INDEPENDENT PEER REVIEW REPORT for DOVER SOLE and SPINY DOGFISH STAR PANEL

Matthew D. Cieri

Center for Independent Experts

August 2021

EXECUTIVE SUMMARY

The first-ever online STAR Panel met from May 3 to May 7, 2021, to review the Dove sole and Spiny dogfish assessment produced by two different STAT groups. Both STAT groups did very well, presenting information to the Panel in an intelligible and easily digestible way. The meeting itself was well run with good organization of the meeting structure as well as the documents and background material.

Both models used the latest version of Stock Synthesis version 3.30.16 and employed sex-specific growth. This is where the similarities ended, however, as the models took different approaches and had varying degrees of data available.

Dover sole was the first stock assessment reviewed. It was last reviewed by the STAR in 2011. Important changes included the use of the “Francis method” for weighting, a change in how natural mortality was estimated for males and fixed for females, a simplified fleet structure by collapsing data and catches from Oregon and Washington into a single fleet, as well as the treatment of select biological parameters: natural mortality rate (M) and the maturity-at-length.

A total of five requests were made by the Panel, with each done satisfactorily by the STAT. The model appeared relatively robust to most of these requests, and overall the Panel and this reviewer were happy with the base model being used to provide management advice. Overall, strengths included a well-done bridging and sensitivity analysis as well as the overall model diagnostics.

There were a few problems, however, most notably the high level of cryptic biomass due to the stock not being available to either fishery or surveys. Additionally, there was a lack of aging data from the California fishery and elsewhere after 2009/2010, as well as tension in the model's different data elements that pushed it in an unfeasible direction for natural mortality. There are also issues with lack of knowledge surrounding the stock’s distribution and problematic fits to the survey data in the most recent years.

Multiple research recommendations were put forward to address some of these issues. In particular, some method of examining the voracity and magnitude of the cryptic biomass is likely to be the most important. Studies should also be conducted on catchability, distributions of the stock, as well as increases in fishery sampling; as recommended.

The second stock to be reviewed was Spiny dogfish. Like Dover sole, it had made vast improvements since it was last reviewed in 2011. These included estimations of growth in the model, changing the fecundity given recent research on the length of gestation for this stock, using aging data in the model, the estimation of discards in the 1960-2002 fishery by examining effort in the Sablefish fishery, and the use of VAST. Important strength included all of the previously mentioned improvements which likely contributed to improvement in the retrospective pattern seen in the 2011 assessment.

A total of 12 requests were made by the Panel of the STAT, each was done satisfactorily. Given the model's performance, diagnostics, and sensitivities, the Panel, and this reviewer were happy with the base model and recommend it to provide management advice, though not without some reservation.

The issue of worn spines on older and larger fish meant that fish >80 cm could not be used in the growth model, potentially biasing the results. The model also had tension within its data streams, with abundance information from the indices pulling the model one way while the length data from those

same indices pulling the model in the other. This model is very sensitive to assumptions on natural mortality, which further compounds the aging uncertainty as max-age is one of the best ways to ground truth natural mortality. Additionally, there were concerns dealing with the overall estimate of discards in the 1960-2002 Sablefish fishery and how it was estimated.

Like with Dover sole, but perhaps more importantly here, there is a lack of knowledge on the distribution of this stock particularly in relationship to Canada. This, as mentioned in the report, can have dramatic implications for model success and should be a focus of ongoing research. A multitude of other research recommendations was made. These include improved sampling, as well as more intensive work on the issue with aging. Catchability studies would also be important to conduct as these may shed light on proper bounds within the model to estimate other parameters. Likewise, a meta-study of natural mortality could also be important. Each of these, in this reviewer's opinion, should be conducted before reviewing this model again.

After reviewing both stocks, some take-home messages include concerns that uncertainty might be inflated to account for other "unknown" uncertainties, as well as a general pattern that distributions and movements of both stocks are not well understood; fueled in part by reduced sampling and lack of collaboration with Canada. More importantly, the analysis conducted by both STATs dramatically improved the robustness of these assessments when compared to 2011.

Overall, this review was an enjoyable experience; the information was well organized, the TOR reference document was clear (and very well done), and there were few technical problems given it was the first online meeting for the STAR. That said, while more convenient, cheaper, and safer than an in-person meeting, there was certainly something lost by not being in the same room as other participants. While fine for this STAR given the state of the world, this reviewer is uncertain if an online-only meeting to review stocks should continue.

INTRODUCTION

The STAR panel for Dover Sole (*Microstomus pacificus*) and Spiny Dogfish (*Squalus suckleyi*) was held via online meeting from May 3 to May 7, 2021. The Panel operated under the Pacific Fishery Management Council's (PFMC) Terms of Reference for the Groundfish and Coastal Pelagic Species Stock Assessment Review Process for 2021-2022. A complete review is provided below (See Conclusions) on how this process played out in the first online STAR Panel meeting. Participants (Listed in Appendix X) included two CIE reviewers, an outside member from Oregon State University, and the Chair who also serves on the SSC (Science and Statistical Committee).

Because both stocks were reviewed independently, this report has been subdivided into two sections: A for Dover Sole, and B for Spiny Dogfish. Each of these includes subsections on Model formulation, Request for the STAT (Stock Assessment Team), the Strengths and Weaknesses of the approach used, Research Recommendations, and Conclusions particular to this assessment.

Section A: DOVER SOLE

Introduction and model structure

Dover Sole is found from Baja California to the Bearing Sea and beyond to the Aleutian Islands. A typical flatfish, Dover sole is benthic in orientation and lives to at least 69 years. Dover sole has a complex migration pattern, moving deeper as they grow larger and older but also shifting from inshore feeding grounds to off-shore spawning habitat. However adult movement is limited, and as such stock substructure cannot be ruled out. Despite this, the stock is characterized as a unit stock within US waters, in part due to their long larval dispersal.

Approach:

The STAT used Stock Synthesis, version 3.30.16 to produce the assessment as a two-sex model, given the dimorphic growth by sex in this species. The model period started in 1911 and went through 2020.

Overall, the model structure was similar to the 2011 assessment except:

- Used the double normal selectivity parameterization for the fishery, Triennial Survey, and West Coast Groundfish Bottom Trawl Survey (WCGBTS) fleets in the model where the female sex-specific selectivity parameters were estimated as full offsets with a scale parameter relative to the male selectivity (offset parameters for the peak, ascending width, descending width, final selectivity, and a scale parameter).
- Selectivity of the NWFSC and AFSC Slope Surveys were modeled using a cubic spline selectivity form, the same as the 2011 assessment.
- Simplified the fleet structure by collapsing data and catches from Oregon and Washington into a single fleet.
- A minor change in parameterization from the 2011 assessment was the change in estimating male biological parameters (natural mortality, growth) as offsets from the female parameters.

- Data weighting approaches and applications have evolved considerably since 2011 when the last assessment of Dover sole was conducted. The base model was weighted using the "Francis method", which was based on equation TA1.8 in Francis (2011).
- The final major changes relative to the 2011 assessment was the treatment of select biological parameters: natural mortality rate (M) and maturity-at-length. The maturity-at-length was updated based on new research conducted by Melissa Head (NOAA, NWFSC).
- The method of developing an M prior has changed since the last assessment conducted in 2011. The current approach used for stock assessments of West Coast groundfish is based on Hamel (2015). Additionally, this assessment did not estimate female M and fixed the parameter at the median of the prior, 0.108 yr-1 compared to the 2011 assessment which estimated both female and male M directly.

Data elements:

The assessment used a plethora of data sources including landings data and discard estimates; survey indices of abundance, length- and/or age-composition data for each fishery or survey (with conditional age-at-length data used for the surveys); information on weight-at-length, maturity-at-length, and fecundity-at-length; information on natural mortality and the steepness of the Beverton-Holt stock-recruitment relationship; and estimates of aging error.

Parameterization:

The model was parameterized as outlined in Table A.1 (Below).

Table A.1: Specifications and structure of the base model.

Model Structure	Base Model
Starting year	1911
Population characteristics	
Maximum age	60
Gender	2
Population lengths	5-60 cm by 1 cm bins
Summary biomass (mt)	Age 3+
Data characteristics	
Data lengths	8-60 cm by 2 cm bins
Data ages	1-60 ages
Minimum age for growth calculations	1
Maximum age for growth calculations	60
First mature age	0
Starting year of estimated recruitment in main period	1975
Fishery characteristics	
Fishing mortality method	Hybrid F
Maximum F	3.5
Catchability	Analytical estimate
CA Trawl Selectivity	Double Normal, Female Offset
OR/WA Trawl Selectivity	Double Normal, Female Offset
AFSC Slope Survey	Cubic Spline, Male Offset
Triennial Survey	Double Normal, Female Offset
NWFSC Slope Survey	Cubic Spline, Male Offset
NWFSC WCGBT Survey	Double Normal, Female Offset
Fishery time blocks	
CA Trawl Selectivity	1911-1984, 1985-1995, 1996-2020
CA Trawl Retention	1911-1947, 1948-2010, 2011-2014, 2015-2020
OR/WA Trawl Selectivity	1911-1984, 1985-1995, 1996-2020
OR/WA Trawl Retention	1911-2001, 2002-2010, 2011-2020

Sensitivities:

The STAT ran a wide array of sensitivities including:

Data Sensitivities

- Remove California commercial length data
- Remove Oregon/Washington commercial length data
- Remove AFSC Slope Survey length data
- Remove Triennial Survey length data
- Remove NWFSC Slope Survey length data
- Remove WCGBTS length data
- Remove California commercial age data
- Remove Oregon/Washington commercial age data
- Remove NWFSC Slope Survey age data
- Remove WCGBTS age data
- Remove AFSC Slope Survey index
- Remove Triennial Survey index
- Remove NWFSC Slope Survey index
- Remove WCGBTS index

Structural Sensitivities

- Estimate Lorenzen natural mortality (M) by sex with the change of M occurring at approximately age at-50-percent maturity of 10 years old 35
- Allow the model to estimate female M
- Fix M values by sex at the estimates from the 2011 assessment of Dover sole
- Fix M values for each sex at the median of the prior distribution (0.108)
- Mirror the selectivity by sex for the California and Oregon/Washington fishing fleets
- Assume the 2011 parameterization of fishery selectivity
- Assume the 2011 parameterization of survey selectivity
- Fix the NWFSC Slope Survey female selectivity to asymptote at 1.0
- Assume the 2011 length-at-maturity estimate
- No estimation of recruitment deviations
- Use of the McAllister and Ianelli method for data-weighting
- Use of the Dirichlet Multinomial method for data-weighting
- Fix the extra standard deviation parameters for the AFSC Slope Survey, NWFSC Slope Survey, and Triennial Survey at a near negligible level (0.01)

In addition, the STAT ran a retrospective analysis that sequentially removed one year of data from 2020 to 2010.

Overall, model performance to these sensitivities was as expected. Of particular usefulness was the extensive “Bridging Analysis” where data streams were reverted sequentially (or removed) to the 2011 assessment. Overall, there were few convergence issues. These sensitivities and their implications will be discussed further (below).

Requests

A total of five requests were made by the Panel to the STAT. Below are those requests, the rationale, the responses, Panel comments, and this reviewer's comments.

Request No. 1: Provide historical catches by state (this can be provided after the STAR panel).

Rationale: There is a request to show these catches for the record and to assist in a future catch reconstruction in WA.

STAT Response: An excel file "Dover_sole_catches_by_state.xlsx" provides the input state-specific catches used in the model. The worksheet titled "Catch by State" are the fully processed catches incorporating all historical reconstructions, PacFIN catches, and any adjustments required (e.g., fish landed in California from Oregon and Washington waters). The worksheet "CA Hist Catch to ORWA" provides the total landings identified by Don Pearson (SWFSC) from 1948 - 1968 that were excluded from the California catch reconstruction, because the catch area was identified to either be in Oregon or Washington (provided by John Field, SWFSC). Catch history by state will be included in the revised assessment document.

Request No. 2: Investigate a time block for CA selectivity - explore 2011 (IFQ implemented) and 2003 (RCAs implemented).

Rationale: To attempt a better model fit to the CA composition data.

STAT Response: Three runs were conducted that explored additional blocks in the California fleet selectivity: 1) add a block from 2003 - 2020, 2) add a block from 2011 - 2020, and 3) add two blocks 2003 - 2010 and 2011 - 2020. The estimated selectivity curves by sex for each of these runs are shown in Figure A.1 below.

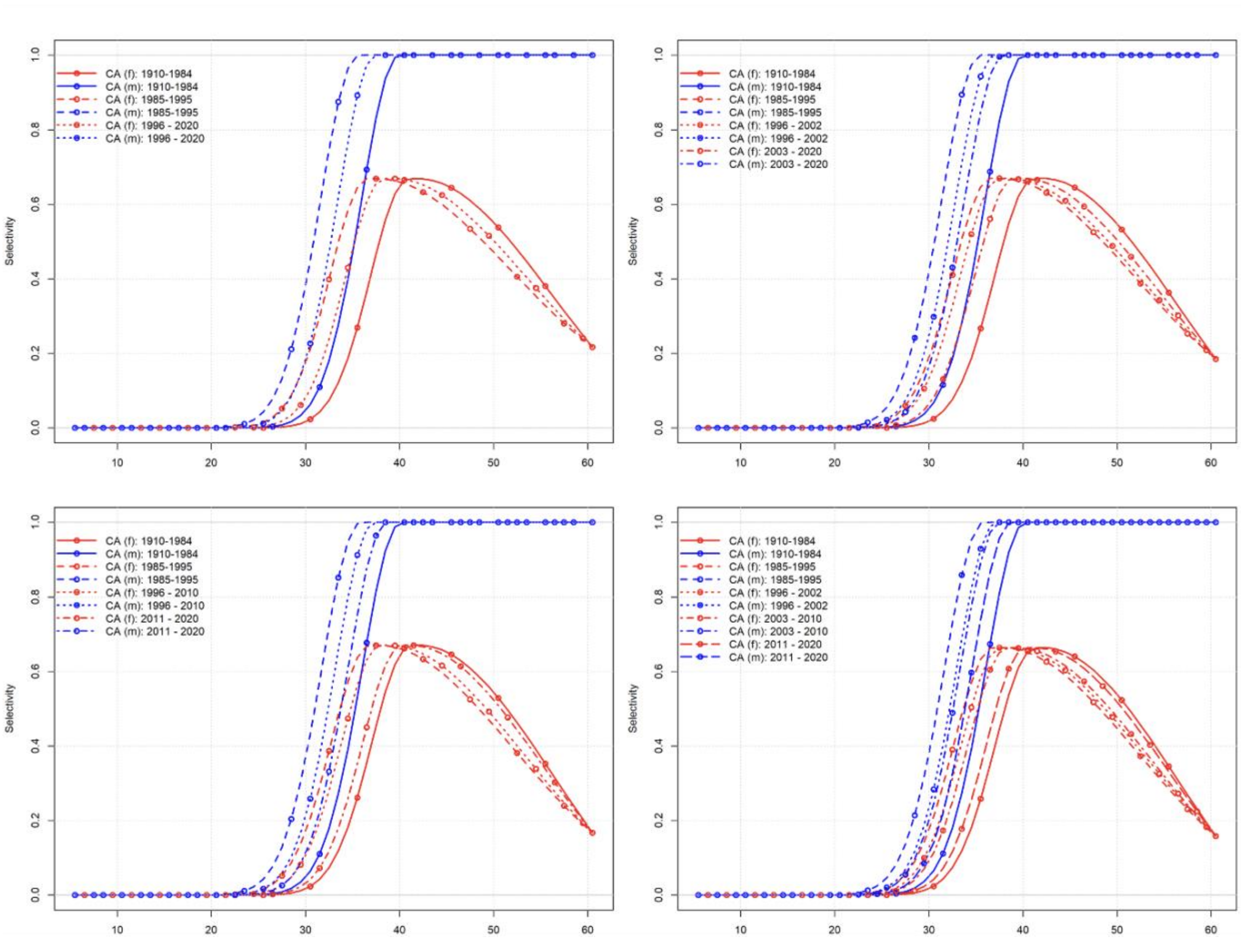

Figure A.1: Estimated sex-specific selectivity curves for each of the alternative blocking for the California fleet. The top right panel is the base model, the top left is the sensitivity with a block from 2003 - 2020, the bottom left is the sensitivity with a block from 2011 - 2020, and the bottom right is the sensitivity with a block from 2003 - 2010 and 2011 - 2020.

Selectivity for the sensitivity runs that either applied a 2003 - 2020 or a block from 2011 - 2020 each had an estimated right-ward shift (selecting slightly larger fish) in selectivity for both sexes relative to the selectivity estimated from 1996 - 2002 or 1996 - 2010. The estimated length at peak selectivity for each block is provided below in Table A.2.

Table A.2: Parameter estimates of the length at peak selectivity by sex.

	Base Model: 1996 - 2020	**2003 - 2020**	**2011 - 2020**	**2003 - 2010 and 2011 - 2020**
Length at peak selectivity for males (cm)	37.4	37.9	38.6	37.5 38.7
Length at peak selectivity for females (offset, cm)	0.9	1.1	1.6	0.8 1.5

The predicted fit to the mean length by year for the California fishery lengths for each sensitivity is shown in Figure 2 below. The sensitivities that applied a selectivity block from 2011 - 2010 (including the sensitivity with two blocks: 2003 - 2010 and 2011 - 2020) to the California fleet appeared to have the best visual fit to the increase in mean lengths in the final years of the model.

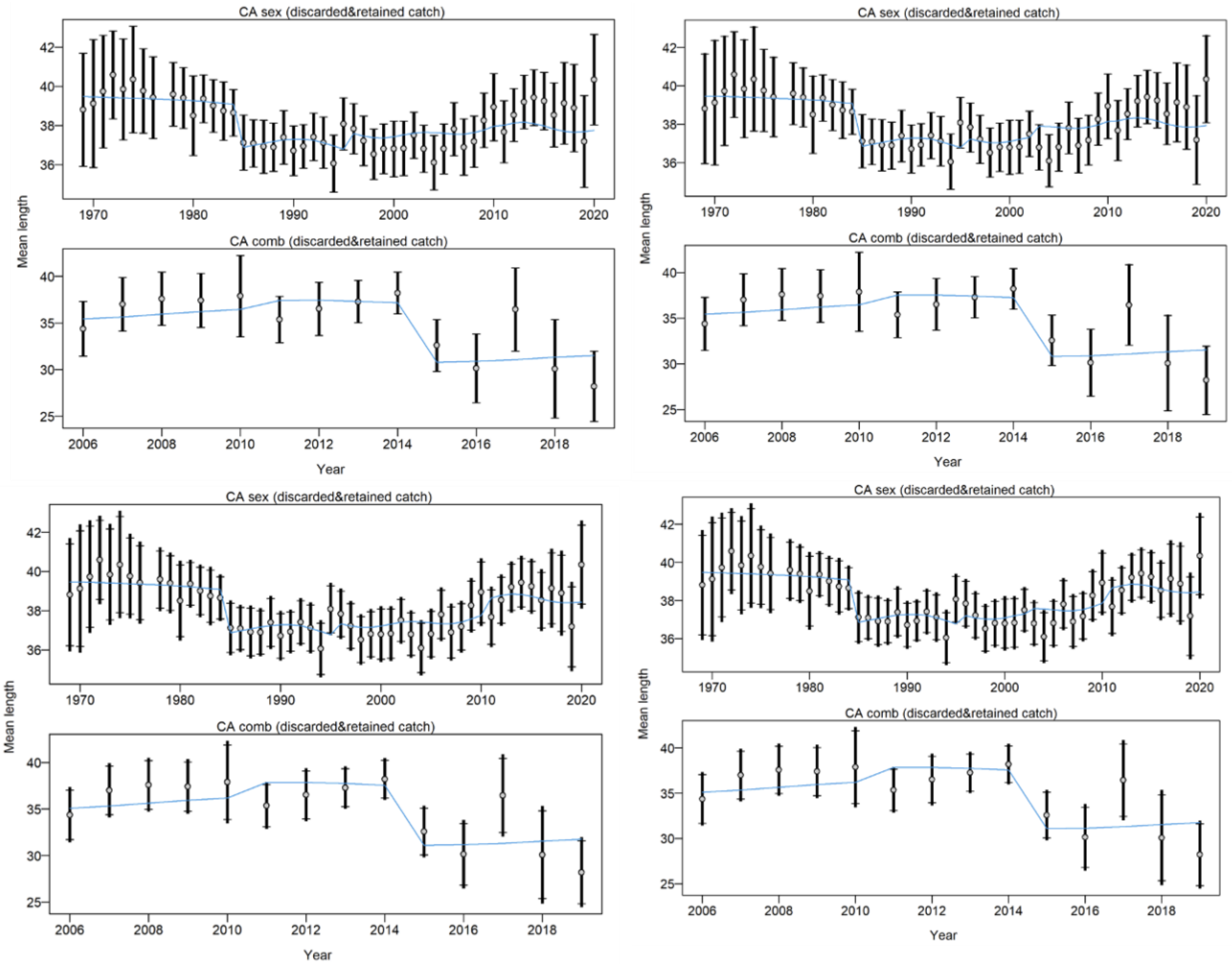

Figure A.2: Observed mean length by year for the California fleet (points) and the model expected mean length (blue line). The top right panel is the base model, the top left is the sensitivity with a block from 2003 - 2020, the bottom left is the sensitivity with a block from 2011 - 2020, and the bottom right is the sensitivity with a block from 2003 - 2010 and 2011 - 2020.

The Pearson residuals for each of the sensitivities are shown in Figure A.3 below. Similar to the mean length figures, the sensitivities that included a block from 2011 - 2020 appeared to decrease the pattern of model expectations exceeding the observations (open circles) at the end of the time series for fish less than 30 cm.

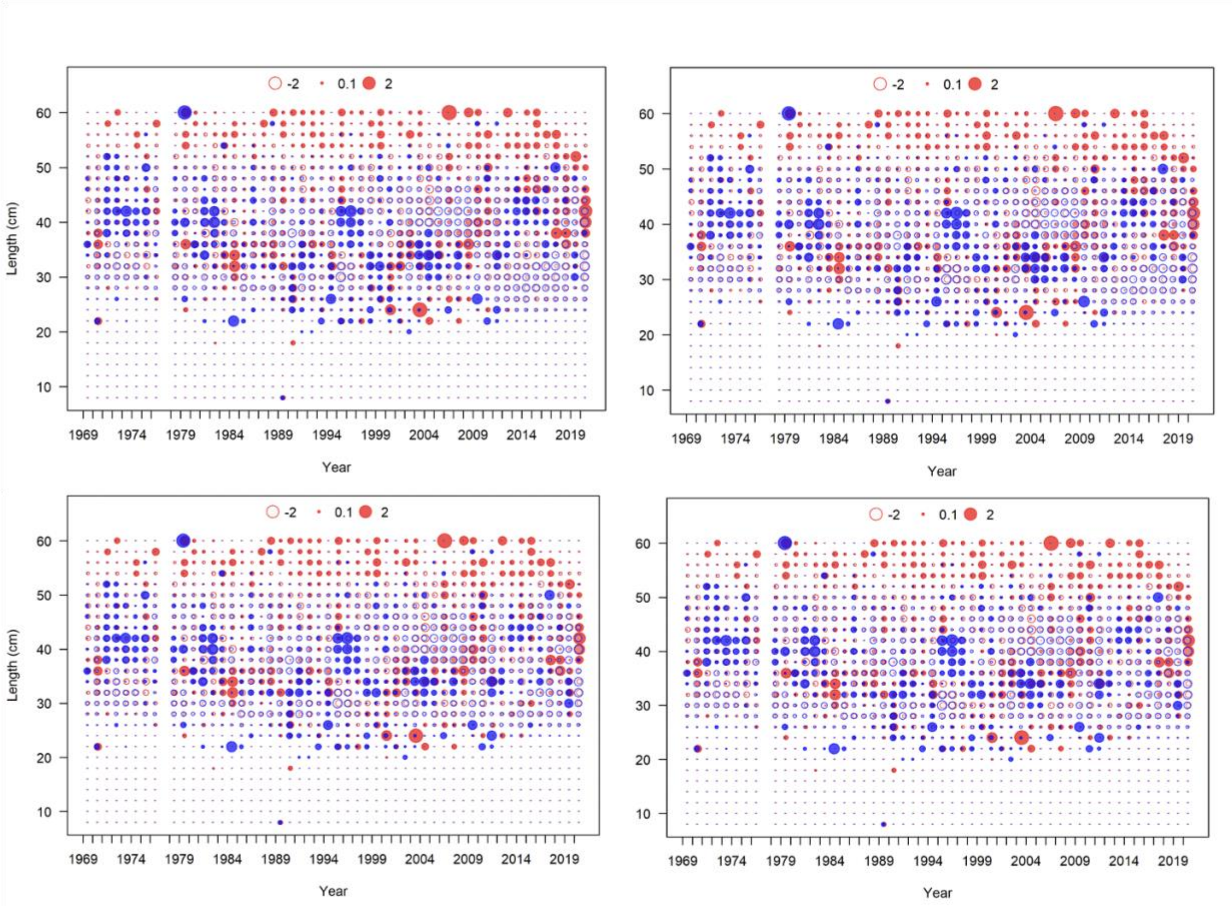

Figure A.3: Pearson residuals of length data by year for the California fleet. The top right panel is the base model, the top left the sensitivity with a block from 2003 - 2020, the bottom left is the sensitivity with a block from 2011 - 2020, and the bottom right is the sensitivity with a block from 2003 - 2010 and 2011 - 2020.

The estimates of spawning biomass and fraction unfished across the selectivity block sensitivities and the base model are shown in Figures 4 and 5. The estimated spawning biomass and fraction unfished were similar across sensitivities runs. The change in the negative-log-likelihoods (NLL) relative to the base model is shown in Table A.2. The sensitivity which added a block for 2011 - 2020 had the lowest NLL, approximately 5 units lower than the base model, but this improved fit to the data requires two additional selectivity parameters.

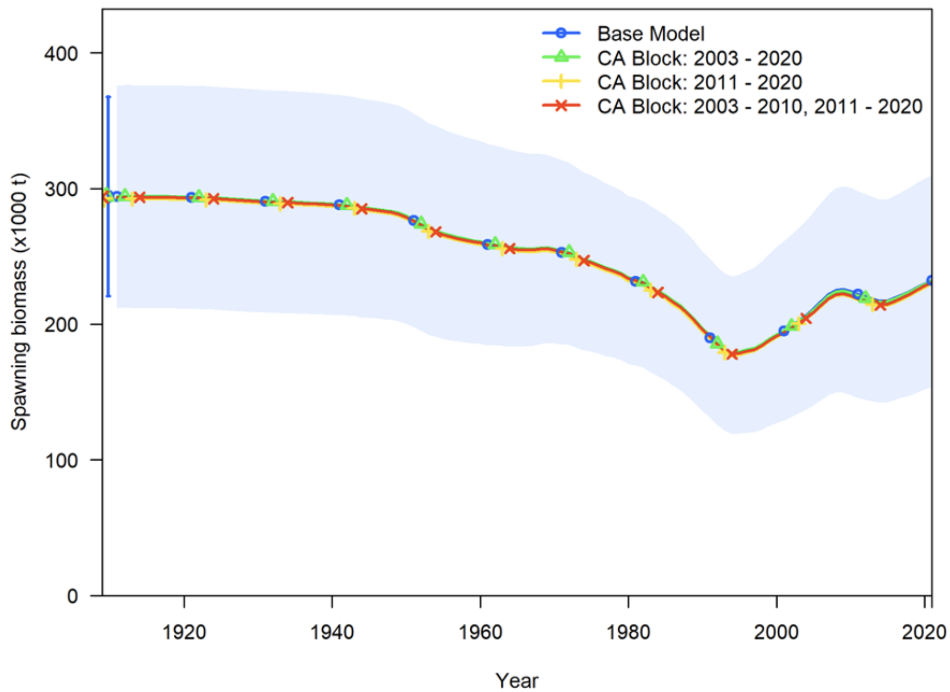

Figure A.4: Spawning biomass estimated across selectively block sensitivities.

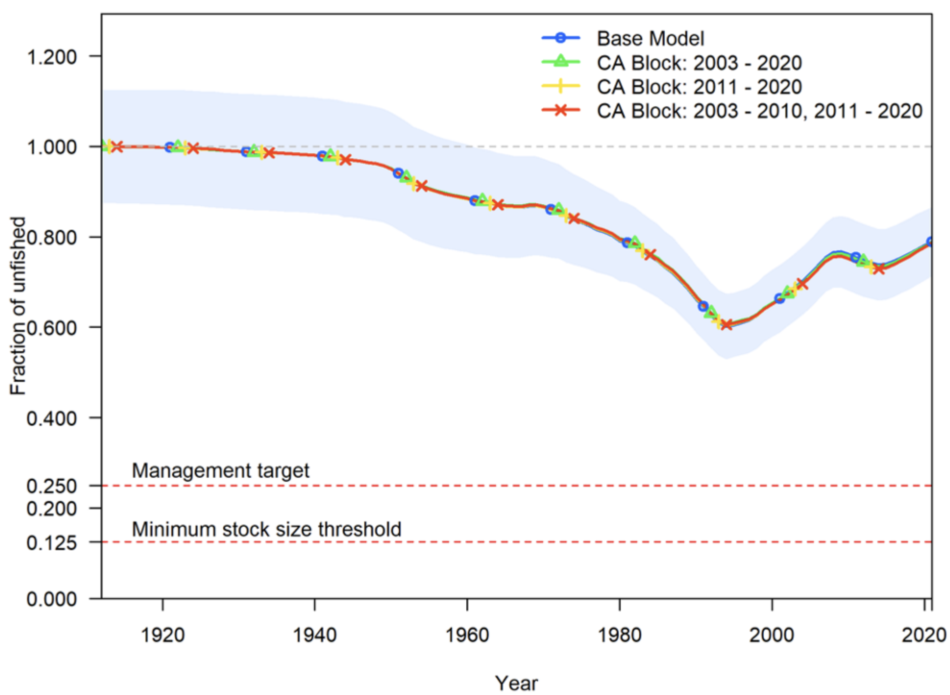

Figure A.5: Fraction unfished estimated across selectively block sensitivities.

Table A.3 Table of likelihoods and estimates across each of the block sensitivities.

	Base Model	CA Block: 2003 - 2020	CA Block: 2011 - 2020	CA Block: 2003 - 2010, 2011 - 2020
Change in NLL	**0**	**1.94**	**5.21**	**5.12**
Total Likelihood	1028.95	1027.01	1023.74	1023.83
Survey Likelihood	-49.59	-49.65	-49.83	-49.82
Length Likelihood	338.18	337.88	334.86	335.11
California	*149.93*	*148.30*	*143.45*	*143.31*
OR/WA	*165.95*	*166.14*	*165.73*	*165.76*
AFSC Slope	*28.41*	*28.51*	*28.49*	*28.48*
Triennial	*23.68*	*23.73*	*23.67*	*23.69*
NWFSC Slope	*11.31*	*11.41*	*11.35*	*11.38*
NWFSC WCGBT	*116.84*	*117.01*	*116.76*	*117.04*
Age Likelihood	909.45	908.18	908.88	908.74
Recruitment Likelihood	-15.67	-15.78	-15.87	-15.88
Forecast Recruitment Likelihood	0	0	0	0
Parameter Priors Likelihood	0.04	0.04	0.04	0.05
log(R0)	12.27	12.27	12.27	12.27
SB Virgin	294070	294299	292524	293626
SB 2020	232065	231982	229959	230935
Fraction Unfished 2021	0.79	0.79	0.79	0.79

Panel Conclusions:
While the new 2011-2020 selectivity block produced a significantly better fit (i.e., 25.21=10.42 22 units; Pr (22>10.42)=0.0055), adding this selectivity block did not result in a change in the estimated stock size and status. It was unclear to the STAT what may have driven a shift in selectivity in the California fishery during this period (the introduction of the IFQ in 2011 is captured via a shift in retention) and a similar shift in selectivity was not observed in the Oregon/Washington fleet. Recent length sample sizes (i.e., since 2017) from the CA fleet have been low, so this needs to be further explored as well. There is a potential that the discrepancy in the model fit is related to a sampling process rather than a real change in fishery selectivity. The STAT and Panel agreed to not adopt the 2011-2020 selectivity block and highlight this issue for future research (see below).

Reviewer Comments: In general, this reviewer agreed with the Panel's comments and conclusions. It should be noted, however, that a change in selectivity may be taking place. This may be an area of further investigation.

Request No. 3: Evaluate the sensitivity runs for the WCGBTS to see what may be driving the poor fit at the end of the time series.

Rationale: To understand what is causing the poor fit.

STAT Response: The STAT went through 1) all sensitivities model runs that were provided in the assessment document, 2) many of the other sensitivity runs that were performed during robustness trial examinations of the draft base model, and 3) many of the model runs that were conducted during development of the base model. In general, the STAT did not identify any model structural assumptions that, when evaluated in isolation, led to an improved fit in the mean age for the WCGBTS (Table A.3).

Table A.3: Main list of general model types that were 'visually' examined for improved fit to the mean age data for the WCGBTS relative to the base model. This list is not exhaustive of all models examined but is representative of general findings.

Better Fit	Little to No Difference	Worse Fit
Severely upweight WCGBTS age data	2011 maturity Alternative M fix/estimate Alternative weighting methods NWFSCslope_female selectivity asymptote at 1 Aging error assumptions Increase growth CVs Remove recruitment deviations Parameter offset methods Other selectivity sensitivities not mentioned elsewhere All other data source sensitivities not mentioned elsewhere	2011 fishery selectivities 2011 survey selectivities Mirror commercial selectivities Remove WCGBTS ages

Models resulting in the largest change in fit to WCGBTS mean age included severely (and artificially) up-weighting these data relative to other data sources (lambda = 10.0, or a 10-fold increase in relative weight) relative to the base model (Figure 6) and removing these data altogether (Figure 7). Clearly, there is a tradeoff between fitting WCGBTS length data versus age data in this model, and this is the case in general as well as by specific parameters (e.g., see profile plots for key parameters, Figures 153-164 in the draft assessment document). In general, input sample sizes were specified as 3.09 * number of tows for WCGBTS length data and was specified as the number of fish for WCGBTS conditional age-at-length data (CAAL; further details at the top of page 12 in draft assessment document). The range of input sample sizes for length across the WCGBTS was 402 (2004) to 1829 (2018), and this range was 1 fish to 78 fish per year-sex-length bin for CAAL. The Francis data weighting approach used in the base model resulted in a 4-fold higher relative weighting of input sample sizes for lengths as compared to ages (i.e., Francis weight of 0.41 compared to 0.11 for WCGBTS length and ages, respectively).

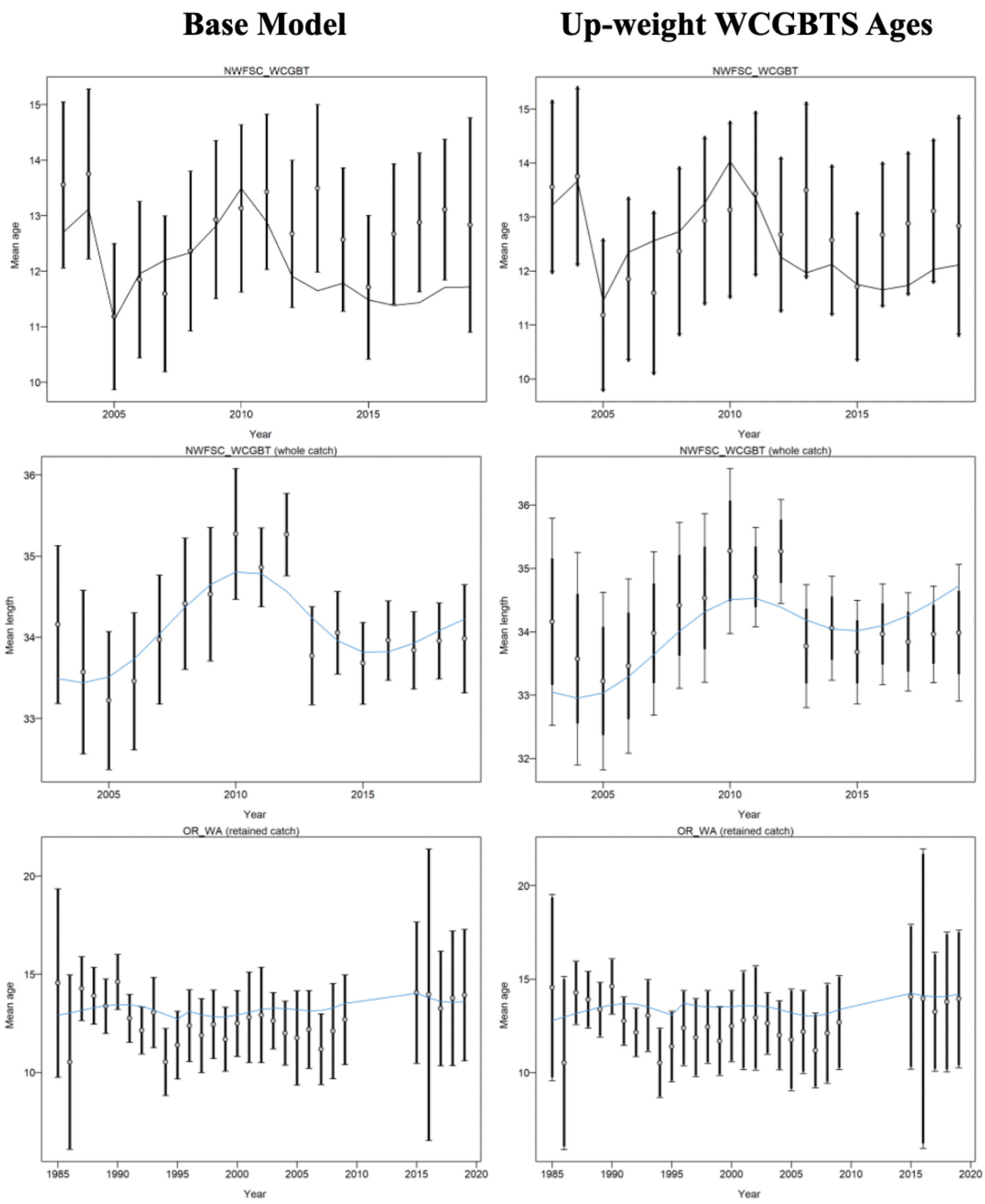

Figure A.6: Fits to WCGBTS mean age for a model that severely up-weighted these data (top right) relative to other data sources (lambda = 10.0, or a 10-fold increase in relative weight) as compared to the base model weight for this data source (top left). Fits to WCGBTS mean length and OR_WA mean age is also shown for comparison.

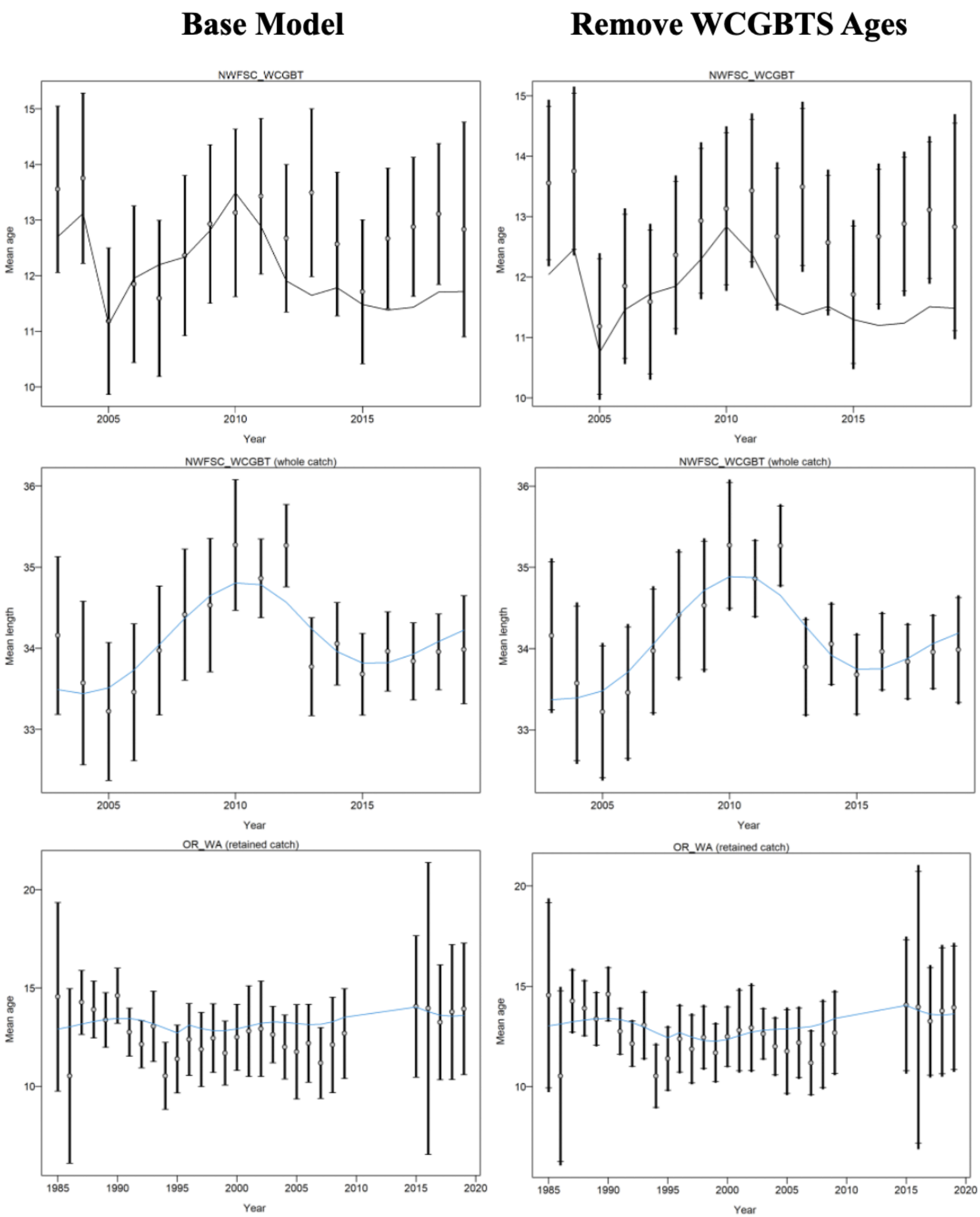

Figure A.7: Fits to WCGBTS mean age for a model that removed these data (top right) as compared to the base model (top left). Fits to WCGBTS mean length and OR_WA mean age is also shown for comparison.

In general, the base model fits the WCGBTS mean length data well at the expense of not fitting the WCGBTS age data as well. When the model is forced to fit the WCGBTS age data more so than in the base model, the fit of length data becomes worse (Figure A.6) as does recent fits to the WCGBTS index (Figure A.8). This change also results in an *a priori* unexpected stock trajectory and the undesirable property of autocorrelation in early recruitment deviations (Figure A.8). The base model attempts to balance WCGBTS length and age data. Ideally, an assessment model would provide unbiased and risk-neutral estimates (i.e., the equal likelihood of being above or below the true state). The Pacific Fishery Management Council applies a precautionary approach when adopting Annual Catch Limits to avoid exceeding the true and unknown Overfishing Limit (OFL) of the stock. Forcing the model to fit the WCGBTS age data results in a dramatic shift in the estimated stock size and status would result in large changes of the estimated OFL and Acceptable Biological Catch (ABC) relative to the base model. Hence, the under-fitting of the age data in the base model results in a de facto precautionary approach compared to forcing the model to fix the mean age of the WCGBTS (i.e., the estimated OFLs and ABCs from the base model would be well below those estimated from this alternative model).

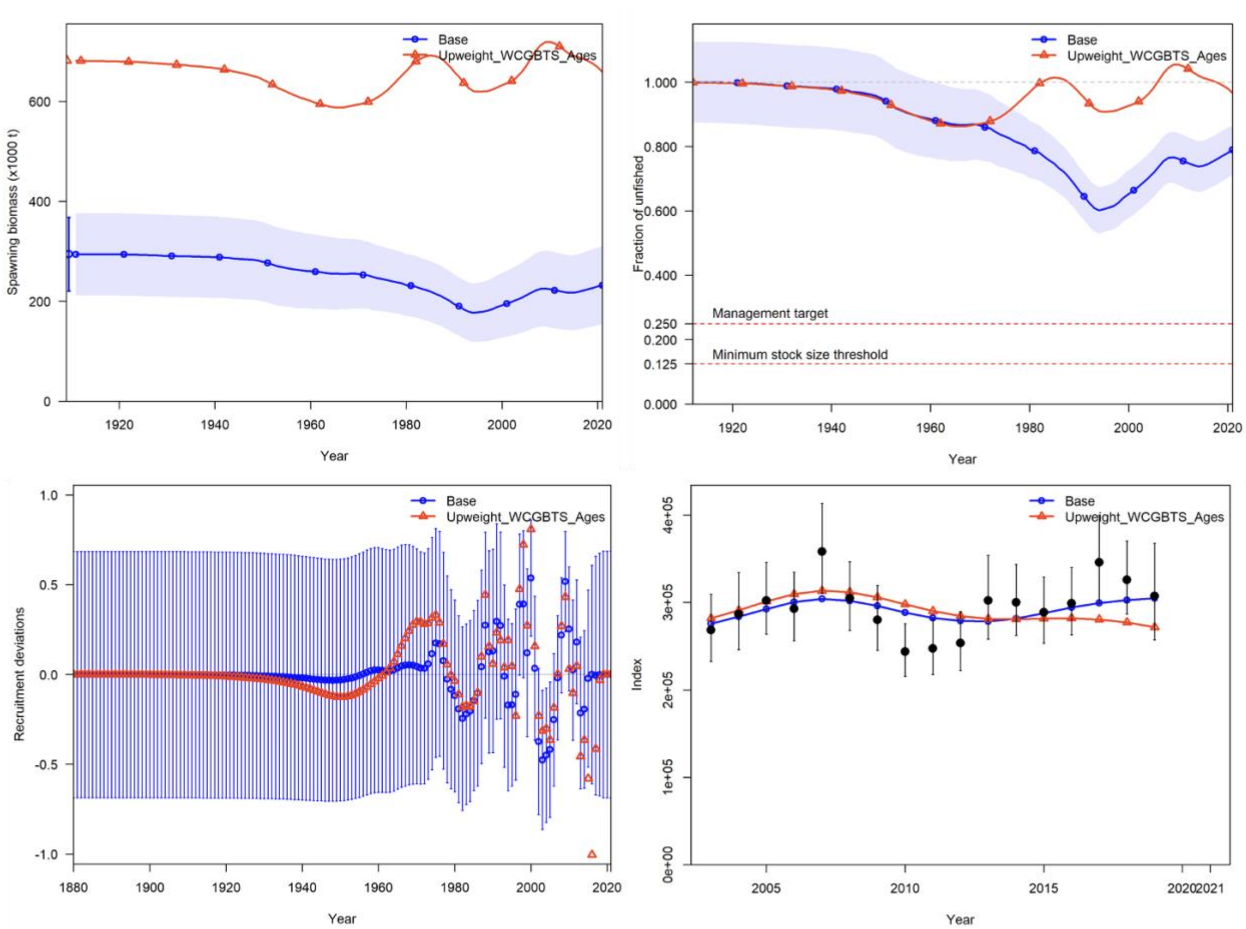

Figure A.8: Comparison of spawning biomass (top left), stock status (top right), recruitment (lower left), and the fit to the WCGBTS index (lower right) for the base model and the sensitivity model where the WCGBTS mean age data are forced to fit better than in the base model.

Panel Conclusions:

Both the STAT and the Panel agreed that these results did not provide evidence to support a change in model formulation. A possible mechanism for the lack of fit is a change in growth rates and size-at-age.

Annual estimates of mean length-at-age were examined and substantial differences were not evident. However, a small change in growth rates may be enough to account for the lack of fit to the WCGBTS age and length data. This issue requires future research (see below).

Reviewer Comments: This reviewer agrees with this conclusion that model formulation should remain unchanged. It is also agreed that a change in growth rate is likely the cause, though more research is needed. This is, however, a troubling uncertainty that suggests that further research should be conducted.

Request No. 4: Provide a likelihood profile of M including the priors in the base model.

Rationale: To explore a range of M estimates for states of nature in the decision table.

STAT Response: Since requests 4 and 5 are closely linked, we will respond to each request in a single response.

Request No. 5: provide an alternative run with M estimated with a tight prior (SE = 0.219).

Rationale: To explore a range of M estimates for states of nature in the decision table.

STAT Response: Since requests 4 and 5 are closely linked, we will respond to each request in a single response. The West Coast groundfish Terms of Reference (TOR) requests that decision tables identify the low and high states using one of the following options:

"One method bases uncertainty in management quantities for the decision table on the asymptotic standard deviation for the current year spawning biomass from the base model. Specifically, the current year spawning biomass for the high and low states of nature is given by the base model mean plus or minus 1.15 standard deviations (i.e., the 12.5th and 87.5th percentiles). A search across fixed values of R0 is then used to attain the current year spawning biomass values for the high and low states of nature. Another method to provide reasonable alternative models uses the 12.5% and 87.5% quantiles of the likelihood profile of an estimated parameter (the value of 0.66 reflects the chi-square distribution with one degree of freedom) to determine the major axis of uncertainty. Expert judgment may also be used as long as it is fully explained, justified, and documented."

Before the meeting, the STAT explored the viability of defining low and high states for a potential decision table based on either the base model uncertainty or the profile across values of female natural mortality rate (M). Request 4 and 5 attempt to provide information to select M values to create the low and high state of nature in the decision table. Below is a table of changes in the total negative-log-likelihood (NLL) across values of female M around the median of the prior (0.108 per year):

Table A.5: Changes in the negative log-likelihood across female M values using either no prior likelihood contribution, prior likelihood contributions from the default prior, or a tighter prior (SE = 0.219) on female M.

Base Model: Profile across natural mortality with the prior likelihood excluded:							
	M = 0.07	M = 0.08	M = 0.09	M = 0.10	M = 0.108	M = 0.11	M = 0.12
NLL	1024.14	1021.75	1022.54	1025.44	1028.96	1029.98	1035.9
Δ Base	-4.82	-7.21	-6.42	-3.52	0	1.02	6.94
Request 4: Change in NLL with the prior likelihood included (Default Prior SE)							
	M = 0.102	M = 0.104	M = 0.106	M = 0.108	M = 0.110	M = 0.112	M = 0.114
NLL	1026.23	1027.08	1027.98	1028.96	1029.99	1031.07	1032.21
Prior	0.033	0.033	0.034	0.036	0.040	0.045	0.050
Δ Base	-2.73	-1.88	-0.98	0	1.03	2.11	3.25
Request 5: Change in NLL with the prior likelihood included with a tighter SE							
	M = 0.102	M = 0.104	M = 0.106	M = 0.108	M = 0.110	M = 0.112	M = 0.114
NLL	1026.26	1027.09	1027.99	1028.96	1029.99	1031.08	1032.24
Prior	0.060	0.044	0.037	0.036	0.043	0.056	0.075
Δ Base	-2.7	-1.87	-0.97	0	1.03	2.12	3.28

In initial explorations of the original profile (Base without prior contribution), the change in NLL across values of M resulted in a relatively steep profile where small changes in M resulted in changes in the NLL that would quickly exceed the 12.5% and 87.5% intervals around the base model (0.66 reflects the chi-square distribution with one degree of freedom). Low and high states of nature, based on M profiles did not seem to capture the range of structural uncertainties in the model, as evidenced through sensitivity analyses.

An alternative method presented in the TOR for defining low and high states of nature would be to identify the 12.5% and 87.5% quantiles around the final year spawning biomass. Using the asymptotic standard deviation for the current year spawning biomass from the base model, the range of low and high spawning biomass in 2021 would range from 186,336 - 277,794 around the base model value of 232,065 which corresponds to female M of 0.093 and 0.1144 per year. The guidance in the TOR clearly states that the low and high states should be identified using the asymptotic standard deviation from the base model. However, an alternative approach that could allow one to capture a larger uncertainty interval for models with low estimated model uncertainty would be to use the default category 1 sigma value of 0.50 to identify the 12.5% and 87.5% quantiles. Using this higher level of uncertainty, the range of low and high spawning biomass values in 2021 would range from 130,584 - 412,410 corresponding to female M values of 0.084 and 0.126 per year. The spawning biomass and fraction unfished trajectories from both approaches are shown below relative to the base model.

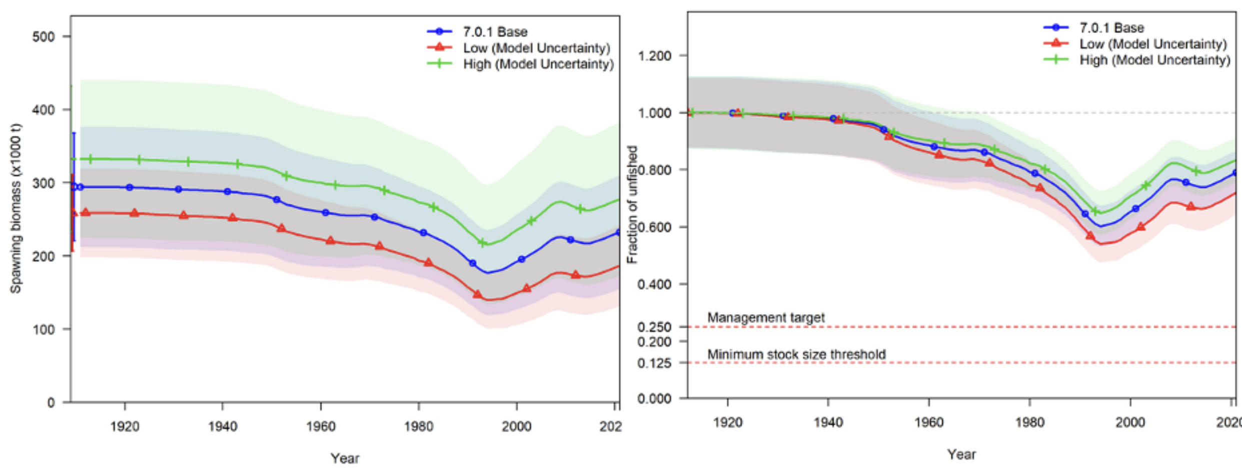

Figure A.9: The estimated spawning biomass and fraction unfished from the base model and low and high states of nature determined based on the 12.5% and 87.5% quantile from the uncertainty around spawning biomass in 2021 from the base model.

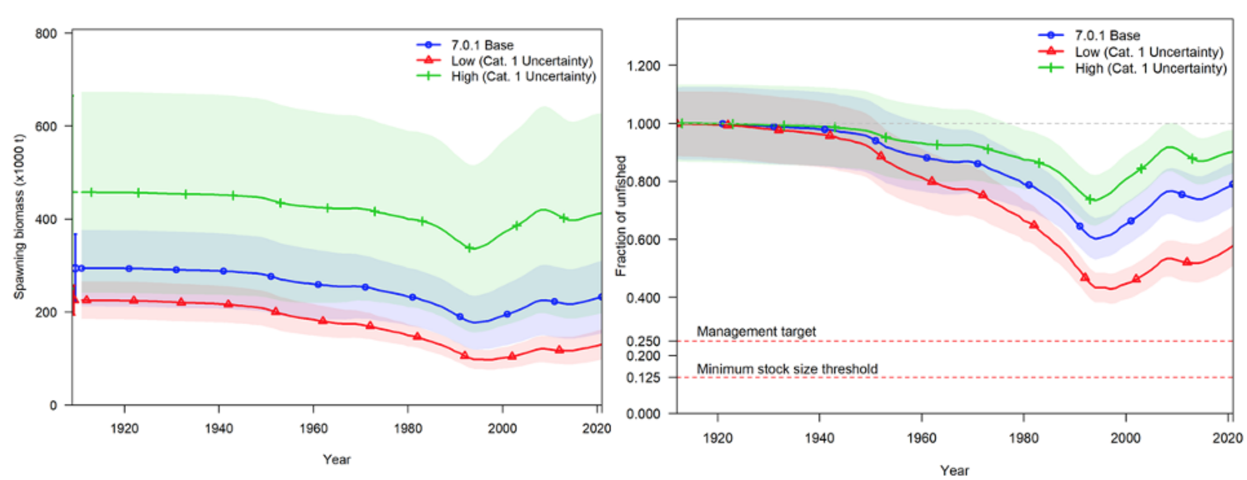

Figure A.10: The estimated spawning biomass and fraction unfished from the base model and low and high states of nature determined based on the 12.5% and 87.5% quantile from the category 1 default sigma value of 0.5.

Panel Conclusions:
The STAT and Panel agreed that the low and high states of nature in Figure 9 reflected the range of structural uncertainties in the model, as evidenced through sensitivity analyses. The WCGBTS catchability parameter estimate for the Low run was about 1.98, and the catchability (Q) estimate for the High run was about 0.56. This range of Q's seems plausible. The STAT and Panel agreed that the female M values used to produce the Low and High runs in Figure 9 will be used to define low and high states of nature for the decision table.

Reviewer Comments: Overall, this reviewer agreed mostly with the conclusions of the Panel. However, it is noted that the resulting catchabilities of 0.56 on the low end and 1.98 on the high end may be too large of a range. Throughout the discussion it was clear that members of Panel and SSC were interested in capturing this range of uncertainty in Natural Mortality as a proxy for other uncertainties in the model; especially in the context of providing advice. While this is likely a very precautionary approach, this reviewer is not at all comfortable with "baking in" uncertainty in the same aspects of a model as a proxy for other uncertainties. Additionally, catchabilities close to 2 are, in this reviewer's opinion, somewhat beyond plausible without direct evidence of herding or another mechanism.

Strengths

Overall, model fit and diagnostics were good with this model, helping to increase the confidence of the results. There was little if any retrospective pattern, which also increases confidence.

A chief strength was the depth and breadth of the sensitivity analysis and the "Bridging Analysis" that was done for this stock. The assessment highlighted the major structural and data uncertainty with this approach and framed the overall uncertainty well using the 12.5% and 87.5% quantiles around the final year spawning biomass. In short, the STAT used a logical progression on sensitivities and diagnostics to highlight the model's uncertainties in a believable and direct way.

Weaknesses

Chief among the weaknesses of the data, and assessment approach resulting from those data, are the level of cryptic biomass associated with this assessment. Clearly older, larger females are not available to either the fishery or the surveys as documented in the assessment report. This cannot be understated, even small changes in the assumed M resulted in near implausible catchability coefficients and dramatic changes in stock biomass. While there are concrete and well-documented reasons for this level of cryptic biomass, the sheer level is disturbing. Particularly when it comes to setting advice for this stock.

The lack of aging data from the California fishery and elsewhere after 2009/2010 is also of concern. Lack of aging data can play into problems of model fit if the few samples that are collected are not representative. Additionally, lack of aged otoliths can otherwise mask changes in the fishery's selectivity or in fish growth that might otherwise be discernable had those ages be available.

There is an obvious tension in the model, with it preferring a low M (0.082) compared to the fixed M from life-history approaches (0.108). As the model preferred value is not plausible for this stock given the maximum age, it does create some uncertainty about model misspecification.

The lack of fit to the most recent survey data is another uncertainty. While this could be the result of sampling, changes in growth, or changes in spatial distribution, this lack of fit is another weakness that should be investigated further.

Stock structure and spatial productivity dynamics are not well understood. While tagging data suggest limited adult movement, genetic mixing appears to be the result of larval dispersal. However, this can also lead to spatial changes in stock productivity that can make assessments difficult, particularly in the face of lower or biased sampling regimes. Part of the lack of fit in the most recent survey data could be the result of differences in growth among locations; and inflated by smaller sample sizes for both fishery-dependent and independent data.

Research recommendations

Higher priority

- Consider studies to verify the magnitude of the cryptic biomass.
- Improved understanding of survey catchability could be provided via trawl escapement and herding studies. This is linked to a 2011 recommendation.
- Improved size and age fishery sampling south of Pt. Reyes should be provided, to investigate possible differences in age, size, and sex structure by depth and latitude. More generally, increase collection and reading of age compositions for the fishery to improve the application of an age-structured assessment model.
- Investigate the spatial and temporal dynamics, seasonality, and ontogenetic movement that could help to capture what is happening with Dover Sole regarding the distribution of ages in the bottom trawl survey. Investigate if there is seasonality or annual environmental factors that could potentially change distribution patterns and how those patterns changes overlap with the bottom trawl survey.

Lower priority

- Consider using the AFSC Slope Survey age data as conditional age-at-lengths.
- Conduct spatiotemporal analysis of maturity-at-length and length-at-age, and examine if trends are significantly different. This is linked to a 2011 recommendation.
- Conduct additional genetic and tagging studies to examine stock structure and connectivity of the stock across its whole range.
- Consider if existing tagging information provides useful assessment information about growth and/or mortality rates.

Reviewer Comments:

Overall, this reviewer agreed with the research recommendations as put forth. It is noted that while studies to verify the magnitude of the cryptic biomass are very important, the likelihood of designing a project to do so seems remote. Nonetheless, if accomplished, such a study could go a long way to reduce the level of uncertainty in the assessment.

Of the recommendations given, of most importance is the catchability/herding work. Studies investigating this issue should help to determine if, in fact, herding occurs and its impact. Of next importance is the collection and processing of aged samples; particularly from the California fishery. There are suggestions in the model on changing growth rates and this sampling could be vitally important to help determine if this change in growth is occurring, its impact on the modeling efforts, and potential links to climate variability.

Similarly, investigations on seasonal/ontogenetic factors which may affect the migration/movement patterns are also important. Such changes can affect the survey catchability as well as fishery selectivity, and while less important than other research, are desperately needed for further model improvement.

Conclusions

Dover sole is a challenging stock to assess. Despite this, the STAT did an excellent job setting up and documenting the model runs, as well as providing ample sensitivities and discussion of the potential uncertainties. Overall, this reviewer agreed with the consensus of the Panel that the base run represented the best available data and analysis to give management advice of stock status, reference points, and harvest policy.

That said, there are several uncertainties and areas of difficulty in the assessment of Dover sole. Chiefly among these is the level of cryptic biomass. This is not surprising given that as the survey strata gets deeper there is little trail-off of the abundance of female Dover sole. This has vexing consequences for modeling. As stated previously, even small changes in the assumed natural mortality can lead to rather large changes in stock status as well as resulting management advice for this stock.

Another area of concern is the potential misspecification of the model. Clearly, the model is looking for an implausibly low natural mortality value according to the likelihood profiles. This may indicate tension between some data streams in the model and others. While this reviewer agrees to fixing natural mortality at the level the STAT suggested, this may indicate a larger underlying problem.

The “low hanging fruit” as it were, to improve the model would be to first increase the number of aged fishery-dependent sampling with a complete reanalysis of growth, and second to conduct survey work to determine the best range of potential catchability for the groundfish survey. Both studies could help inform and possibly detect spatial changes in growth. Additional tagging work could also help to nail down the migration and movement rates across the latitudinal range of the stock.

A final area of concern deals with the characterization of uncertainty. While this reviewer agrees with the approach as put forth by the STAT, as well as the overall Groundfish Terms of Reference, caution is advised. It can be easy to expand certain aspects of uncertainty, often erroneously, to compensate for other “unaccounted for” uncertainties. While such a practice is likely conservative, it seems less transparent. A better approach, at least in this reviewer's opinion, would be to use realistic measures of uncertainty along the axis of uncertainty and use a precautionary buffer for any "unaccounted for" structural or data uncertainties.

While there is certainly more work to do and some modeling issues, this model has dramatically been improved since the 2011 assessment. The STAT should be commended for their work on this model, for their diligence in honestly presenting the sensitivities and unresolved problems, as well as their advancement of this stock's assessment.

Section B: SPINY DOGFISH

Introduction and model structure

Spiny dogfish is a widely distributed shark species found from Baja California to the Gulf of Alaska. Here, the stock assessment focused on that part of the stock bounded by the US Canadian border to the North and the US Mexico border to the south. Dogfish can range far offshore to depths approximating 1200 m.

Dogfish are slow-growing, late to mature, have low natural mortality, less fecund, and have a long lifespan, resulting in a high degree of vulnerability to fishing pressure. While there was a large targeted fishery (the Vitamin A fishery) early in the 20th century, since then dogfish have been harvested principally as incidental bycatch in other fisheries.

Approach:

In this assessment, the dogfish population was modeled using Stock Synthesis, version 3.30.16 as a two-sex approach: due to the dimorphic growth by sex. Sex-specific growth was estimated within the model. It should be noted that females larger than 80 cm were removed from this growth analysis as there was considerable uncertainty in aging due to worn spines (see below). The start year for the model was 1916 and assumed an unfished equilibrium state at that time.

The most recent assessment was very similar to the 2011 assessment, except for:

- Updated fisheries- and survey-related data
- Abundance indices estimated using the VAST approach
- Revised historical discard estimates using the sablefish fishery
- Updated selectivity assumptions from asymptotic to dome-shaped with sex-specific offset
- Updated biological parameters, and updated tuning for age data
- Fecundity parameters were half of the values used in the 2011 assessment to account for the 2-year (22-24 months) gestation period

Data elements:

Eight fleets were included in the model: 1) bottom trawl landings, 2) bottom trawl discard, 2) midwater trawl catches, 4) bycatch in at-sea Pacific hake fishery, 5) non-trawl landings, 6) non-trawl discard, 7) non-trawl catches within historical Vitamin A fishery, and 8) recreational removals. In addition, there were five fishery-independent survey indices including; the AFSC (Alaska Fishery Science Center) Triennial, the AFSC slope, the NWFSC (North West Fishery Science Center) slope, and the WCGBTS (West

Coast Groundfish Bottom Trawl Survey), and one set-line survey conducted by the IPHC (International Pacific Halibut Commission).

Only aging data from the WCGBT was used. For the rest of the fisheries and indices, length composition was the only data available for growth.

Parameterization:

Table B.1: Parameters used in the base model for Spiny Dogfish

Model strcuture	Base model
Starting year	1916
Population characteristics	
Maximum age	95
Gender	2
Population lengths	10-136 by 2 cm bins
Summary biomass (mt)	Age 1+
Data characteristics	
Data lengths	12-132 by 4 cm bins
Data ages	0-71 ages
Minimum age for growth calculations	0
Maximum age for growth calculations	999
First mature age	1
Fishery characteristics	
Fishing mortality method	Hybrid F
Maximum F	4
Catchability	
AFSC triennial survey	Estimated parameter
AFSC slope survey	Analytical solution
NWFSC slope survey	Analytical solution
WCGBTS	Estimated parameter, Fixed at estimated value
IPHC survey	Analytical solution
Selectivity	
Bottom trawl landings	Double normal, Female offset
Bottom trawl discard	Double normal, Female offset
Midwater trawl	Double normal, Female offset
At-sea hake fishery bycatch	Double normal, Female offset
Nontrawl landings	Double normal, Male offset, Assumed asymptotic
Nontrawl discard	Double normal, Female offset
Nontrawl catch during vitamin A fishery	Mirrored to Nontrawl discard
Recreational catch	Double normal
AFSC triennial survey	Double normal, Female offset
AFSC slope survey	Double normal, Male offset
NWFSC slope survey	Mirrored to AFSC slope survey
WCGBTS	Double normal, Female offset
IPHC survey	Double normal, Male offset, Assumed asymptotic

Sensitivities:

The STAT ran a large number of sensitivities including:

- 2011 assessment discard estimates
- 50% increase in bottom trawl and non-trawl discards
- 50% decrease in bottom trawl and non-trawl discard
- 100% and 50% discard mortality for bottom trawl and non-trawl discard fleets
- 35% discard mortality in the non-trawl fleet
- 6% mortality for non-trawl discard fleet and 5% for bottom trawl discard fleet
- 2011 weight-length relationship, fecundity, natural mortality (0.064), and growth parameters
- Natural mortality for males while keeping female natural mortality fixed at 0.065
- Natural mortality is estimated for females while males are assumed to have the same M as females
- Growth parameters all age data from all sources
- As well as a retrospective run which drops the last year of data sequentially for 10 years

Requests

A total of 12 requests were made by the Panel to the STAT. Like with Dover sole, below are those requests, the rationale, the responses, Panel comments, and this reviewer's comments.

<u>Request No. 1</u>: Provide a time series plot of the residuals of the total catch relationship between Sablefish and Spiny Dogfish from the observer data.

<u>Rationale</u>: The relationship is assumed to not vary by year and this needs to be checked

<u>STAT Response</u>:
The STAT provided Figures B.1 and B.2 as requested. Figure B.2 demonstrates that there is no pattern or bias in the residuals and only one year had an outlier.

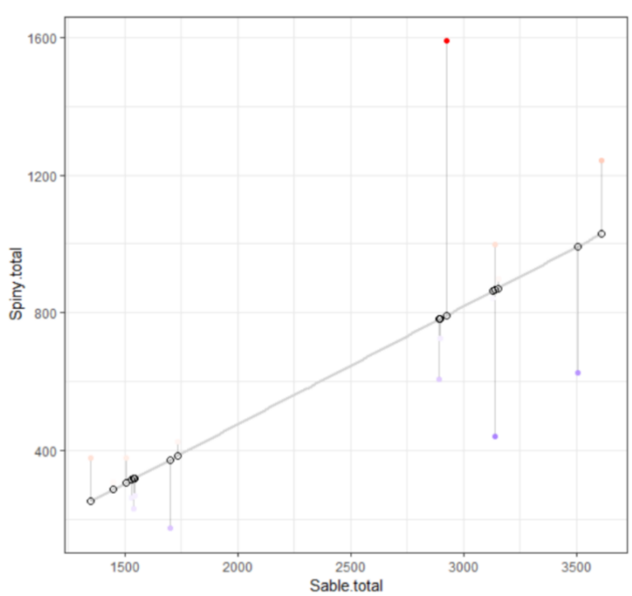

Figure B.1: Residuals against the linear relationship between catches by bottom trawl of Spiny Dogfish and Sablefish.

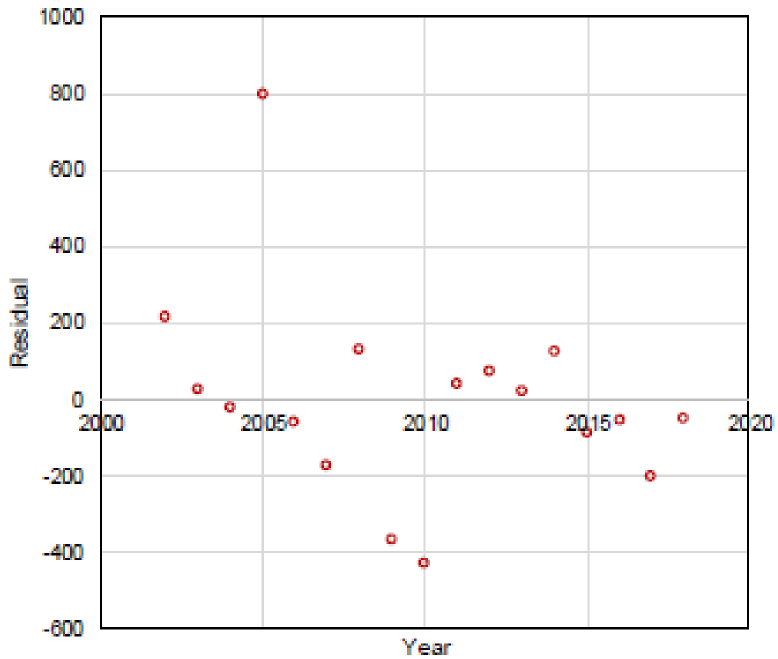

Figure B.2: Residuals of the linear relationship between catches by bottom trawl of Spiny Dogfish and Sablefish by year.

Additionally, the STAT examined using just Sablefish landings, as opposed to catch (both landings and estimated discards of Sablefish), to estimate dogfish discards. As shown in Figure B.3, the predictive value of Sablefish landings was rather poor.

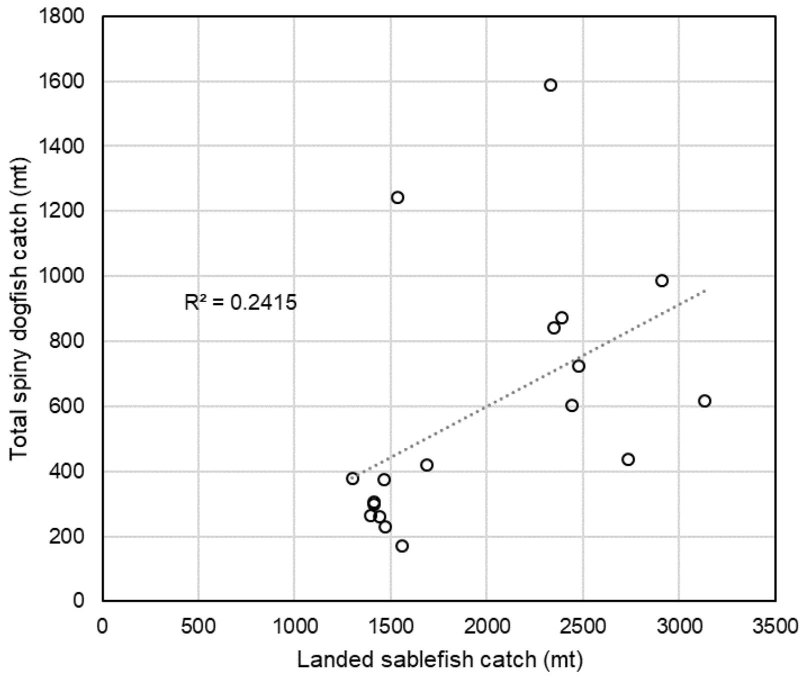

Figure B.3: Relationship between catches of Spiny Dogfish and landings of Sablefish by bottom trawl.

Panel Conclusions**:**
The Panel agreed with the STAT that there appeared to be little bias and no time pattern in the residuals that could indicate a change in the relationship between Sablefish catch and Spiny Dogfish discards. While there were some concerns about using Sablefish catch as a predictor of dogfish discards during 1960 -2002 given the seasonal differences in occurrence between these two species, overall the Panel concluded that the proposed procedure was an improvement compared to the 2011 assessment which assumed a static catch to discard ratio.

Reviewers Comments:
This reviewer agreed with the Panels conclusions that this represents the best information. While improvements in future assessments are recommended, this is a vast improvement over the 2011 assessment approach.

Request No. 2: Provide a sensitivity of the gamma vs. log-normal error distribution of the VAST.

Rationale: There is a need to explore alternative assumptions.

STAT Response:
The STAT provided Figures B.4 and B.5 that compared gamma and log-normal error distributions for runs with fixed and estimated catchability (*q*)

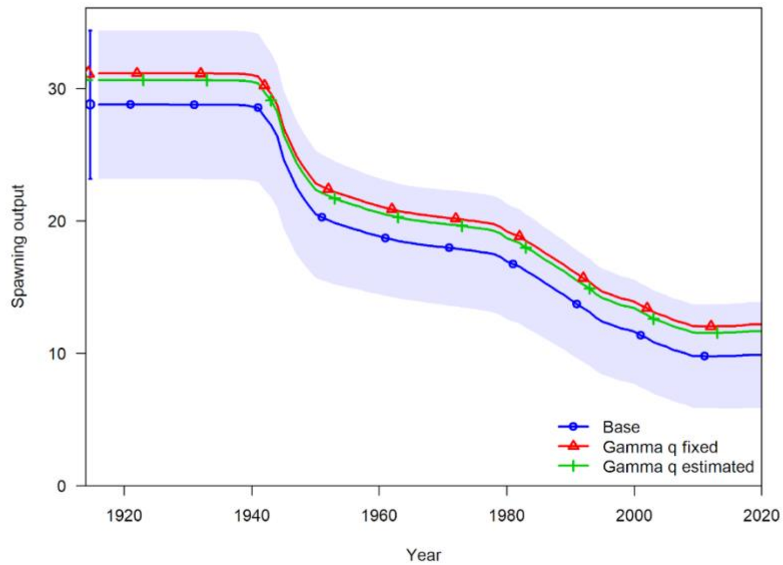

Figure B.4: Time series of spawning output (in millions of fish) associated with abundance indices estimated by using lognormal (base) vs gamma error assumption in the VAST model.

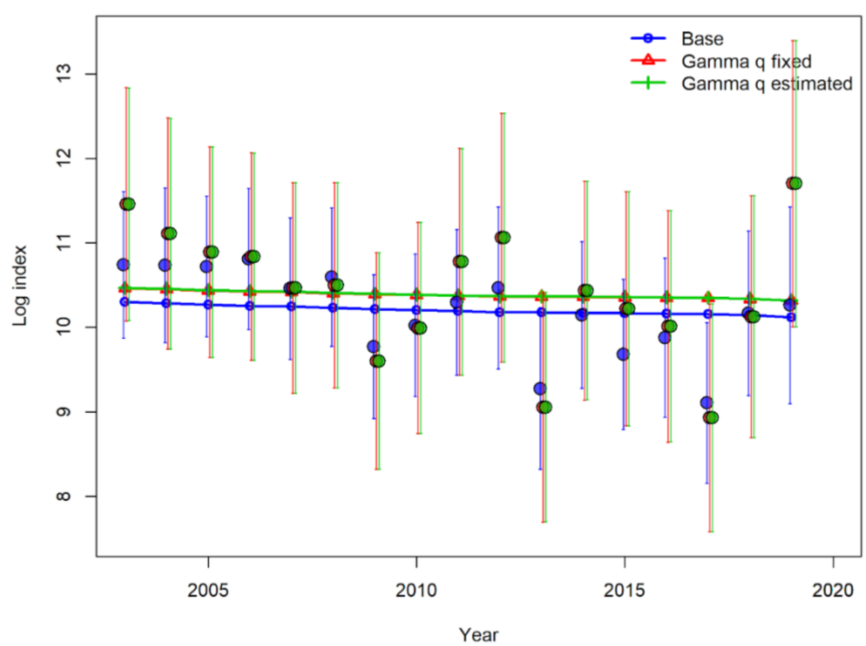

Figure B.5: Comparison of fit to the WCGBT index by using log-normal (base) vs gamma error assumption in the VAST model.

As can be seen in the Figures, switching to a gamma error structure did not change the results of the model by much, nor were the estimated q's for the WCGBT survey very different (0.59 for the bases vs 0.6 for the VAST using Gamma). Additionally, the assessment model fit the WCGBT survey index with the VAST lognormal error structure indices was slightly better than with the gamma error structure.

Panel Conclusions:
The Panel agreed with the STAT's conclusion to use the lognormal error distribution in the VAST for developing the WCGBTS index.

Reviewer comments:
This reviewer agreed with the Panels suggestion, as well as the STAT's that lognormal error was the best approach to use.

Request No. 3: Provide a justification for the 80 cm cutoff in the growth function.

Rationale: To better understand the model selection decisions.

STAT Response:
The STAT reminded the Panel that the length at 50% maturity was 88.2 cm, and also noted that the proportion mature was approximately 12.5% at the base model's 80 cm cut-off. Reducing the cut-off to 70 cm had no effect on the model output as shown in Figure 6. Additionally, the L_{inf} estimated from the alternative run with a 70 cm cut-off was identical to the base run using an 80 cm cut-off.

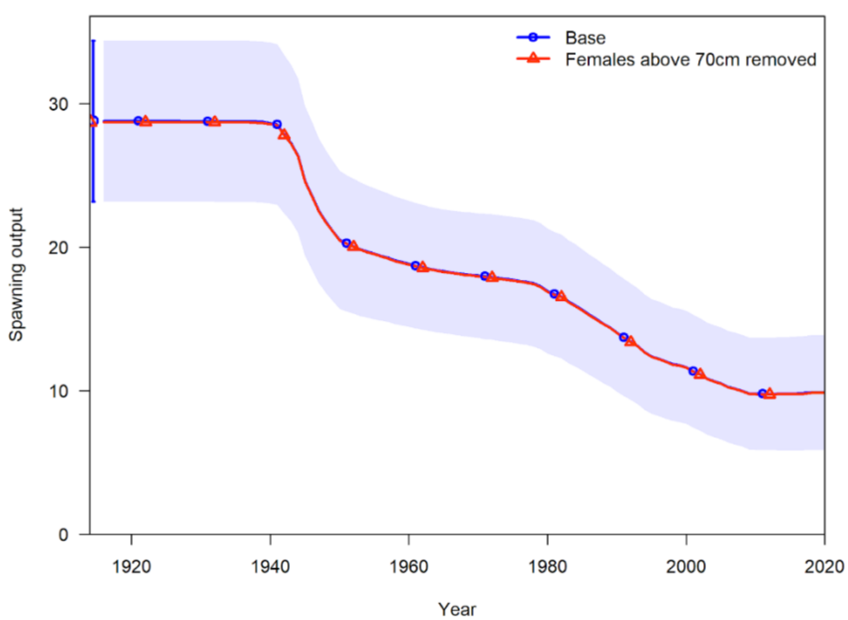

Figure B.6: Time series of spawning output (in millions of fish) associated with growth parameters estimated with females above 70 cm or 80 cm.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request and agreed that an 80 cm cutoff was justified.

Reviewer Comments:
This reviewer agreed with the conclusion that the 80 cm cut-off was the most appropriate given the life history of dogfish.

Request No. 4: Provide the uncertainty intervals of the Spiny Dogfish historical discard estimation.

Rationale: To better understand the realistic bounds of historical dogfish removals. A sensitivity of these bounds may be requested later if the model is sensitive to these assumptions.

STAT Response:

The STAT provided estimates of uncertainty in the discarded fraction from 1960 to 2002 as shown in Figure B.7. The methods for how this uncertainty was calculated appears in the STAT's response to Request 6.

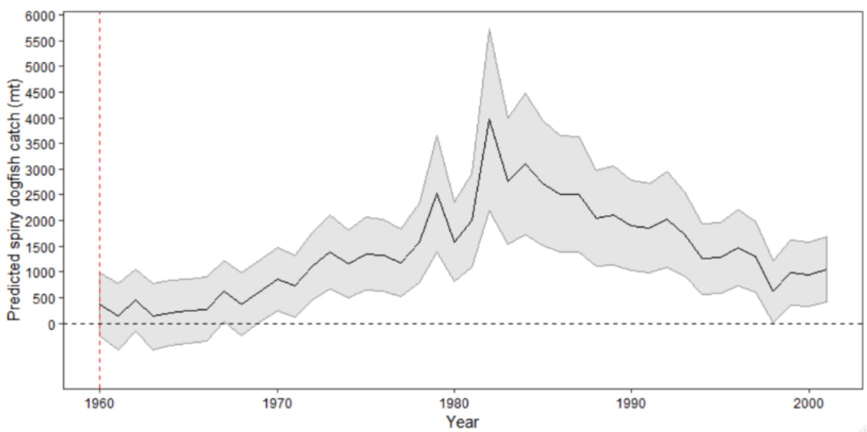

Figure B.7: Historical discard estimates with uncertainty.

Runs were conducted at both the upper and lower bounds in addition to the base model. Results suggested some difference in the initial biomass estimated and a corresponding difference in the estimated depletion (Figures B.8 and B.9). Despite this rather large uncertainty in the discards during this time period, the model didn't appear to be very sensitive to it in its stock status determination.

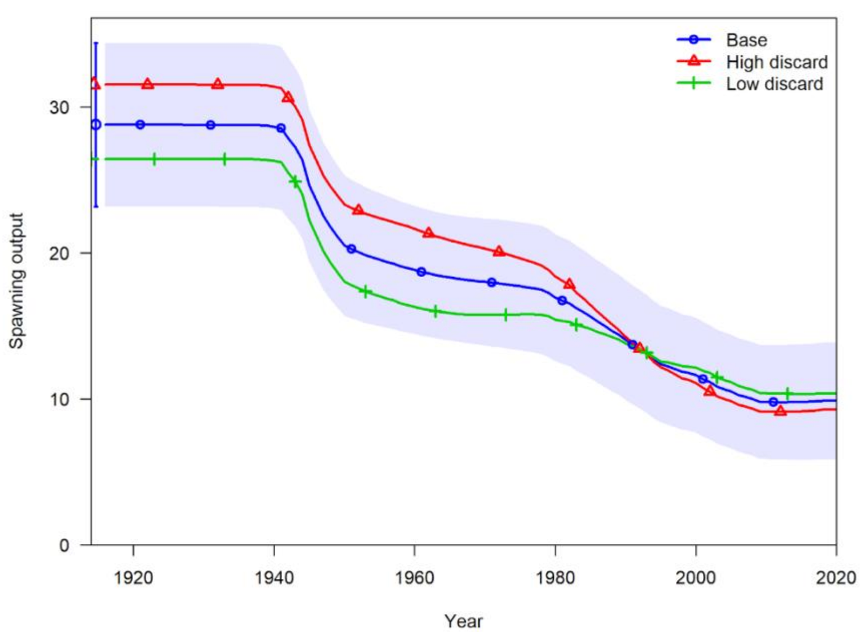

Figure B.8: Time series of spawning output (in millions of fish) associated with different levels of discards.

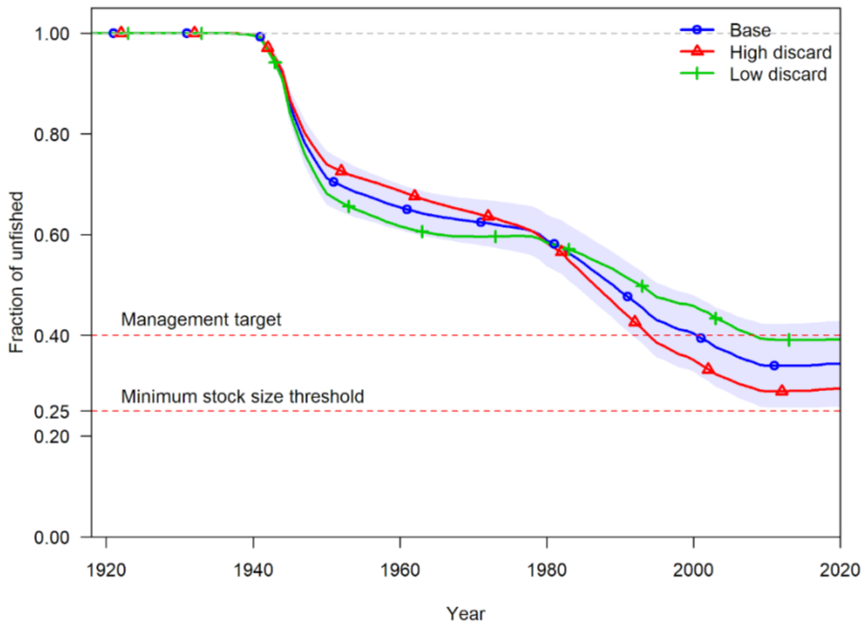

Figure B.9: Time series of spawning depletion associated with different levels of discards.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. While the resulting stock status did not appear to be sensitive to the estimates of historical discards, this issue remains uncertain and unresolved. Estimates of stock status were similar whether q was estimated or fixed. Further research recommendations on this issue appear below.

Reviewers Comments:
Here, the STAT went above and beyond the request by not only providing confidence intervals but runs at both high and low bounds of those intervals. This reviewer agrees with the conclusions of the Panel; the mode doesn't appear very sensitive to the results of using either the upper or lower bounds to a large degree. Although it is noted that there are some larger differences when looking at unfished depletion (Figure B.9), which is in line with expectations.

Request No. 5: Provide the discard rates applied to trawl and non-trawl landings.

Rationale: To better understand these rates

STAT Response:

The STAT provided Figure B.10 and indicated that a table of these rates would be added to the assessment document.

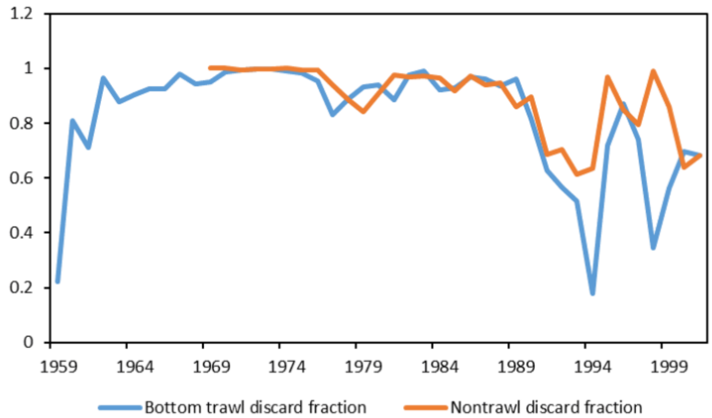

Figure B.10. Historical bottom trawl and non-trawl discard rates.

The STAT also noted that the higher discard rates for the 1980s were supported by Pikitch et al. (1988), while the lower rates in the 1990s were supported by the Enhanced Data Collection Project (EDCP).

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. The STAT also clarified that model-estimated discard amounts used for each sector after 2002 are listed in the draft assessment document.

Reviewers Comments:
This reviewer agreed that this response was met and that the clarification in the document was helpful.

Request No. 6: Provide details on calculating the prediction intervals for the historical bottom trawl discards, and provide the catch streams for the low and high alternative runs (from request #4).

Rationale: To understand how the prediction intervals were calculated, including how the negative values were considered in the low run.

STAT Response:
The STAT indicated that the 95% prediction intervals were calculated using the predict function in base R (https://rpubs.com/aaronsc32/regression-confidence-prediction-intervals) and that a full write-up of these methods including equations would be added to the assessment document. The full time series of discards is shown in Figure 11 with the base, lower, and upper bounds to frame the uncertainty.

.

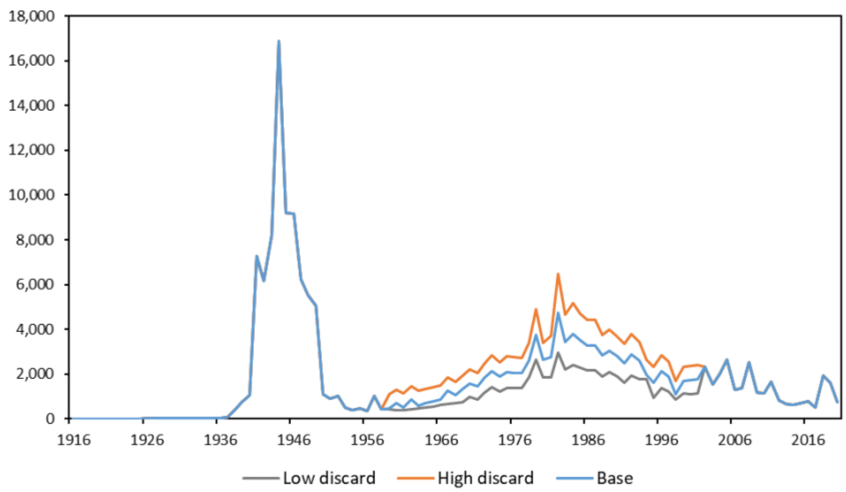

Figure B.11: The full time series of catch streams with low and high historical discard rates.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. A particular issue was the negative confidence intervals near the start of the estimated time period. While these were dropped by the STAT, a better method for this estimation could be conducted. This issue of estimating the historical discards remains an area of uncertainty as outlined in Request 4. The full write-up of the analysis used for dogfish including the above figure should be included in the draft assessment report appendix.

Reviewers Comments:

This reviewer agreed with the comments by the Panel. There is likely a better way to do this sort of analysis; perhaps a more sophisticated analysis of these could be conducted in the future. That said given the model's sensitivity (or lack thereof) there would unlikely be much change in stock status or model results.

Request No. 7: Show the sensitivities from slide #56 (from the day 1 presentation) with the WCGBTS q estimated and a supplemental table displaying the estimated q's for these sensitivities.

Rationale: To understand the behavior of these sensitivities when q is estimated and to see if the q estimates are realistic.

STAT Response:
As shown below, the STAT produced the required figures and table.

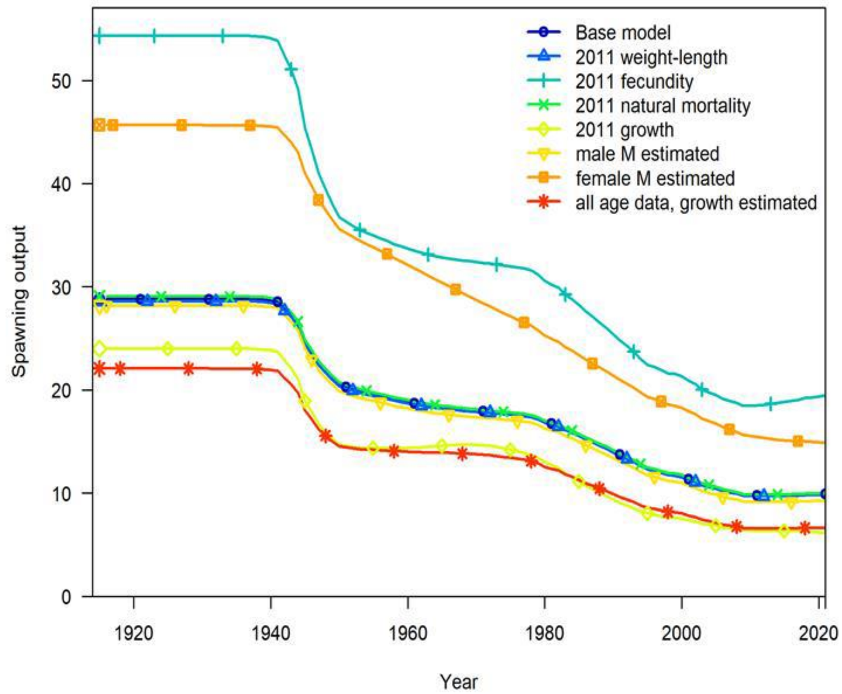

Figure B.12: Sensitivity runs with WCGBTS *q* fixed.

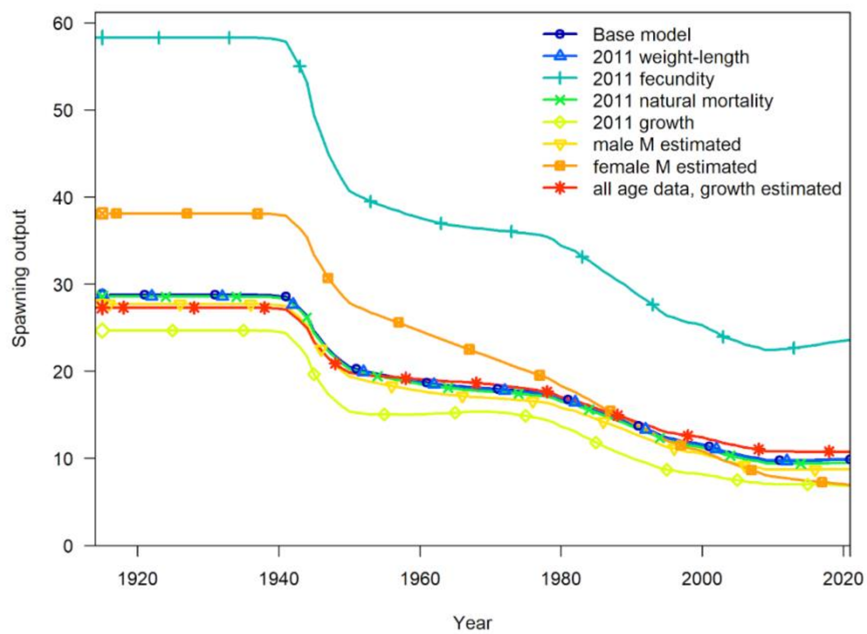

Figure B.13: Sensitivity runs with WCGBTS *q* estimated.

Table B.2: Estimated *q* for the WCGBTS.

	Base model	2011 WL	2011 fecundity	2011 M	2011 growth	Male M estimated	Female M estimated	All age data
WCGBTS q	0.59	0.59	0.49	0.62	0.53	0.61	1.36	0.39

Table B.3: Summary for sensitivities comparing the base model with a run that estimated the *q* for the WCGBTS.

Label	Base model	2011 WL	2011 fecundity	2011 M	2011 growth	Male M estimated	Female M estimated	All age data, growth estimated
TOTAL Likelihood	418.93	418.93	420.138	418.708	454.956	418.891	408.524	703.726
Survey Likelihood Components	-5.21814	-5.2187	-3.92766	-5.24066	-6.3733	-5.41917	-3.91443	-5.31925
Length Likelihood Components	381.387	381.39	381.281	381.194	397.497	381.559	368.102	380.901
Age Likelihood Components	42.7391	42.739	42.7616	42.7318	63.8046	42.7294	42.8571	328.11
Natural mortality females	0.065	0.065	0.065	0.064	0.065	0.065	0.0307681	0.065
Natural mortality males	0.065	0.065	0.065	0.064	0.065	0.0631745	0.026988813	0.065
L1 females	22.064	22.063	22.101	22.040	25.246	22.066	21.460	25.821
Linf females	118.954	118.952	119.912	118.575	109.100	119.058	107.833	122.857
L1 males	22.064	22.063	22.101	22.040	25.246	22.066	21.460	25.821
Linf males	98.680	98.679	99.457	98.417	86.123	98.302	90.048	109.202
von Bertalanffy k females	0.028	0.028	0.027	0.028	0.026	0.028	0.034	0.020
von Bertalanffy k males	0.040	0.040	0.039	0.040	0.052	0.040	0.049	0.025
LnQ_base_WCGBTS(11)	-0.53364	-0.533	-0.717794	-0.482637	-0.6431	-0.489487	0.309194	-0.952754
WCGBTS q	0.59	0.59	0.49	0.62	0.53	0.61	1.36	0.39
SOVirgin (millions of fish)	28.778	28.586	58.323	28.562	24.682	27.706	38.104	27.319
SO 2021 (millions of fish)	9.895	9.829	23.616	9.494	6.849	8.781	6.969	10.733
B ratio_2021	0.343834	0.3438	0.404912	0.332397	0.2775	0.316923	0.182885	0.39289
SPR ratio 2020	0.282184	0.2823	0.230211	0.296184	0.3568	0.311266	0.701907	0.225196

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. The Panel noted that when both M and q were allowed to be estimated, the model tended to settle on the same scaling of biomass; especially in the most recent years. The Panel expressed some reservations that q was estimated implausibly high (1.36) when M for females was also estimated.

Reviewers Comments:
This reviewer agreed with the comments given by the full Panel. The issue of high catchability is a difficult one and it's obvious that the model is tending to have such high catchability when M is also estimated. However, the fact that a similar scaling is achieved was a good sign.

Request No. 8: Provide a sensitivity to the estimated female k values by fixing k at 0.065; estimate male k as an offset.

Rationale: The M/k ratio is atypical for elasmobranchs; a ratio of 1 - 2 is more typical and a fixed k at 0.065 provides a 1.0 M/k ratio.

<u>STAT Response</u>:
The STAT provided the requested runs (see below). The STAT noted that while female growth parameters changed, those for males did not.

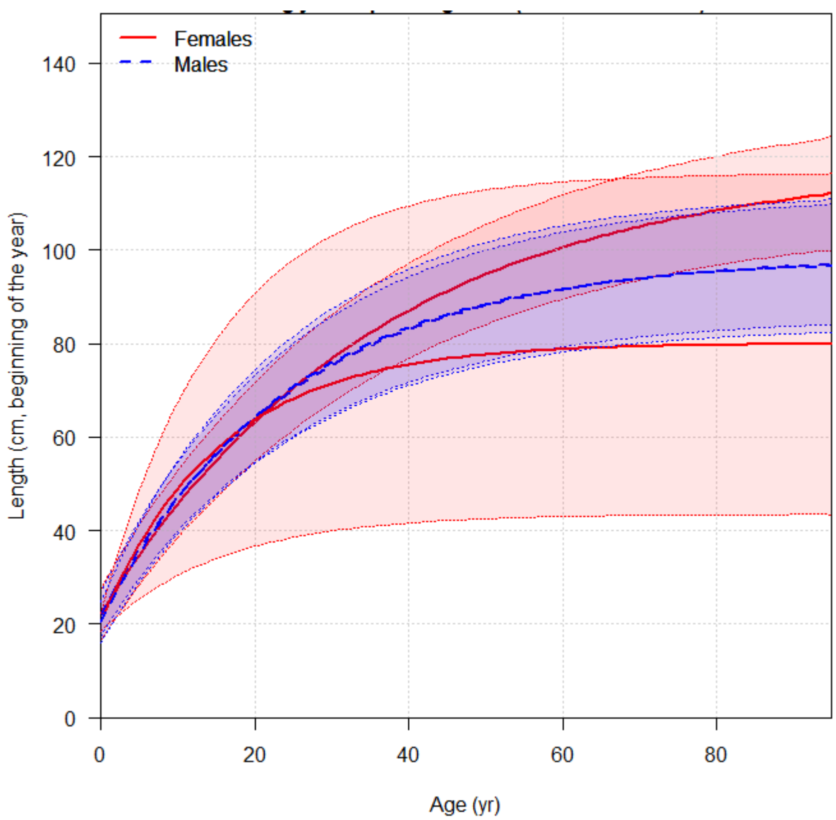

Figure B.14: Growth curves for males and females when fixing female k at 0.065 and estimating male k as an offset.

Moreover, this change in female growth parameters was not consistent with the observed biology of dogfish as this run resulted in females attaining a lower size than males. Additionally, the requested run had an effect on the resulting scale of the model, but little effect on the overall depletion (see below).

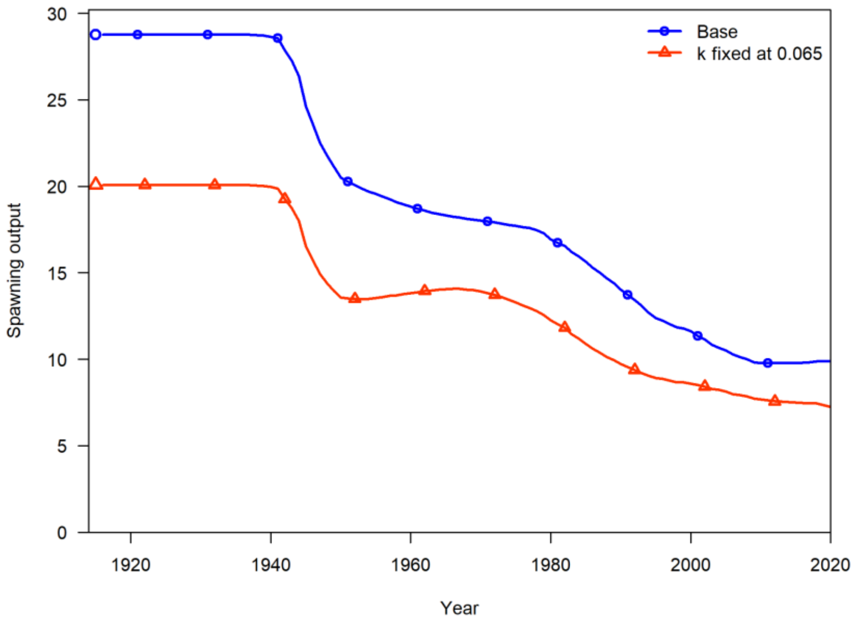

Figure B.15: Time series of spawning output (in millions of fish) associated with different *k* values.

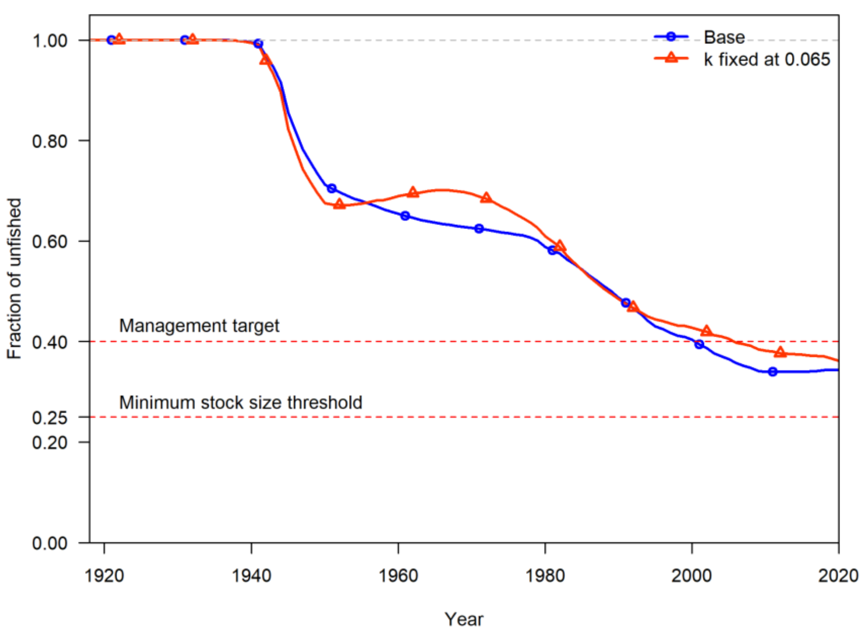

Figure B.16: Time series of spawning depletion associated with different k values.

<u>Panel Conclusions</u>:
The Panel was satisfied with the STAT's response to the request. Moreover, the Panel agreed with the conclusion that fixing female k at 0.065 and estimating male k as an offset was not an improvement to the base model.

<u>Reviewers Comments</u>:
This reviewer agreed with the Panels conclusions to fix the k at 0.065.

Request No. 9: Provide runs where female M is estimated and WCGBTS q is estimated and fixed. Provide fits and other diagnostics for these runs.

Rationale: We need a better rationale for the choice of female M and why model fits appear to improve with lower M.

STAT Response:
The STAT provided the requested runs. When female M was estimated alone and when both female M and q from the WCGBTS have estimated results indicated a higher scaling of the model, as shown in Figure B.17.

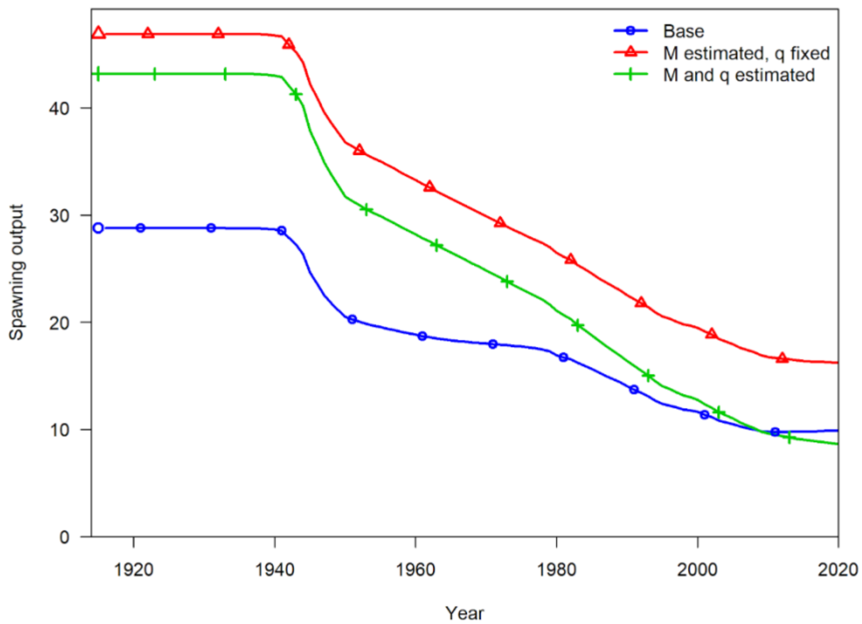

Figure B.17: Comparing spawning output estimated by the base model and two models with different M and WCGBTS *q* settings.

Table B.4: Likelihood of the Base model, M estimated in the model, and both M and *q* estimated in the model.

Label	Base model	M estimated	M and q estimated
TOTAL Likelihood	418.93	415.30	409.90
Survey Likelihood Components	-5.22	-2.86	-4.19
Length Likelihood Components	381.39	373.93	369.49
Age Likelihood Components	42.74	43.08	42.86
Priors Likelihood Components	0.00	1.14	1.72
Natural mortality females	0.065	0.034	0.029
Natural mortality males	0.065	0.034	0.029
L1 females	22.06	21.67	21.54
Linf females	118.95	109.34	107.46
L1 males	22.06	21.67	21.54
Linf males	98.68	90.01	90.89
von Bertalanffy k females	0.03	0.03	0.03
von Bertalanffy k males	0.04	0.05	0.05
LnQ_base_WCGBTS(11)	-0.53	-0.53	0.30
WCGBTS q	0.586	0.586	1.356
SOVirgin (millions of fish)	28.78	46.90	43.18
SO 2021 (millions of fish)	9.90	16.20	8.55
B ratio_2021	34%	35%	20%
SPR ratio 2020	0.56	0.38	0.67

Likelihood profiles showed this contrast in the data (below) where the most recent surveys (WCGBTS and IPHC) were fit best with higher M, while two sources of length comps (non-trawl landing and IPHC) are best fit with lower M.

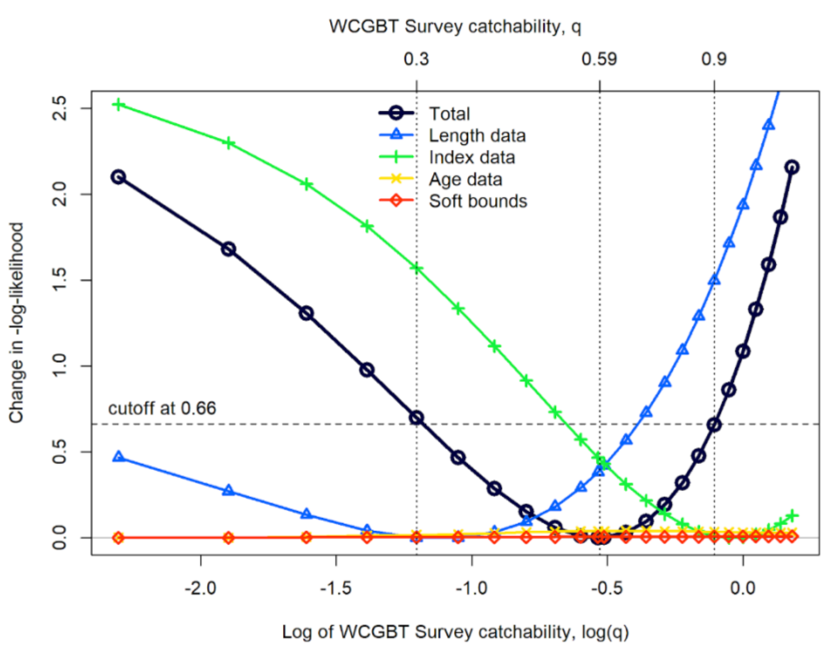

Figure B.18: Likelihood profile over WCGBTS *q*.

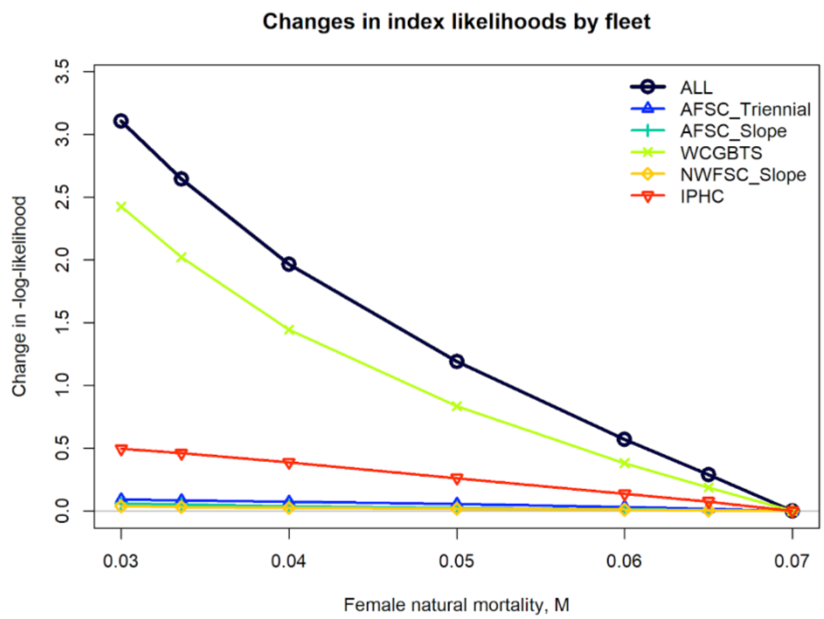

Figure B.19: Likelihood profile over M by survey.

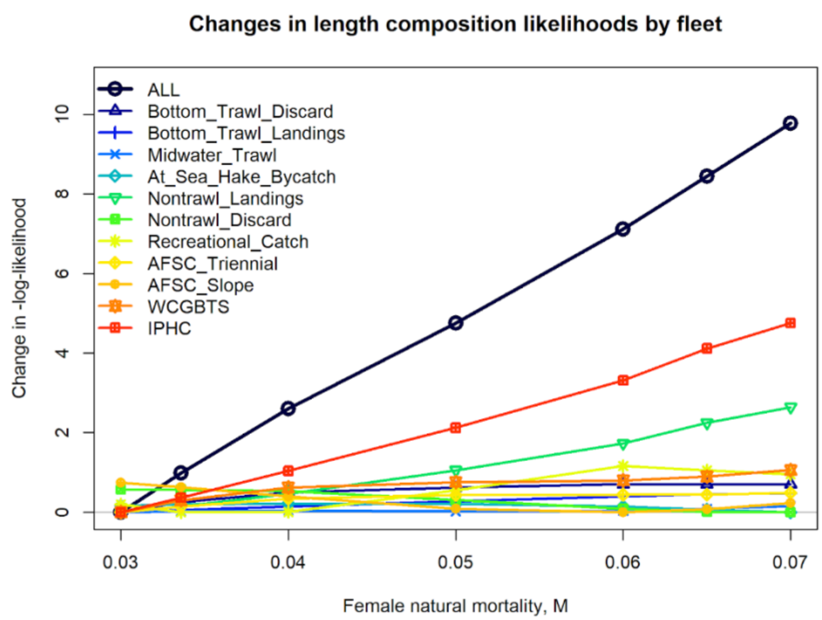

Figure B.20: Likelihood profile over M by fishing fleet and survey.

Overall, changing M did have a large effect on the scale of the model, but it had little effect on either stock status or depletion, as shown in Figures B.21 and B.22.

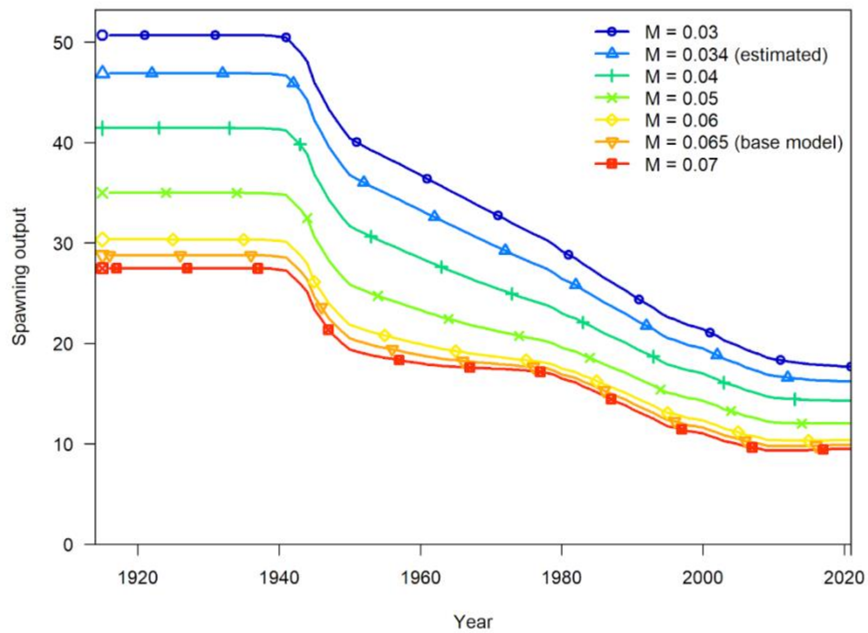

Figure B.21: Time series of spawning output (millions of fish) associated with different values of natural mortality (M).

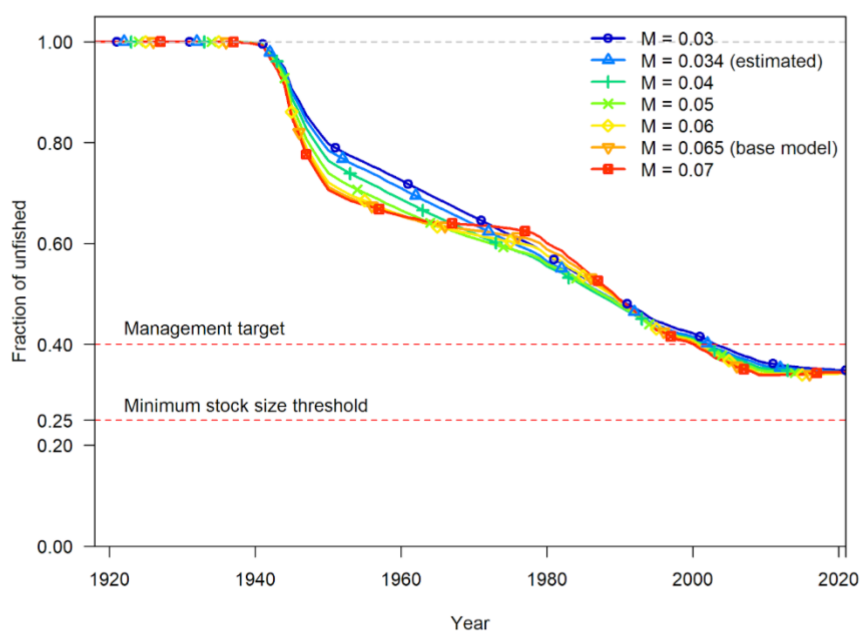

Figure B.22: Time series of spawning depletion associated with different values of natural mortality (M).

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. The Panel agreed with the STAT's conclusion that the index abundance and the index lengths were pulling the M estimates in different directions. This is also shown in the response to Request 10.

Given that the model's scale is sensitive to changes in M, the Panel concluded that this is an unresolved issue worthy of a research recommendation.

Reviewers Comments:
This reviewer agreed with the Panels comments and is satisfied with the STATS response. There is tension in this model between the index abundance data and the indices length composition data. This is somewhat troubling as the data are ultimately coming from the same source, the indices.

Request No. 10: Provide runs with WCGBTS q values of 0.3 and 0.9 with an accompanying likelihood profile.

Rationale: To explore potential values and states of nature for the proposed axis of uncertainty in the decision table.

STAT Response:
The STAT provided the requested runs, displayed below in Figure B.23.

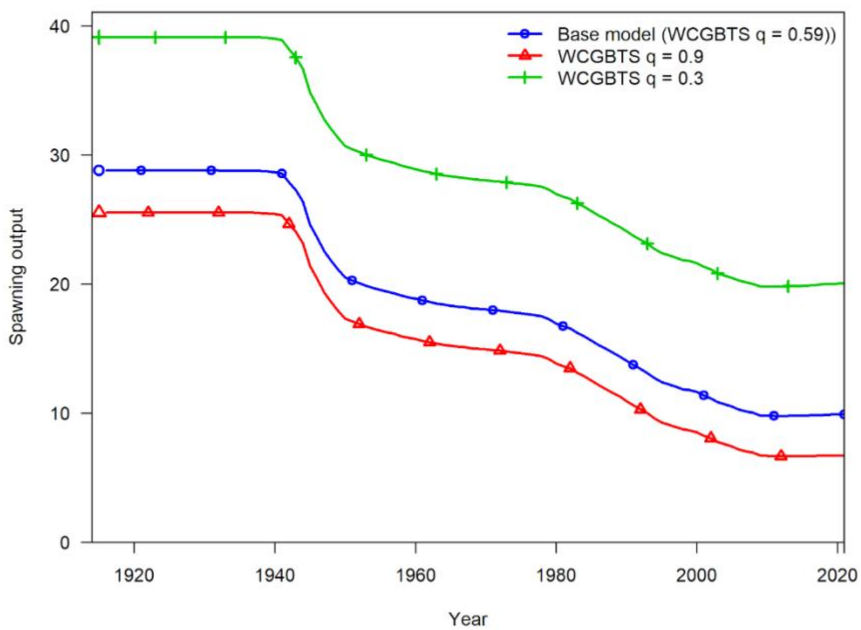

Figure B.23. Time series of spawning output (in millions of fish) associated with different values of WCGBT *q*.

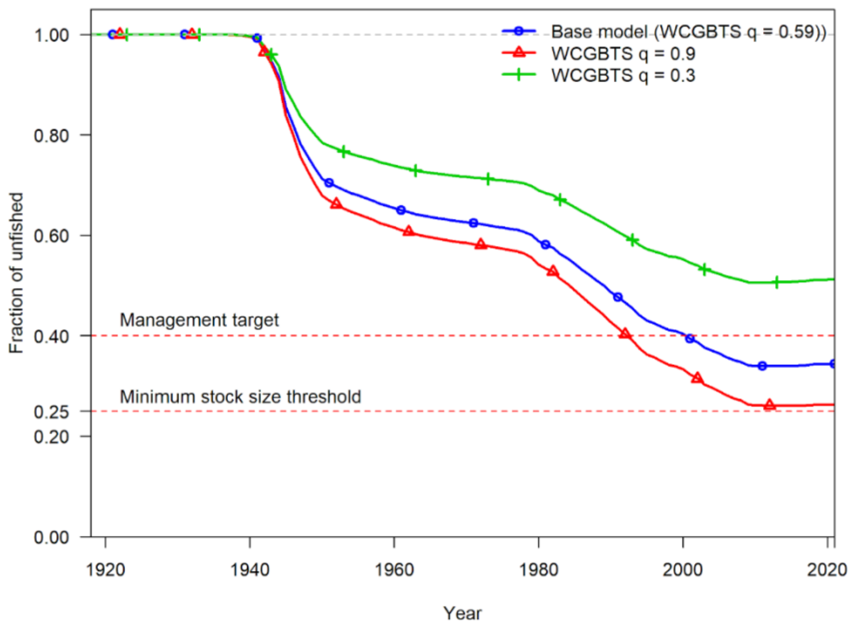

Figure B.24: Time series of spawning depletion associated with different values of WCGBTS *q*.

The STAT proposes these two models (WCGBTS *q* = 0.3 and 0.9) as alternative states of nature with *q* as the axis of uncertainty.

Overall, the likelihood profiles (below) again showed the contrasting signal of the length and index data in this assessment.

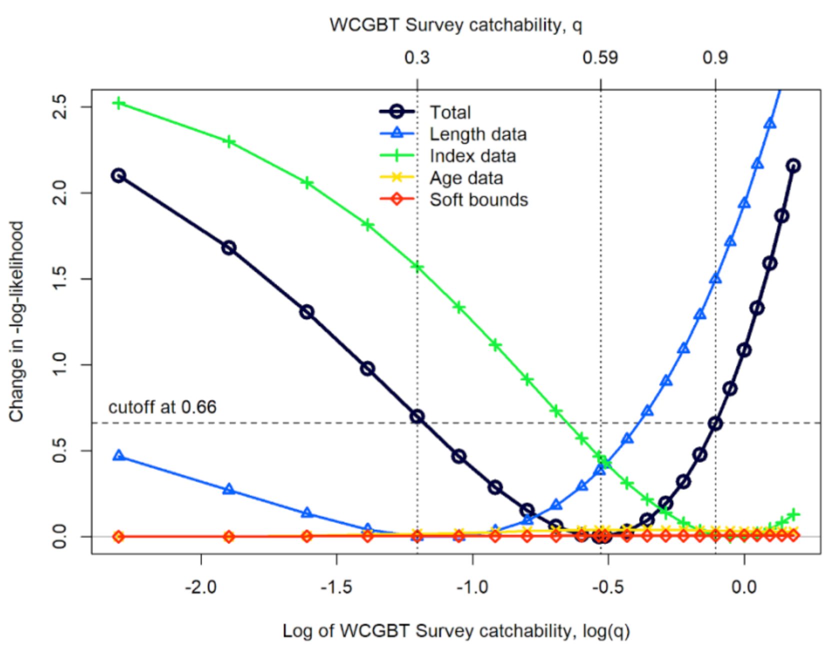

Figure B.25: Likelihood profile over WCGBTS q by data type.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. It also noted that q is a major uncertainty in this model formulation and, in particular, one that affects the overall scale of the model greatly. The tension between length and index data pulling q in opposite directions is an unresolved issue worthy of further research. The Panel further noted that different length data were also in conflict with each other. After discussion, the Panel formulated Request 11 to, in part, attempt to decide if q or R_0 would be best to capture this scaling uncertainty.

During the meeting, the panel also evaluated length composition likelihood contributions by data source (Piner plots), as well as Francis weighting fits to the mean lengths by year. WCGBTS exhibits the increasing mean length, which along with declining index suggests a potential decline in recruitment (consistent with a decline in spawning females), which could also explain why that data source is best fit at higher catchability values associated with lower stock sizes as shown in the likelihood profiles. Increasing trends in both observed and expected values of mean length from composition data are also observed for bottom trawl discards, which are shown in Figures B.26-B.28.

Reviewers Comments:
This reviewer agreed with the Panels conclusions. Overall, this tension in the data, within the index data is worthy of further research (see below). Catchability studies or some other work on the indices could be informative in solving this problem in future assessments.

Request No. 11: Provide runs where ln(R0) is the axis of uncertainty with WCGBTS q estimated with an accompanying likelihood profile.

Rationale: To explore potential values and states of nature for the axis of uncertainty in the decision table.

STAT Response:
The STAT provided the requested runs. They noted that the initial profiles with q fixed resulted in a very narrow range of alternative models. With q estimated, models with $\log(R_0)$ = 9.6 and 10.05 were close to the 0.66 cutoffs typically used. Alternative values of q associated with these runs are very similar to the proposed alternative states of nature, based on q as shown below.

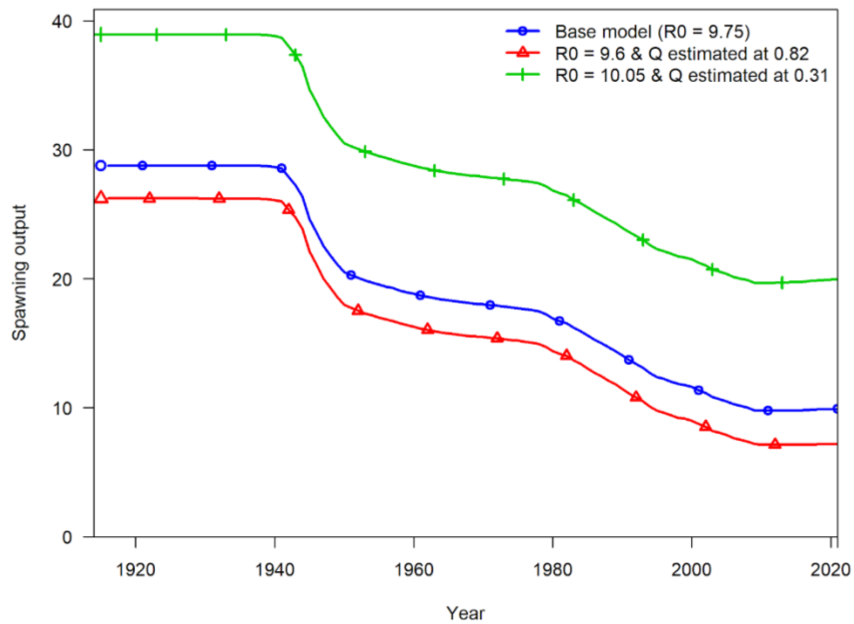

Figure B.26: Time series of spawning output (in millions of fish) associated with different values of R0.

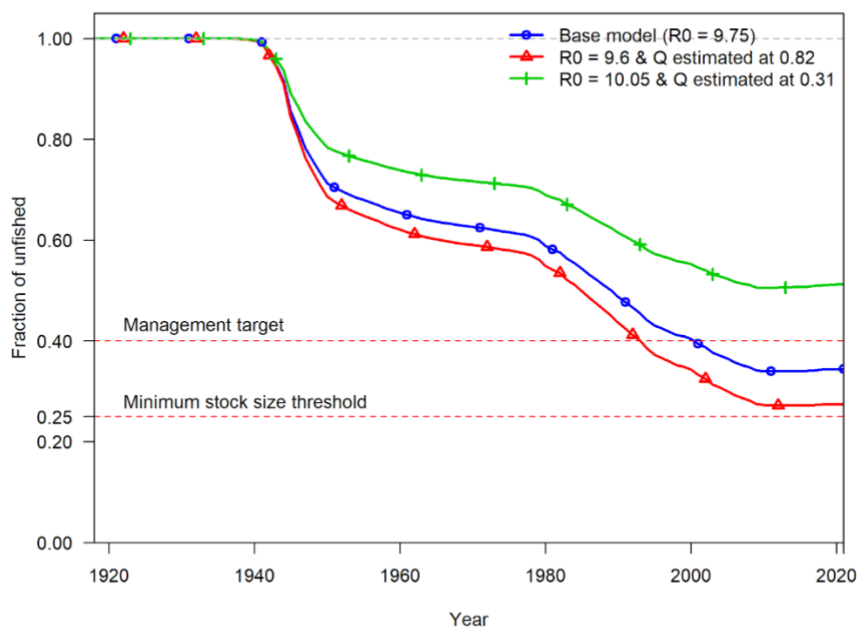

Figure B.27: Time series of spawning depletion associated with different values of R0.

Likewise, the likelihood patterns were very similar to using *q* as the axis of uncertainty. The STAT felt that *q* would be a better axis of uncertainty given that it was more explainable to stakeholders and the general public.

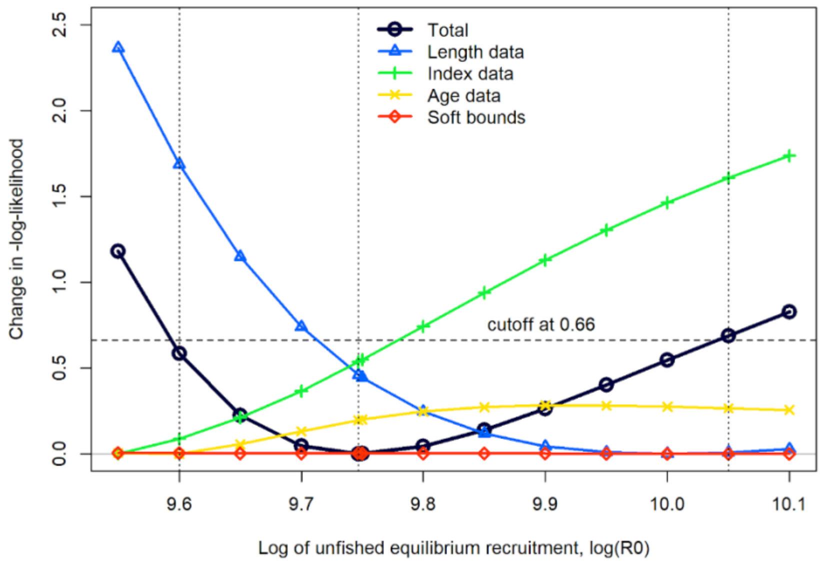

Figure B.28: Likelihood profile over log(R0) by data type.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. They noted and agreed that using R_0 or q is acceptable. They affirmed the recommendation by the STAT that q should be the axis of uncertainty at the values specified above.

Reviewers Comments:
While this reviewer did agree with the comments of the Panel, it is this reviewer's opinion that q is a better axis of uncertainty than R0. R0 is often not that meaningful for stakeholders and managers, while q is more familiar to most. As such, and to help somewhat with buy-in, the axis should be catchability, rather than R0.

Request No. 12: Repeat request #4 and evaluate the sensitivity of the historical discard assumptions under each catch stream when WCGBTS q is estimated. Reproduce the figures under request #4 with an accompanying table of the q values and other model outputs. Also, provide the total biomass time series under each of these scenarios.

Rationale: To examine estimated q among different historical discard assumptions.

STAT Response:
The STAT produced both estimates with q either fixed or allowed to be estimated (below).

With WCGBTS *q* estimated:

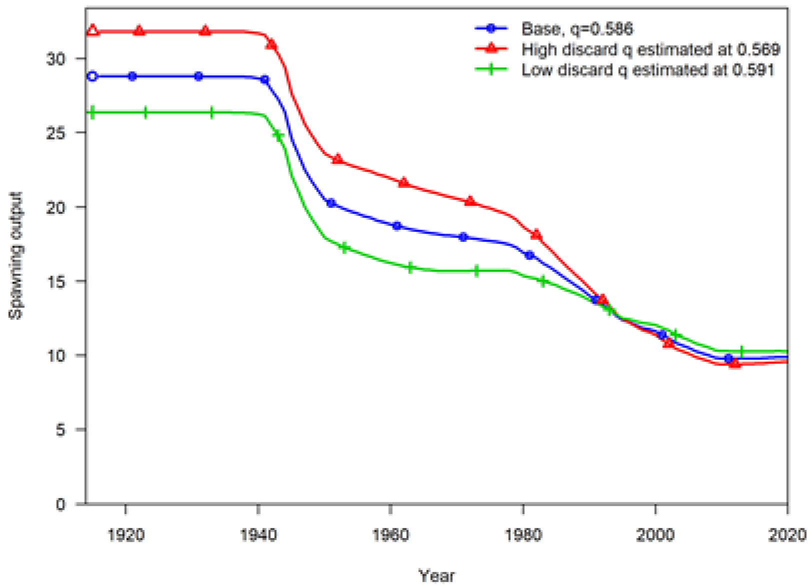

Figure B.29: Time series of spawning output (in millions of fish) associated with different discard rates and estimating WCGBTS *q*.

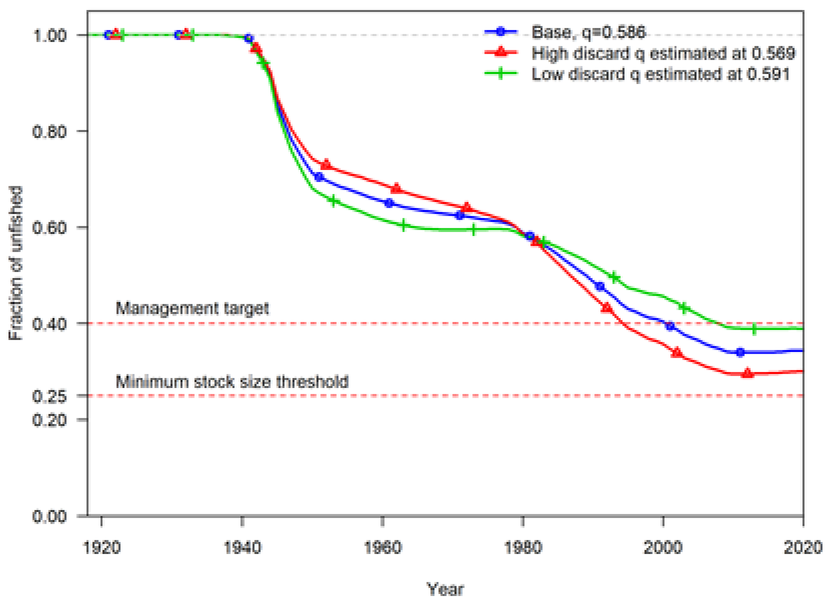

Figure B.30: Time series of spawning depletion associated with different discard rates and estimating WCGBTS *q*.

With WCGBTS *q* fixed:

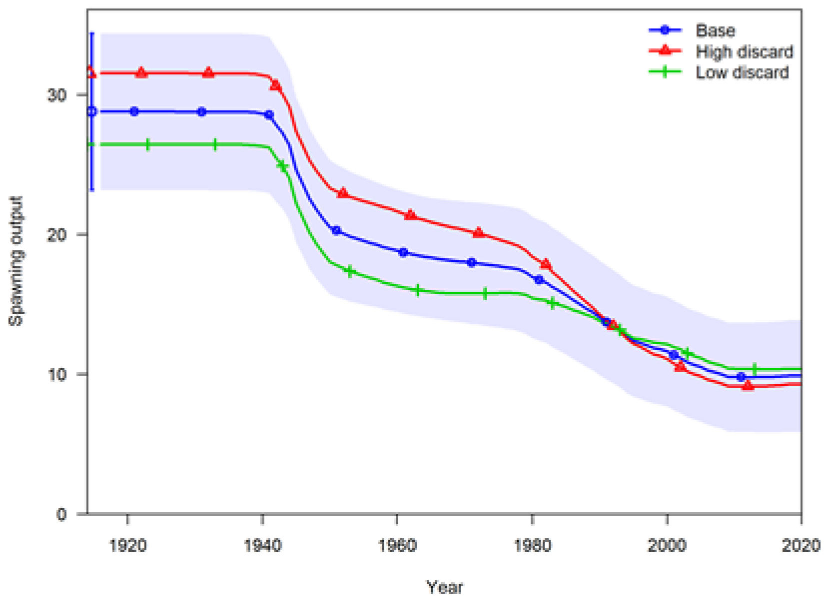

Figure B.31: Time series of spawning output (in millions of fish) associated with different discard rates and WCGBTS *q* fixed.

Figure B.32: Time series of spawning depletion associated with different discard rates and WCGBTS *q* fixed.

As can be seen by Figures B.29 & B.30 vs. Figures B.31 & B.32, there was little difference between allowing the catchability to be estimated or fixed. Estimated catchability was 0.6, similar to the fixed value of 0.59.

Panel Conclusions:
The Panel was satisfied with the STAT's response to the request. After discussion, the Panel concluded that this was a non-issue. The STAT reminded the Panel that the estimate of discards with high uncertainty was mostly from years 1960-2002, well before the use of the WCGBT survey; as such, a change in the effect of catchability would not be likely to occur given the differences in the timing of these two data inputs. The Panel agreed with this conclusion.

Reviewers Comments:

After consideration, this reviewer also agreed with the STAT's response and conclusions. In retrospect, one wouldn't expect there to be much difference in the index's current catchability with a change in estimated discards from 1960-2002, as there have been almost two decades since more complete discard information was available. In short, the estimate of discards happened well before the start of the WCGBT survey.

Strengths

This assessment had a number of strengths associated with it. Overall, the STAT did a thorough job in documenting the sensitivity analysis, data choices, and other aspects of the assessment. The document was clearly laid out and the presentation was professionally conducted. In addition, the STAT was more than accommodating with the sheer number of requests given to them.

A clear improvement in the model from 2011 is the use of the sablefish fishery to estimate dogfish discards from 1960 -2002. This represents a major step forward in estimating historical discards and thus removals for this fishery. While there are some areas of improvement (see below), it is this reviewer's opinion that this analysis represents the best data available for this stock.

Another vast improvement over the 2011 assessment was the incorporation of aging data and the estimation of growth. Again, while there are improvements that can and should be made for the next assessment, this is a dramatic improvement and is the culmination of much work.

Another strength, likely the result of both improved growth and aging data, is the lack of a retrospective pattern. There was a rather large retrospective pattern in the 2011 assessment, and for the current assessment, not much if any bias.

Weaknesses

While the model has improved since 2011 there are a number of ongoing issues and uncertainties.

This model is very sensitive to assumptions catchability, as would be expected. In this case, however, the data tends to pull the model in different directions when it comes to scaling, and this is a major source of uncertainty that should be assessed in earnest in any follow-up assessments. Moreover, the assessment model is looking for much higher catchability as a result of the index abundance (0.9), when compared to the index length composition (0.3).

The model is also sensitive to assumptions on natural mortality. Like with catchability, this is a major uncertainty in part influenced by the aging uncertainties (below), affecting perceptions on growth resilience, and other life-history traits.

Another factor that contributes to uncertainty is aging. Worn spines of large dogfish present a difficult problem to overcome, resulting in the removal of large and older females from the growth estimations. Additionally, both natural morality and maximum age are also very uncertain for this stock. Suggesting that fundamental life-history aspects for this stock are not well understood.

An overarching issue, at least to this reviewer, is the issue of unit stock. As outlined below there are some issues dealing with the stock's assumed range and the data included in the assessment. While it may not be a driving force in the uncertainty, taken together with other aspects, can lead to some major issues if not rectified or at least explored further.

Research recommendations

The Panel, including this reviewer, made several Research recommendations. These were then subdivided into two groups; those that should be done before the next assessment, and those that can wait. Comments on research recommendations by this reviewer are given at the end of this section

Research to be done prior to the next assessment attempt.

The Panel also supported the STAT's recommendation that all ongoing data streams used in this assessment be continued or increased including fishery-dependent sampling for length, age, and maturity, as well as fishery-independent collection and aging. Fishery-dependent samples should be collected in light of changing fleet dynamics and to fully cover the range of the current fishery.

Additionally, the approaches for informing the historical discards of Spiny Dogfish should be reevaluated, and existing literature reexamined. If the preferred method continues to be examining the total catch of Spiny Dogfish in association with the total catch of Sablefish in recent years of at-sea observations, the Sablefish catch data should be parsed to the portion of the fishery on the shelf where Spiny Dogfish occur by excluding trawl efforts on the slope. This could be done by excluding winter trawl effort for Sablefish or by using a MacCall-Stephens approach of filtering out efforts where Sablefish are caught with Dover Sole and thornyheads, which is indicative of slope targeting.

As also recommended by the STAT, the Panel suggests that a vigorous examination of natural mortality via meta-analysis be conducted to help in establishing informative priors for M for future assessments. This analysis should be linked to other parameters such as growth.

Like most other assessments, estimates of catchability (q) are a major source of uncertainty and an unresolved issue for this assessment. This is especially true for dogfish as they appear to be semi-pelagic and may not be consistently available to bottom trawling. As such both the STAT and the Panel recommend future research into the catchability of dogfish in the WCGBTS. These could include depletion studies, video surveillance of trawl operations, or other analyses as appropriate, such as bench-top analysis of co-occurring fishery-dependent/independent data.

Given the issue that worn spines of older females may produce an aging bias, the panel recommends that research be conducted to examine this issue in detail. The Panel suggests a re-examination of existing data, models, and methods used to derive age and growth.

Research needed at some point in the future.

Given the densities of large schools of dogfish adjacent to the US-Canada border, the Panel supported the STAT recommendation that the next assessment is conducted jointly with DFO Canada as a potential transboundary assessment. Before that, research on tagging might be helpful in either reaffirming the current 5% straying rate or updating it.

As outlined in the assessment report, efforts should be devoted to both improving current aging techniques based on dogfish spines and developing new methods using other age structures. Ideally, an alternative method of aging dogfish that does not rely on the estimation of ages missing from worn spines may be necessary. Improvement in aging would contribute to a better understanding of Spiny Dogfish longevity and would help estimate natural mortality as well as inform growth parameters within the assessment model.

Reviewer Comments:

While this reviewer generally agreed with the comments of the Panel, at least for consensus purposes, there are a few areas where full agreement is tenuous. In particular, the discussions around a transboundary assessment as well as the other discussions for follow-up assessment work.

In the first case, transboundary issues are not something new for this stock. While the straying rate may only be 5% or so per year, over the life span of a dogfish, this can result in large uncertainties. Additionally, this straying rate has not been updated recently, and there may be a different outcome given the climate-induced changes in this area recently.

While a full transboundary assessment might be logistically infeasible and politically difficult, it's important to remember that misspecification in the unit stock can present itself in many modeling uncertainties, from selectivity through catchability and beyond. Ensuring a unit stock should be the first order of business for any assessment. In addition, while difficult, some preliminary work and be conducted by using Canadian catch and sampling data as a sensitivity, or perhaps including Department of Fisheries and Oceans membership on the STAT temporarily might be useful.

Another area worthy of discussion is the idea that this assessment should come back for further review during this STAR. While possible, it's important to remember that until some, or most, of the research outlined above, is completed, the results are not likely to change much. Although there could certainly be some improvements in the way discards are calculated those are unlikely to be easy or quick to do in the interim. As such it's important to ensure that many of the highlighted research topics be address prior to bring this assessment back to full peer review.

Conclusions

The STAT did an excellent job presenting the information to the Panel and were very accommodating in the number of requests put to them. Like with Dover sole, despite the sheer number of requests the Panel agreed with the STAT that the base model represented the best available data from which management decisions could be made. This reviewer agrees whole-heartly with that consensus.

There have been major improvements in the mode since the 2011 STAR. Chiefly the use of aging data to allow for the model to utilize growth. The improvement in the retrospective pater cannot be understated. This was a major uncertainty in the 2011 assessment, and the work the STAT put into resolving this issue should not be overlooked. Additional work has been done on estimating discards from 1960-2002. Also, the method employed in this assessment is a vast improvement over the 2011 analysis. While there was some consternation during the Panel meeting, by non-Panel members, on this issue, it is important to remember that this estimate of discards is the least concerning when it comes to model issues.

And this model has issues. In particular, the need to fix natural mortality to get a reasonable catchability. This in turn was compounded by the uncertainty in natural morality estimated outside the model due to uncertainty in the aging. As mentioned elsewhere there is a large conflict between abundance index data and the length data from those very same indices.

Some on the Panel, as well as others, had concerns about fixing catchability in subsequent diagnostics runs. While this is a common practice, the STAT is advised to carefully consider, at least in future assessments, presenting both the fixed and unfixed diagnostics. This can give reviewers the ability to look at the trade-offs between some parameters and others, allowing for a choice on whether changes to model structure give realistic catchability and natural mortalities.

The inclusion of growth in the model was a welcomed improvement over the 2011 assessment. However, the issue of worn spines and potentially biased ages created another issue in the model that is not easily rectified without extensive research and study. This aging uncertainty affects many aspects of this assessment, from growth and natural mortality through estimations of maximum age.

The research recommendations outline elsewhere (above) may well help to “fix” or at least mitigate some of these uncertainties. That said, the issue of unit stock and immigration/migration with portions in Canadian or Gulf of Alaska needs to be examined. While by itself untested here, issues of the unit stock can be a major issue compounding already uncertain parameter estimates.

CONCLUSIONS AND RECOMMENDATIONS

Overall, this was an enjoyable set of stocks to review. Both STATs did a wonderful job organizing and presenting the data and analysis at hand. Both were also very accommodating in the numerous requests by the Panel and went well beyond by providing analysis they thought would help elucidate the major issues. Likewise, Theresa did a great job as chair, keeping Panel members on track and the meeting running smoothly. Panel members as well as others made well-formed and concise comments and overall, the professionalism was beyond reproach. Those other Panel members made this review enjoyable, stimulating, and just plain fun.

However, during one particularly heated exchange during dogfish, non-Panel members' emotions seemed to get the best of them. While understandable, it did create an uncomfortable few moments that could have been avoided. It is well recognized that the dogfish assessment may have difficult management implications, thus it is even more important that professionalism be maintained. It wasn’t a “big deal”, but one wonders if it would have taken place, or been as awkward, had the meeting been in person.

As the (perhaps) first completely online STAR Panel logistics went very well. Reports and materials were provided timely, data and even code were provided to the Panel in the sharefiles or FTPs. Given the COVID situation, it was welcome to not have to travel to this meeting. It did create some issues. The lack of being in the same room as other Panel members made it difficult, at times, to fully understand their points. Additionally, that collaborative atmosphere between Panel and STAT was certainly missing, leading to a more formal interaction than necessary. This reviewer found it difficult to operate in two time zones at the same time, being in the eastern US and attending a meeting in the western US was more difficult than expected. An online meeting was more convenient, likely cheaper, and definitely safer, but may not be better than holding an in-person meeting.

As mentioned previously this process ran very smoothly. This was in part due to the TOR document provided at the start of the Panel deliberations which outlined the best practices and in particular the use of an axis of uncertainty. This type of document was, in short, a monumental improvement over other peer review Panels elsewhere. It clearly laid out the expectations, the products needed from the STAT, as well as how to handle uncertainty for decision making. That said, during the course of the discussions there was a tendency by STATs, as well as others, to over-inflate some of the uncertainties as a precautionary measure for other “unaccounted for” uncertainties. While understandable and not consciously done, this reviewer’s opinion is that uncertainty attributed to a particular issue should be realistic, and an additional “precautionary buffer” used if the managers/SSC feel it is warranted. To do otherwise compromises the transparency of the process and could damage credibility in the future. The use of an axis of uncertainty, and the outline of how to display it in the TOR document, was a great improvement in the process (one that more councils should adopt) but it’s important to place the uncertainty within reasonable contexts.

This reviewer is comfortable recommending both Dover sole and Spiny dogfish base models be used to craft management advice to these stocks. Additionally, the process and the modeling have greatly improved from the 2011 STAR process. This reviewer is looking forward to other Panels during the STAR season over this upcoming summer.

APPENDICES

Appendix 1: Bibliography of materials provided for review

Hicks, Allan C., and Chantel R. Wetzel. 2011. "The Status of Dover Sole (*Microstomus pacificus*) Along the U.S. West Coast in 2011." 7700 Ambassador Place NE, Suite 200, Portland, OR 97220: Pacific Fishery Management Council.

Gertseva, V.V., and Matson, S.E., 2021. Right on target: using data from targeted stocks to reconstruct removals of bycatch species, a case study of longnose skate from Northeast Pacific Ocean. Fisheries Research. Vol. 236 (2021) 105841. https://doi.org/10.1016/j.fishres.2020.105841.

Gertseva, V., Taylor, I.G. 2012. Status of the spiny dogfish shark resource off the continental U.S. Pacific Coast in 2011. Pacific Fishery Management Council, Portland, OR.

Method, R., and Wetzel, C., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142 (2013) 86– 99. http://dx.doi.org/10.1016/j.fishres.2012.10.012.

Method, R., and Wetzel, C., Taylor, I., and Doering, K., 2020. Stock Synthesis User Manual Version 3.30.16. NOAA Fisheries, Seattle, WA.

Miller, S., Stephens, A., Whitmire, C., and Hastie, J., 2021. Overview of West Coast Groundfish Fishery-Independent Surveys. Northwest Fisheries Science Center, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, Washington 98112.

Taylor, I.G, Gertseva, V.V., Matson, S.E., 2013. Spine-based ageing methods in the spiny dogfish shark, Squalus suckleyi: How they measure up. Fisheries Research. Vol. 147 (2013) Pages 83– 92. http://dx.doi.org/10.1016/j.fishres.2013.04.011.

Taylor, I.G, Gertseva, V.V., Methot, R.D., and Maunder, M.N., 2013. A stock–recruitment relationship based on pre-recruit survival, illustrated with application to spiny dogfish shark. Fisheries Research. Vol. 142 (2013) Pages 15-21. https://doi.org/10.1016/j.fishres.2012.04.018.

Thorson, J., and Barnett, L., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science (2017), http://doi:10.1093/icesjms/fsw193.

Appendix 2: A copy of the CIE Performance Work Statement

Performance Work Statement

External Independent Peer Review by the Center for Independent Experts

Stock Assessment Review (STAR) Panel 1 - Virtual

Dover Sole and Pacific Spiny Dogfish

May 3-7, 2021

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards.
(http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).
Further information on the CIE program may be obtained from www.ciereviews.org.

Scope:

The National Marine Fisheries Service and the Pacific Fishery Management Council will hold three stock assessment review (STAR) panels and potentially one mop-up panel if needed, to evaluate and review benchmark assessments of Pacific coast groundfish stocks. The goals and objectives of the groundfish STAR process are to:

1) ensure that stock assessments represent the best available scientific information and facilitate the use of this information by the Council to adopt OFLs, ABCs, ACLs, (HGs), and ACTs;

2) meet the mandates of the Magnuson-Stevens Fisheries Conservation and Management Act (MSA) and other legal requirements;
3) follow a detailed calendar and fulfill explicit responsibilities for all participants to produce required reports and outcomes;
4) provide an independent external review of stock assessments;
5) increase understanding and acceptance of stock assessments and peer reviews by all members of the Council family;
6) identify research needed to improve assessments, reviews, and fishery management in the future; and
7) use assessment and review resources effectively and efficiently.

Benchmark stock assessments will be conducted and reviewed for the Dover sole and Pacific spiny dogfish. These stocks were identified within the top five rankings for assessment consideration during the Pacific coast groundfish regional stock assessment prioritization process, which was based on the national stock assessment prioritization framework (http://www.st.nmfs.noaa.gov/Assets/stock/documents/PrioritizingFishStockAssessments_FinalWeb.pdf).

Dover sole was last assessed in 2011, and estimated stock depletion in that year was 83.7 percent of its unfished biomass at the start of 2011 (Hicks and Wetzel, 2012). A catch-only projection update of that assessment was conducted in 2019, and estimated depletion at that time was 77.6 percent. Dover sole range from Baja California to the Bering Sea and eastern Aleutian Islands; however, the assessment addresses that portion of the stock caught in the fisheries off California, Oregon, and Washington. Dover sole are highly important to the commercial fishery; however, modeling difficulties arise from the fact that females grow to be much larger than males and display ontogenetic movement to deeper waters as they age, making the older females unavailable to the fishery and to the West Coast Bottom Trawl Survey. The attainment for Dover sole is constrained by the fishery for Sablefish, as well as by market considerations.

Pacific Spiny Dogfish off the U.S. West Coast was last assessed in 2011 (Gertseva and Taylor 2012), which estimated stock depletion to be 63.2 percent of unfished spawning biomass at the start of 2011. The species range is from Baja California to the Bering Sea; however, the assessment addresses the portion of the stock caught in the fisheries off California, Oregon, and Washington. Seasonal movement of some dogfish between the higher-density areas off Washington and British Columbia is likely in many/most years. Because dogfish lack otoliths, traditional methods of aging used for other groundfish species are not available. Instead, dorsal spines are used to determine age. Although these spines exhibit readable annuli, they are subject to wear, over time, which increases aging uncertainty, particularly for older fish. Additionally, preparing the spines for aging is a time-consuming, multi-step process, which has severely limited the availability of age data for use in assessments. Consequently, age data were not included directly in the 2011 model, nor will that be the case in the new assessment. Pacific Spiny Dogfish are sporadically targeted, but are more often a bycatch species in the commercial trawl fishery, with the vast majority of retained catch being exported, mostly to Asian markets.

Assessments for these stocks will provide the basis for the management of the groundfish fisheries off the West Coast of the U.S., including providing scientific basis for setting OFLs and ABCs as mandated by the Magnuson-Stevens Act. The technical review will take place during a formal, public, multiple-day virtual meeting of fishery stock assessment experts. Participation of an external, independent reviewer is an essential part of the review process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements:
Two CIE reviewers will participate in the stock assessment review panel. One CIE reviewer, requested herein, shall conduct an impartial and independent peer review of the assessments described above and in accordance with the Performance Work Statement (PWS) and ToRs herein. Additionally, one “common” CIE reviewer will participate in all STAR panels held in 2021 and the PWS and ToRs for the “common” CIE reviewer are included in **Attachment A**.

The CIE reviewers shall be active and engaged participants throughout panel discussions and able to voice concerns, suggestions, and improvements, while respectfully interacting with other review panel members, advisors, and stock assessment technical teams. The CIE reviewers shall have excellent communication skills in addition to working knowledge and recent experience in fish population dynamics; with experience in the integrated-analysis modeling approach, using age- and size- (and possibly spatially-) structured models, and methods for quantifying uncertainty. Familiarity with environmental, ecosystem and climatic effects on population dynamics and distribution may also be beneficial. The CIE reviewer’s duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Tasks for Reviewers:
The CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

<u>Prior to the Peer Review</u>: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the NMFS Contracting Officer Representative (COR), who forwards this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the PWS and ToRs to the CIE reviewer. The NMFS Project Contact is responsible for providing the CIE reviewer with the background documents, reports, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the PWS in advance of the panel review meeting. Any changes to the PWS or ToRs must be made through the COR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance with the PWS scheduled deadlines specified herein. The CIE reviewer shall read all documents in preparation for the peer review.

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- The current draft stock assessment reports;
- Previous stock assessments and STAR Panel reports for the assessments to be reviewed;
- The Pacific Fishery Management Council's Scientific and Statistical Committee's Terms of Reference for Stock Assessments and STAR Panel Reviews;
- Stock Synthesis (SS) Documentation;
- Additional supporting documents as available;
- An electronic copy of the data, the parameters, and the model used for the assessments (if requested by reviewer).

Test: Additionally, two weeks prior to the peer review, the CIE reviewers will participate in a test to confirm that they have the necessary technical specifications provided in advance of the panel review meeting.

Panel Review Meeting: The CIE reviewer shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the review panel's virtual meeting, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., video or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: The CIE reviewer shall complete an independent peer review report in accordance with the PWS. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. The CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Tasks – Contribution to Summary Report: The CIE reviewer should assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. The Chair is not provided by the CIE under this contract. A CIE reviewer is not required to reach a consensus with other members of the Panel, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Place of Performance:
The CIE reviewers shall conduct an independent peer review during the panel review meeting scheduled for the dates of May 3-7, 2021. Due to current uncertainties in the state of the COVID-19 pandemic at that time, this meeting will be conducted as a virtual meeting, with technical assistance provided by staff from the Pacific Fishery Management Council.

Period of Performance:
The period of performance shall be from the time of award through **July 2021**. The CIE reviewers' duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables:
CIE shall complete the tasks and deliverables described in this PWS in accordance with the following schedule.

Schedule	**Milestones and Deliverables**
Within two weeks of the award	Contractor selects and confirms reviewers. This information is sent to the COR, who then transmits this to the NMFS Project Contact
Approximately two weeks later	Contractor provides the pre-review documents to the CIE reviewers
May 3-7, 2021	**Virtual Panel Review Meeting**
Approximately two weeks later	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final CIE independent peer review reports to the COR

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel:
No travel is necessary, as this meeting is being held remotely.

Restricted or Limited Use of Data:
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Andi Stephens, NMFS Project Contact
National Marine Fisheries Service,
Newport, OR 97365
Andi.Stephens@noaa.gov
Phone: 843-709-9094

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Performance Work Statement
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Stock Assessment Review (STAR) Panel 1

The specific responsibilities of the STAR panel are to:

1. Become familiar with the draft stock assessment documents, data inputs, and analytical models along with other pertinent information (e.g., previous assessments and STAR panel report when available) prior to review panel meeting.
2. Discuss the technical merits and deficiencies of the input data and analytical methods during the open review panel meeting.
3. Evaluate model assumptions, estimates, and major sources of uncertainty.
4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
6. When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches and technical issues, differentiating between the short-term and longer-term time frame.
7. Provide a brief description on panel review proceedings highlighting pertinent discussions, issues, effectiveness, and recommendations.

Annex 3: Tentative Agenda

Final Agenda to be provided two weeks prior to the meeting with draft assessments and background materials.

Stock Assessment Review (STAR) Panel 1 - Virtual

May 3-7, 2021

Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Participants Dover Sole

STAR Panel Members

Tien-Shui Tsou, Washington Department of Fish and Wildlife (Chair)
Fabio Caltabellotta, Oregon State University
Matt Cieri, Center for Independent Experts
Noel Cadigan, Center for Independent Experts

Stock Assessment Team (STAT) Members

Chantel Wetzel, National Marine Fisheries Service Northwest Fisheries Science Center
Aaron Berger, National Marine Fisheries Service Northwest Fisheries Science Center

STAR Panel Advisors

Whitney Roberts, Washington Department of Fish and Wildlife, Groundfish Management Team representative
Gerry Richter, B&G Seafoods, Groundfish Advisory Subpanel representative
John DeVore, Pacific Fishery Management Council representative

Participants Spiny Dogfish

STAR Panel Members

Tien-Shui Tsou, Washington Department of Fish and Wildlife (Chair)
Fabio Caltabellotta, Oregon State University
Matt Cieri, Center for Independent Experts
Noel Cadigan, Center for Independent Experts

Stock Assessment Team (STAT) Members

Vladlena Gertseva, National Marine Fisheries Service Northwest Fisheries Science Center
Ian Taylor, National Marine Fisheries Service Northwest Fisheries Science Center
John Wallace, National Marine Fisheries Service Northwest Fisheries Science Center

Sean E. Matson, National Marine Fisheries Service West Coast Region

STAR Panel Advisors

Whitney Roberts, Washington Department of Fish and Wildlife, Groundfish Management Team representative
Gerry Richter, B&G Seafoods, Groundfish Advisory Subpanel representative
John DeVore, Pacific Fishery Management Council representative

Agenda

PROPOSED AGENDA

Stock Assessment Review (STAR) of
Dover Sole and Spiny Dogfish
Pacific Fishery Management Council
Via Webinar
All Times are Pacific Daylight Time and Subject to Change During the Course of the Meeting at the Discretion of the STAR Panel Chair
May 3-7, 2021

Monday, May 3, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues

(8:30 a.m.)

Welcome and Introductions

1. Roll Call and Introductions Theresa Tsou, Chair
2. Review Terms of Reference Theresa Tsou
3. Review and Approve Agenda
4. Review Virtual Format Operational Guidelines John DeVore
5. Assign Writing Duties Theresa Tsou

(8:45 a.m.)

Overview of the Spiny Dogfish Assessment

(9:30 a.m.)

1. Biology, Fisheries, Data, and Inputs Used Vlada Gertseva

BREAK (10:00 – 10:15 a.m.)

2. Assessment Modeling, Performance, and Current Status Vlada Gertseva
3. STAR Panel Requests to the Stock Assessment Team (STAT-1)

LUNCH BREAK (12:30 – 1:30 p.m.)

Overview of the Dover Sole Assessment

(1:30 p.m.)

1. Biology, Fisheries, Data, and Inputs Used Chantel Wetzel & Aaron Berger
2. Assessment Modeling, Performance, and Current Status Chantel Wetzel & Aaron Berger

2
BREAK (3:00 – 3:15 p.m.)
3. STAR Panel Requests to the Stock Assessment Team (STAT-2)

Public Comments
(3:30 p.m.)
STAR Panel Discussion/Planning (as needed)
(4:00 p.m.)
Adjourn for the Day
(4:30 p.m.)

Tuesday, May 4, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues
(8:30 a.m.)
Responses to Panel Requests for Spiny Dogfish
(8:45 a.m.)
1. Presentation of Modeling Results Vlada Gertseva
2. Further Discussion of Modeling Results

BREAK (10:00 – 10:15 a.m.)
3. Additional STAR Panel Requests to STAT-1

LUNCH BREAK (11:30 a.m. – 1:00 p.m.)
Responses to Panel Requests for Dover Sole
(1:00 p.m.)
1. Presentation of Modeling Results Chantel Wetzel & Aaron Berger
2. Further Discussion of Modeling Results

BREAK (2:15 – 2:30 p.m.)
3. Additional STAR Panel Requests to STAT-2

Public Comments
(3:30 p.m.)

3

STAR Panel Discussion/Planning (as needed)

(4:00 p.m.)

Adjourn for the Day

(4:30 p.m.)

Wednesday, May 5, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues

(8:30 a.m.)

Responses to Panel Requests for Spiny Dogfish

(8:45 a.m.)

1. Presentation of Modeling Results Vlada Gertseva
2. Further Discussion of Modeling Results

BREAK (10:00 – 10:15 a.m.)

3. Additional STAR Panel Requests to STAT-1

LUNCH BREAK (11:30 a.m. – 1:00 p.m.)

Responses to Panel Requests for Dover Sole

(1:00 p.m.)

1. Presentation of Modeling Results Chantel Wetzel & Aaron Berger
2. Further Discussion of Modeling Results

BREAK (2:15 – 2:30 p.m.)

3. Additional STAR Panel Requests to STAT-2

Public Comments

(3:30 p.m.)

STAR Panel Discussion/Planning (as needed)

(4:00 p.m.)

Adjourn for the Day

(4:30 p.m.)

4

Thursday, May 6, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues

(8:30 a.m.)

Responses to Panel Requests for Spiny Dogfish

(8:45 a.m.)

1. Presentation of Modeling Results Vlada Gertseva
2. Further Discussion of Modeling Results

BREAK (10:00 – 10:15 a.m.)

3. Agreement of a Preferred Model Between the STAR Panel and STAT-1
4. STAR Panel Requests for Model Runs for the Decision Table

LUNCH BREAK (11:30 a.m. – 1:00 p.m.)

Responses to Panel Requests for Dover Sole

(1:00 p.m.)

1. Presentation of Modeling Results Chantel Wetzel & Aaron Berger
2. Further Discussion of Modeling Results

BREAK (2:15 – 2:30 p.m.)

3. Agreement of a Preferred Model Between the STAR Panel and STAT-2
4. STAR Panel Requests for Model Runs for the Decision Table

Public Comments

(3:30 p.m.)

STAR Panel Discussion/Planning (as needed)

(4:00 p.m.)

Adjourn for the Day

(4:30 p.m.)

Friday, May 7, 2021 – 8:30 AM

Early Log-In to Resolve Connection Issues

(8:30 a.m.)

Consideration of Remaining Issues

(8:45 a.m.)

1. Discussion of Proposed Base Models
2. Review Decision Tables for All Assessments

BREAK (10:00 – 10:15 a.m.)

3. Review Any Possible Disagreements from GMT, GAP, and PFMC Advisors
4. Identify Research and Data Needs

Public Comments

(11:00 a.m.)

LUNCH BREAK (11:30 a.m. – 1:00 p.m.)

Review Draft STAR Panel Report

(1:00 p.m.)

1. Discuss Deadlines for Report Submission
2. Review and Discuss Draft Report

BREAK (2:15 – 2:30 p.m.)

STAR Panel Discussion/Planning (as needed)

(2:30 p.m.)

STAR Panel Adjourns

(4:30 p.m.)