent very little removals and noisy data. Removals of these data did not appreciably change the results for this assessment and their selectivity showed similar patterns to other fleets. Removing these as separate fleets would likely to make the modelling simpler with no loss of signal.
- Select one or more fleets (as run-time allows) and create conditional age at length data in order to inform growth and selectivity from more than just the most recent years where survey data are available.

General recommendations for all assessments

- A specific data workshop, perhaps for all species prioritized for assessment, could examine information across a broad range of species due for assessment, and would also assist with the development of more specific documentation of protocols used to compile best available data sets for stock assessment, continue acceptance of agreed procedures for standardization of abundance indices, and also begin work on procedures for the development of alternative data series that capture uncertainty, particularly for historical catch and discards.
- Additional work is required to further develop an objective procedure for evaluating the chosen stock boundaries across all rockfish (and potentially all other) assessments may be beneficial, and also more directly point to required directions for future research or assessment collaboration across national/international political boundaries. Further investigation is required for whether the stock boundaries assumed in the assessment are appropriate for management as well as scientific assessment.
- There is a need for more detailed examination of input data independent of the stock assessment, particularly in relation to sample size and representativeness. An examination of data sources by year and sub-area in particular may suggest appropriate methods for post-stratification of composition data. Potential stratification that should be considered should include season, latitude, depth, and boat type.
- Reports should include a section on how the recommendations from any data workshop and previous CIE reviewers were addressed. This would be an extension of the section for addressing previous STAR Panel recommendations (Appendix B in the 2014 Terms of Reference).

- Additional work is required in developing catch histories. An evaluation of the plausible range of proportions of species in the aggregated catch on the reconstructed catch time series is recommended. It would be most useful not only to provide single best estimates, but to define ranges suitable for use in bracketing uncertainties and sensitivity analyses.
- The state of Washington still needs a formal catch reconstruction to standardize approaches across assessments and ensure the best available estimates are being used.
- An objective procedure for identifying sub-model error structure (usually gamma or lognormal here) is required for delta-GLM procedures. Consistency is required for the model selection process, preferably using *a priori* candidate models rather than a stepwise selection. The standard delta-GLM procedure should allow for different factors to be considered in the binomial and sub-models. A standard set of diagnostics should be provided to review panels for each abundance index including: plots of the binomial and positive catch rate year effects in addition to the combined year effects; plots of all estimated effects; production of year:area interactions. The effects of the standardization on the "nominal" or unstandardized indices should also be shown and explained (i.e., which variables have caused a shift in the trend). It may be far more efficient to produce and review this output for groups of species together, rather than try to include it in each species-specific review.
- Reporting the extreme catch encounter probability in the ECE models would be an interesting diagnostic and additional piece of information for understanding how frequent exceptional catches are estimated to be.
- Expand and weight conditional age at length data to accurately represent both the sampling process and the numbers of fish predicted in each strata. This achieves logical consistency among data sets, choices of the number of length bins, and imparts the greatest amount of orthogonal information possible to the assessment model. Where length sampling is random, marginal length data should be associated with conditional age at length data (by year and fleet) such that the recruitment information contained in the ages is not lost relative to what would be included if using marginal age compositions.
- An objective procedure using maximum age for a natural mortality prior needs to be developed and fully explored. Specifically, unless age samples were collected prior to significant exploitation, the oldest fish may be missing from the observed data, ageing error will tend to bias the maximum age to higher values and there may be sampling bias (e.g. domed selectivity). A percentile based method (e.g. use the 90^{th} percentile) is likely to be more robust than using the single maximum age. Other information (e.g., GSI, growth, exploitation rates etc.) should be included, where possible in the derivation of the *M* prior.
- Where there is significant uncertainty in a very recent recruitment estimate (informed by very little data), it may be helpful to perform a likelihood profile over the strength of that year-class (running the model adjusting that deviation in the par file and using a "noest" option) to see where the primary signal was coming from.
- Aggregated residual plots (weighted and combined across all fleets would help to understand whether the model is fitting the available data adequately, even where patterns in residuals might show trade-off among fleets. This could reduce the over interpretation of residual patterns within the fit to a single fleet.
- Where there are marginal age- and length-compositional data being used from the same fish, iterative reweighting should be done first, then at the final stage an additional multiplier of 0.5 should be applied and no additional iteration performed. This retains the goals of both logically consistent reweighting and down weighting the doubly-used data.

- MCMC results are a useful tool to measure uncertainty and diagnose problems in the assessment as well as provide an alternative to MLE-based results, which can differ appreciably in terms of point estimates and uncertainty (Stewart et al. 2013). MCMC should be routinely used and reported where possible.
- Producing at least one model run with the full time-period of estimated recruitment deviations would be a very helpful diagnostic and could be a plausible base case depending on the model behavior.
- Triggering a future full assessment could be based on monitoring the most reliable indices, such as surveys with confidence bounds. In comparing observed and predicted values, values outside confidence limits may suggest a higher priority for more immediate assessment.

In addition to current R4SS and SS3 functions, the following additional features and standardized procedures should be developed:

- Procedures for examining sources of information on recruitment events is required. This could include profile over recruitment events or partition likelihood components.
- A method to examine observed and expected sex ratio by age and through time would resolve questions about the consistency of sex ratios being produced for the modeled population.
- Developing residual plots that are weighted across data sources would allow comparisons to be made that might help to identify common patterns.
- Removal of the re-scaling to 1 problem after weighting is applied to composition data
- Development of standard procedures for the selection of the most appropriate weighting system that should be applied to input data (additional sd for indices, harmonic mean/Francis/other for length and marginal age comps, harmonic mean/Francis/other for conditional age-at-length data).

Acknowledgements

The STAR Panel thanks the public attendees, Maggie Somers, and the STAT, GMT and GAP representatives.

References

Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* **68**: 1124–1138.

Minte-Vera, C. V., T. A. Branch, I. J. Stewart and M. W. Dorn. 2005. Practical application of meta-analysis results: avoiding the double use of data. *Canadian Journal of Fisheries and Aquatic Sciences* **62**: 925-929.

Stewart, I. J., A. C. Hicks, I. G. Taylor, J. T. Thorson, C. Wetzel and S. Kupschus. 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fisheries Research 142(0): 37-46.

Lorenzen, K. 2005. Population dynamics and potential of fisheries stock enhancement: practical theory for assessment and policy analysis. Philosophical Transactions of the Royal Society of London. Fisheries Theme Issue 2004