2020 Pacific halibut (*Hippoglossus stenolepis*) stock assessment: Development

PREPARED BY: IPHC SECRETARIAT (I. STEWART, A. HICKS & P. CARPI; 22 MAY 2020)

PURPOSE

To provide the IPHC's Scientific Review Board (SRB) with a response to requests made during SRB015 (IPHC-2019-SRB015-R) and to provide an update of the 2020 assessment development.

INTRODUCTION

The 2019 stock assessment included a complete re-evaluation of all data sources and modelling choices as part of a full stock assessment analysis. A summary of results (IPHC-2020-AM096-09 Rev_2) was presented to the Commission during AM095 (Stewart et al. 2020). Full assessment (Stewart and Hicks 2020) and data overview (Stewart and Webster 2020) documents were posted directly to the stock assessment page of the IPHC's website. The 2019 scientific review comprised both the standard SRB reviews in June (SRB014) and September (SRB015), as well an external peer review (Stokes 2019).

For 2019, there were two primary improvements to the existing data sources:

1) sex-ratios at age based on genetic assays of port sampled Pacific halibut were available for commercial fishery landings made in 2017 and 2018; and
2) a revised modelled index of abundance reflecting the 2019 FISS sampling and expansions (in IPHC Regulatory Areas 3A and 3B).

The stock assessment comprised an ensemble of four equally weighted models: two long time-series models, reconstructing historical dynamics back to the beginning of the modern fishery, and two short time-series models incorporating data only from 1992 to the present, a time-period for which estimates of all sources of mortality and survey indices are available for all regions. The most salient changes to the assessment models included the estimation of male commercial fishery asymptotic selectivity parameters to accommodate the newly available sex-ratio data, re-weighting of sample sizes for age-composition data, and re-tuning time-varying parameter (recruitment, catchability and selectivity) constraints to be consistent with the variability estimated in the models. In aggregate these results produced slightly higher terminal biomass estimates as well as higher estimates of recent fishing mortality; however, model results remained consistent with previous assessments conducted since 2012.

For 2020, the Secretariat intends to conduct an updated stock assessment, including recreational fishery sex-ratio data, additional commercial fishery sex-ratio data from 2019 as well as newly available data collected during 2020, but not making large structural changes to the ensemble or individual assessment models prior to the next full review cycle.

SRB REQUESTS AND RESULTS

The SRB made the following requests and recommendations during SRB015:

SRB015 (para. 33): *"The SRB REQUESTED that for SRB016 (2020), the IPHC Secretariat:*

a) provide a more detailed evaluation and profile of steepness values. Specifically, this should show the different data and model components that inform the steepness parameter, and also the interaction with sigmaR. This should also help inform the SRR relationship to be used in the operating model for MSE work;

b) consider examining the relative impact of different fleets (sources of mortality) on historical SSB (e.g. set fleet x F = 0, replay, then fleet x and y, etc.)."

SRB015–Rec.04 (para. 34): *"NOTING the discussion of recommendations arising from the external peer review of the IPHC stock assessment (Section 4), the SRB RECOMMENDED that the IPHC Secretariat:*

a) Update data weighting for the 2019 assessment;

b) For SRB016:

i. evaluate the types of weightings (e.g., Dirichlet-multinomial) for compositional data;

ii. advise on the impact of data re-weighting as new information arises. This could be more sensitive as new sex-composition data are included;

iii. keep apprised of new software developments (e.g. CAPAM meeting in NZ) and report on potential future directions (e.g. if alternatives provide improved Bayesian integration or adaptations for simulation testing etc.)."

Each of these requests and recommendations are addressed below. In order to provide comparability between these results and all subsequent steps working toward the final 2020 stock assessment, this evaluation began with the final 2019 models. First, each of the four assessment models was extended by one year, including projected 2020 mortality from all sources based on the mortality limits set during AM096. Next, the stock synthesis software was updated to the most recent version available, 3.30.14 (Methot Jr et al. 2019). The changes from the version used for the 2019 stock assessment (3.30.13) were unimportant to the Pacific halibut stock assessment, but maintaining a current version (when possible and efficient) reduces the likelihood of compatibility issues with plotting and other software and reduces the cumulative transitional burden (which was substantial for the 2019 stock assessment) when future changes are added. All requests are therefore based on the extended and updated models, ready for the 2020 stock assessment data as it becomes available.

Request a – steepness evaluation and profile

The steepness parameter (*h*) defines the relative recruitment predicted to occur at 20% of the unfished spawning biomass based on the stock-recruitment curve. If recruitment deviations are estimated about this curve, then as the variance of the recruitment deviations (σ_r) goes to infinity the parameters of the curve are redundant to the predictions and therefore inestimable. This statistical reality, in addition to the low information content of most fisheries data sets on the stock-recruitment relationship due to low contrast in stock size and uncertain information of the absolute scale of recruitments, leads to steepness being difficult or impossible to estimate in many stock assessments (e.g., Lee et al. 2012; Thorson et al. 2018). However, steepness is importantly linked with modelled quantities of management interest, particularly reference points (Mangel et al. 2013).

For the 2019 stock assessment, a value of 0.75 was used for steepness in all four individual stock assessment models, and this assumption was investigated in the preliminary assessment prepared for SRB014 (IPHC-2019-SRB014-07), and further for SRB015 (IPHC-2019-SRB015-07).This evaluation repeats that performed in 2019, extending it to include component-specific likelihood profiles as well as an investigation of maternal effects (Berkeley et al. 2004; Shelton et al. 2015), or an increased egg-output/survival per body mass based on the age of the fish.

Four alternative models were run for each of the four stock assessment models (in addition to the base model where *h* = 0.75) holding steepness constant at 0.65, 0.85, 0.99, and estimating steepness using a uniform prior from 0.2-1.0. When steepness was estimated, the maximum likelihood estimate for steepness was equal to 1.0 for three of the four models, and for the Areas-As-Fleets (AAF) long time-series model it was equal to 0.85 (rounded to the second decimal place). As previously found in earlier analyses, the AAF long model produced very similar time-series results under alternate values for steepness (Figure 1), as did the AAF short time-series model (Figure 2).

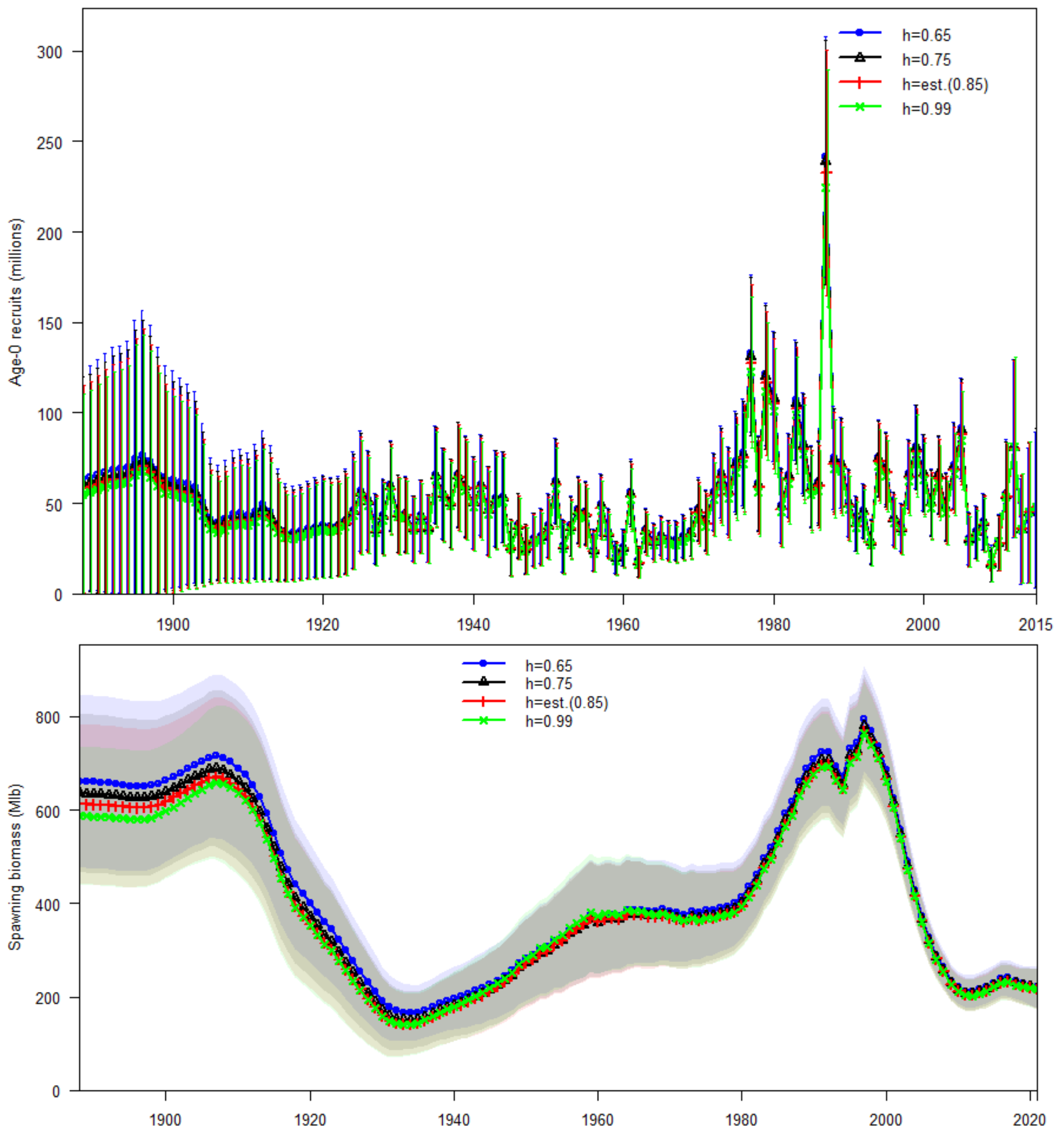

Figure 1. Comparison of AAF long time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) using alternative values for the steepness of the stock-recruitment relationship.

In contrast to previous results, the coastwide short model showed some sensitivity to steepness, with higher values (0.85 and 1.0) corresponding to lower spawning biomass over the time-series and slightly higher recruitment estimates near the end of the time series where the information content of the age data is relatively low (Figure 3). It is possible that the increased sensitivity of this model relative to the preliminary 2019 assessment may be due to the data informing the 2011 and 2012 year classes which are estimated to be larger than predicted by the stock-recruitment relationship. At higher values for steepness the spawning biomass can be slightly lower and still produce these larger year classes. A similar but more pronounced relationship was found for the coastwide long time-series model (Figure 4). In that model, the very low spawning biomass estimated in the 1930s may be creating a 'bottleneck' that causes the scale of the biomass to be dependent on the model's ability to produce recruitments to match the age and mortality information during that time period.

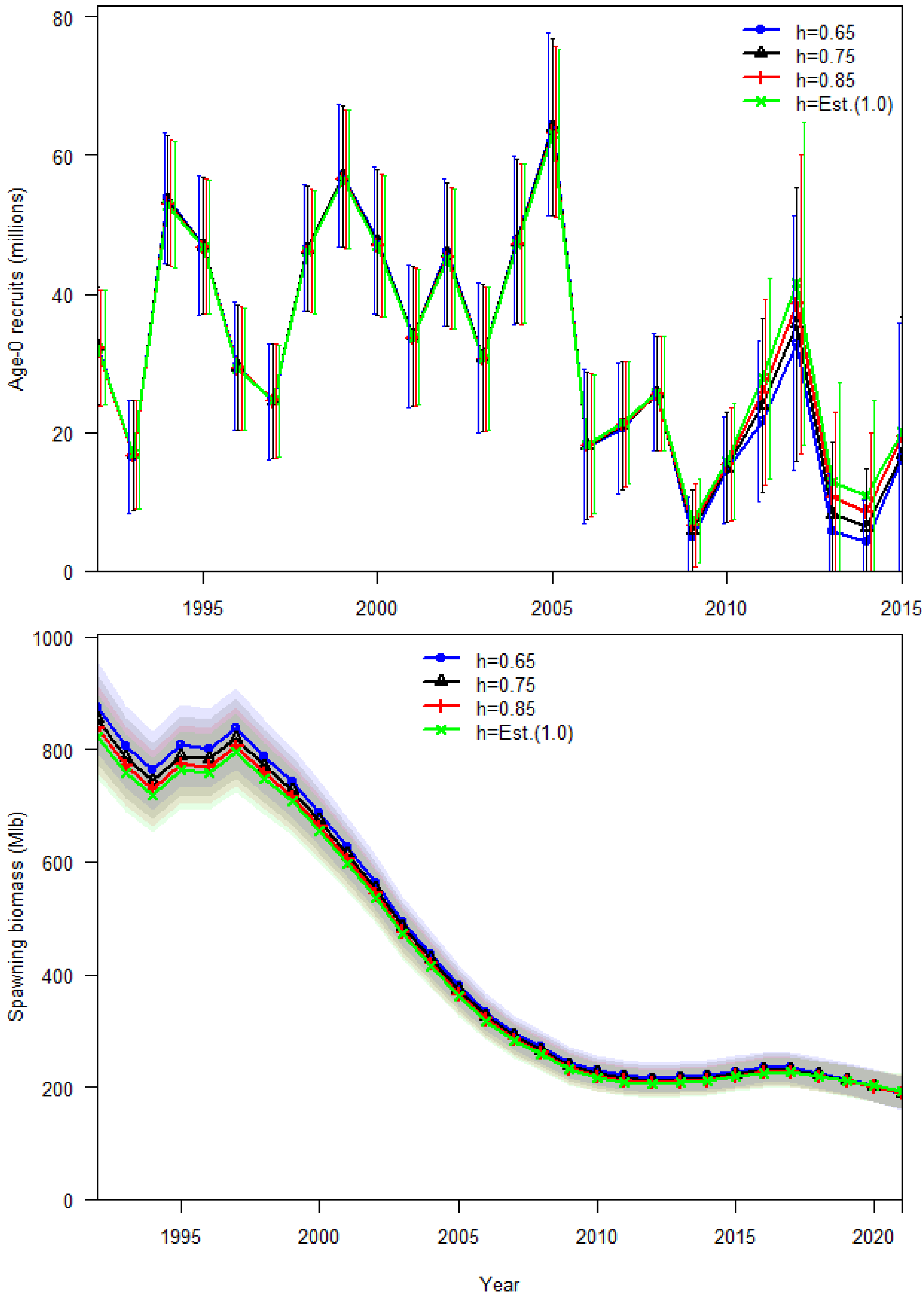

Figure 2. Comparison of AAF short time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) using alternative values for the steepness of the stock-recruitment relationship.

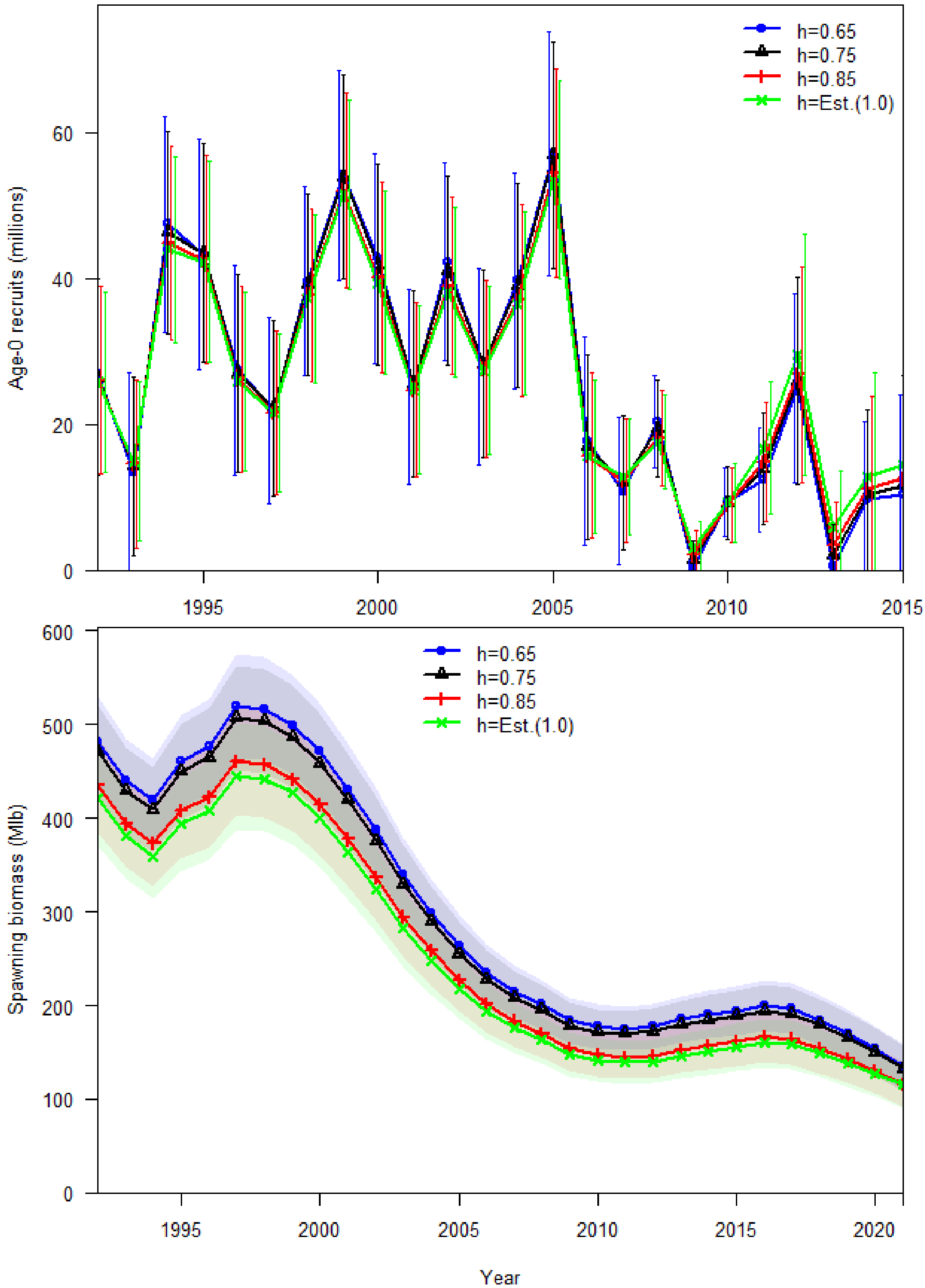

Figure 3. Comparison of coastwide short time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) using alternative values for the steepness of the stock-recruitment relationship.

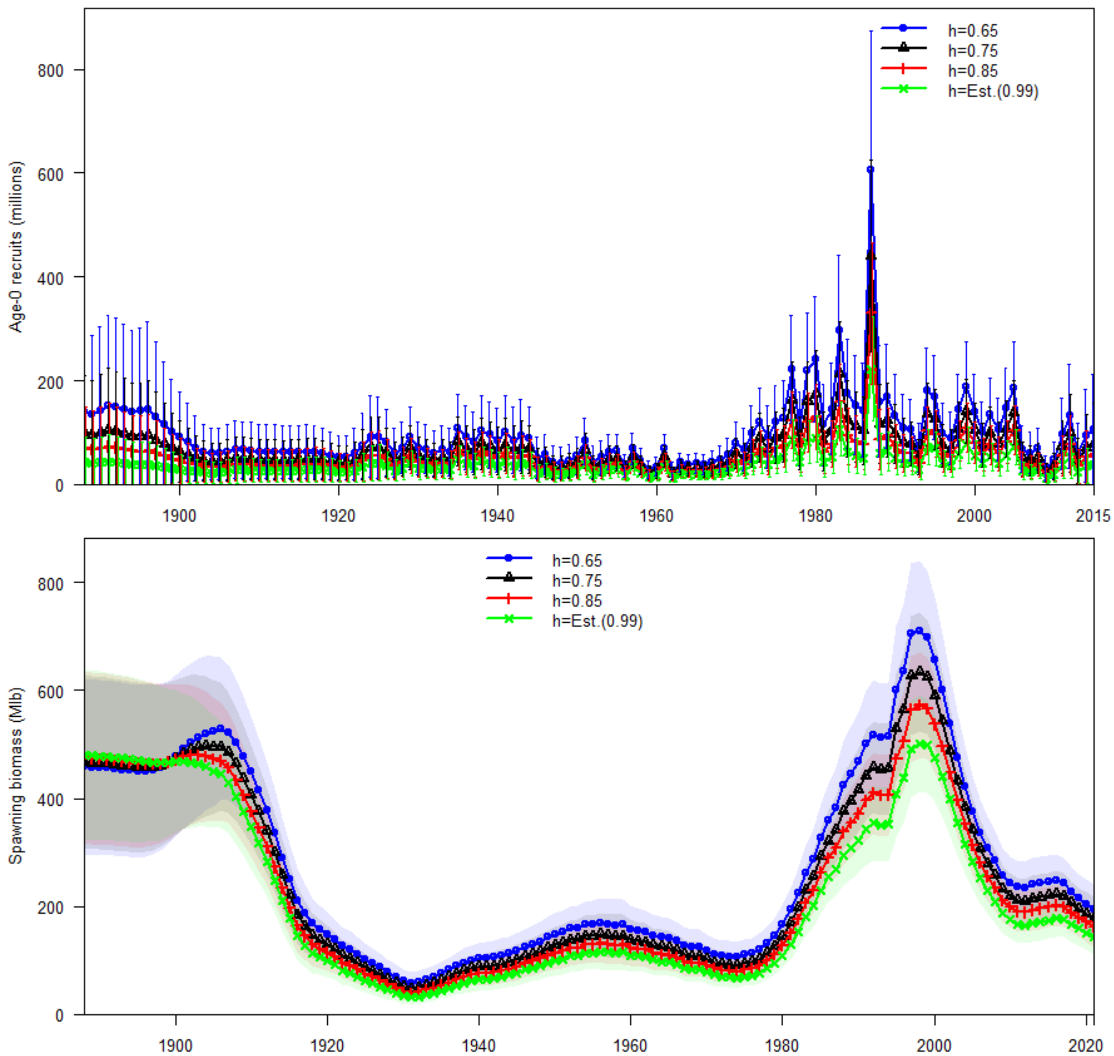

Figure 4. Comparison of coastwide long short time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) using alternative values for the steepness of the stock-recruitment relationship.

In order to better understand which sources of data and/or model constraints were informative regarding steepness, the negative log-likelihood (NLL) values from the alternate models reported above were evaluated (all values are reported as delta NLL as the absolute scale differs among models and data sources). Each model had four sources of information: age composition data, index data, the constraint for deviations around the stock-recruitment relationship (σ_r), and the constraints on other time-varying parameters based on the variances assigned to each. Of particular interest is whether the data provide information beyond the model constraints. It is important to consider that likelihood profiles cannot be interpreted in a strict statistical sense for fisheries models as the error distributions (in this case sample sizes for age composition data) have been iteratively reweighted and so the scale of the profiles depends on the weighting.

The coastwide short time-series model had the greatest difference in NLL values over steepness ranging from 0.65 to 1.0 (Figure 5), with the information coming from the age-composition data and σ_r in similar proportions. All of the fleet-level components within the age data showed similar but weak information favoring higher values of steepness. Time-varying parameter deviation variances were not iteratively re-tuned for each level of steepness, therefore some of the change in NLL from these components may be meaningless. When the input value for σ_r was reduced to account for the slightly improved fit of the stock-recruitment relationship at a steepness of 1.0, the time-series of recruitment and spawning biomass were largely unchanged (Figure 6). In contrast, the only appreciable information for steepness in the coastwide long time-series model was from σ_r (Figure 7) and when σ_r was retuned (from a value of 0.54 to 0.51) there was little change in the estimated time-series (Figure 8).

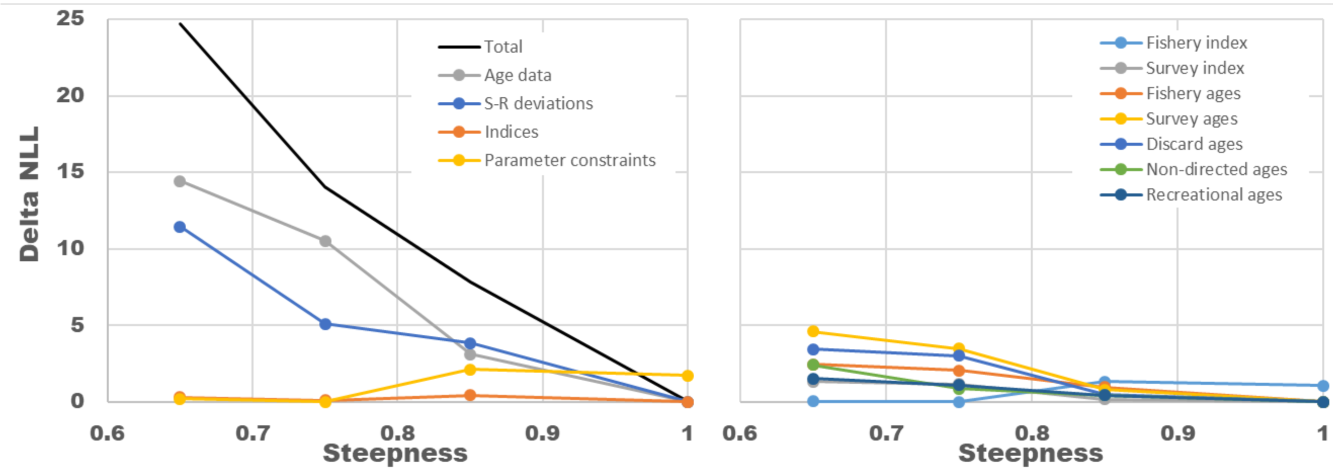

Figure 5. Comparison of coastwide short time-series likelihood profiles by type of data (left panel) and source within type for index and age data (right panel).

The AAF short time-series model had less information on steepness overall than the short coastwide model, but the age composition data and σ_r still represented the majority of the change in NLL (Figure 9). The AAF long time-series model had essentially no information favoring steepness values between 0.65 and 1.0, despite a maximum likelihood estimate of 0.85 (Figure 10; tested with alternative starting values due to the apparently flat likelihood surface). In aggregate, these likelihood profiles illustrate why steepness has not been estimated in these models.

With regard to management quantities the use of steepness equal to 0.75 represents an intermediate value, not clearly risk-prone or averse. Although the spawning biomass estimates are estimated to be slightly lower at a steepness of 1.0 for two of the four assessment models comprising the ensemble, there is a counter effect on the estimate of relative spawning biomass via the feedback through the calculation of the unfished biomass. At higher steepness values, the effect of fishing is reduced (a value of h=1.0 results in no recruitment effect on the unexploited biomass). An illustration of this is that the probability that the spawning biomass was below $SB_{30\%}$ at the beginning of 2020 goes from an estimate of 46% using a steepness of 0.75, down to 37% using a steepness of 1.0.

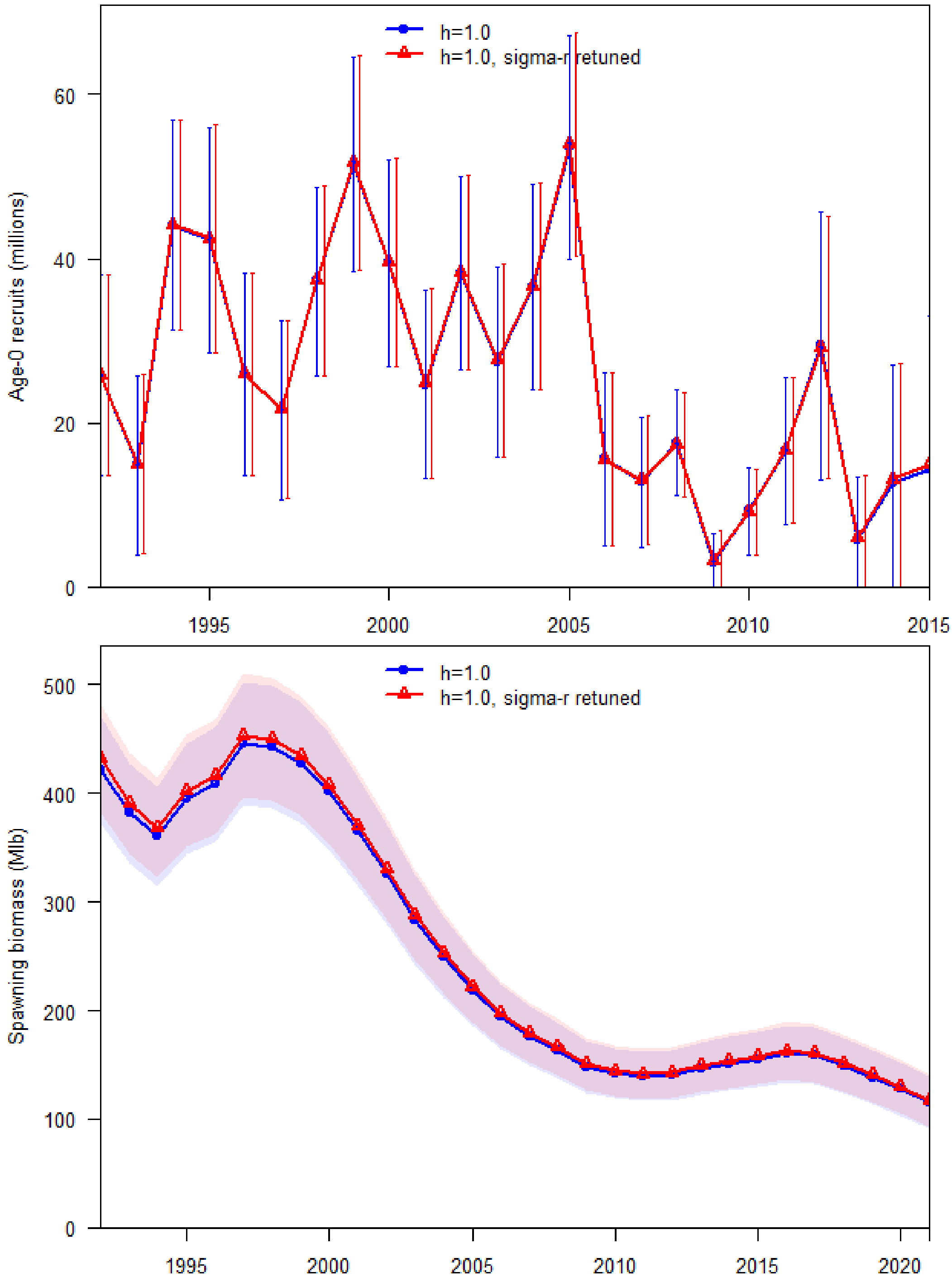

Figure 6. Comparison of coastwide short time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) with steepness equal to 1.0 and retuning the value for σ_r.

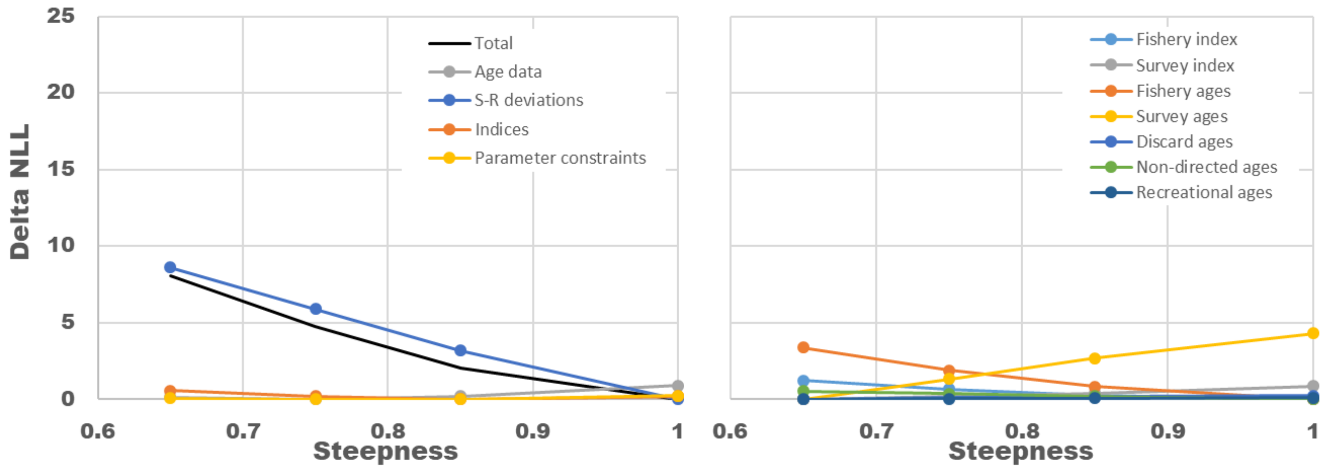

Figure 7. Comparison of coastwide long time-series likelihood profiles by type of data (left panel) and source within type for index and age data (right panel). Note that the y-axis of this figure has been scaled to be identical to Figure 5 for comparability.

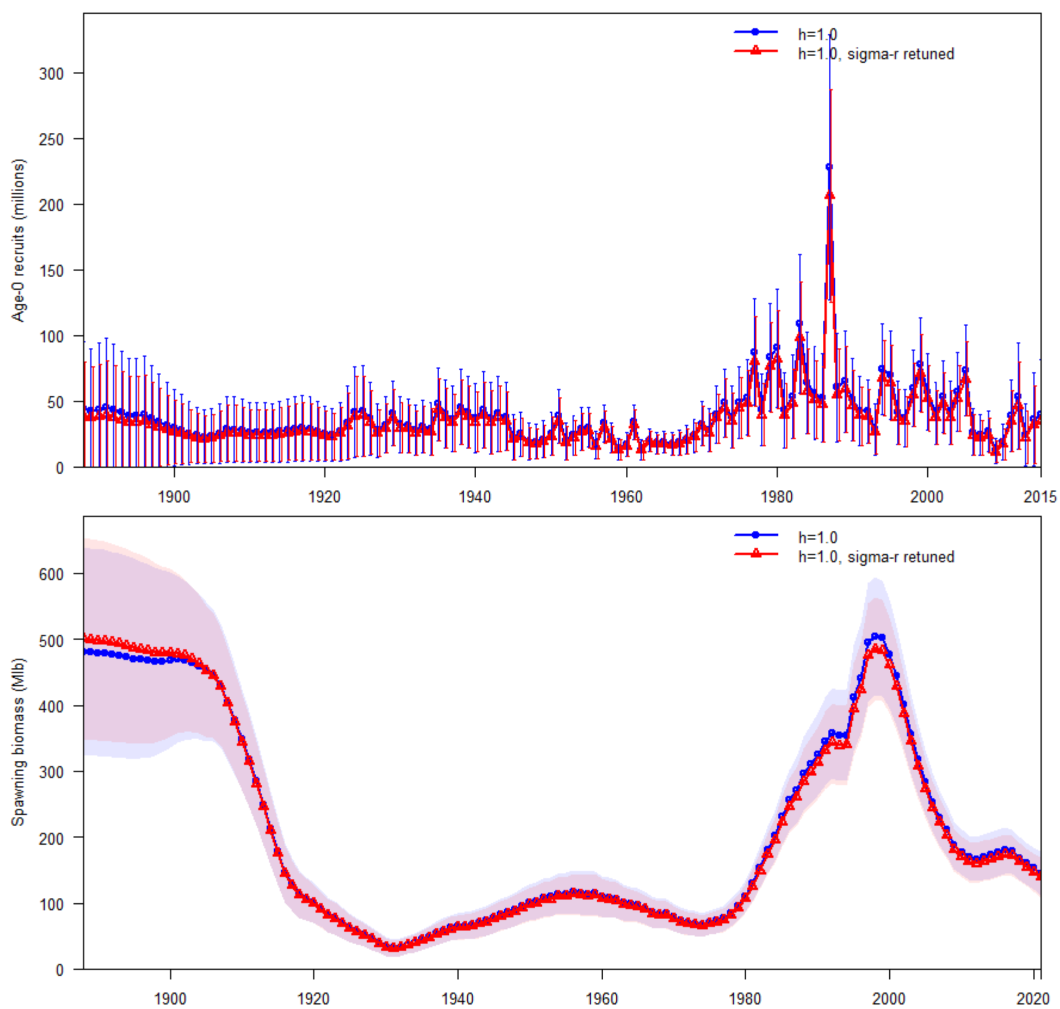

Figure 8. Comparison of coastwide long time-series model results for recruitment (upper panel) and female spawning biomass (lower panel) with steepness equal to 1.0 and retuning the value for σ_r.

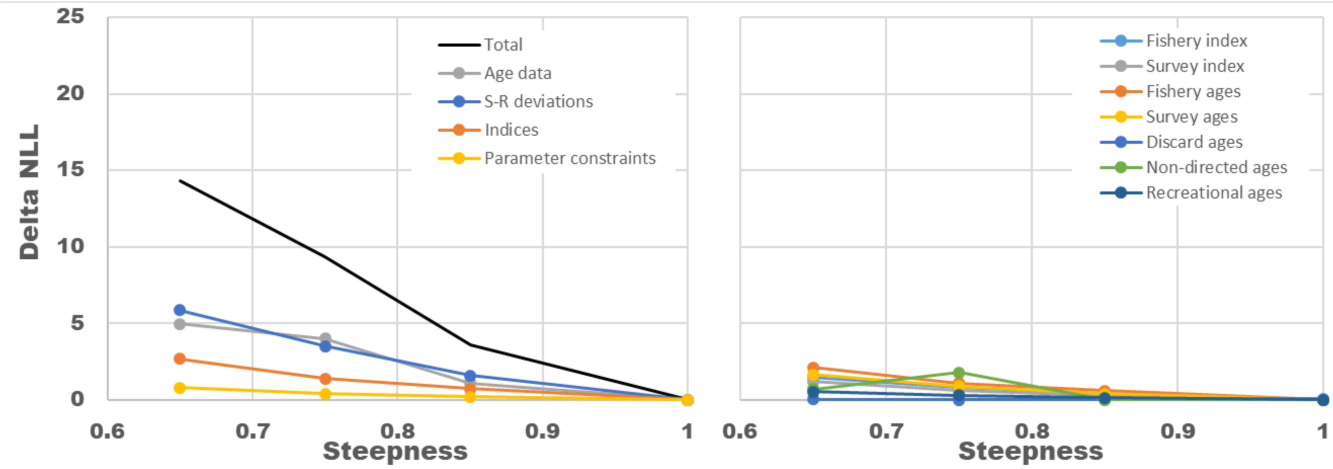

Figure 9. Comparison of AAF short time-series likelihood profiles by type of data (left panel) and source within type for index and age data (right panel). Note that the y-axis of this figure has been scaled to be identical to Figure 5 for comparability. Individual model fleets representing Biological Regions have been aggregated to the same level as the coastwide model results.

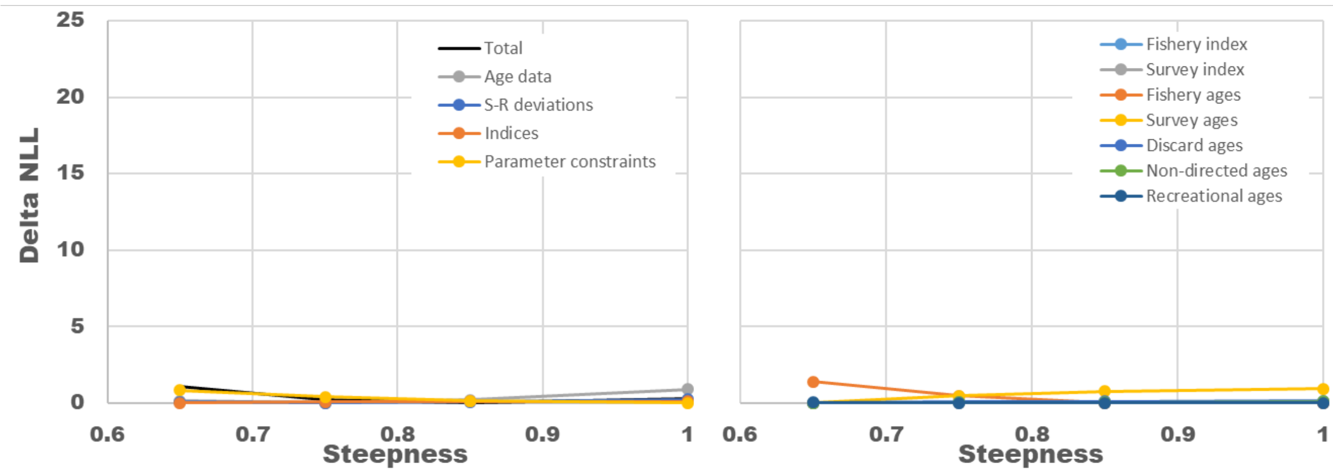

Figure 10. Comparison of AAF long time-series likelihood profiles by type of data (left panel) and source within type for index and age data (right panel). Note that the y-axis of this figure has been scaled to be identical to Figure 5 for comparability. Individual model fleets representing Biological Regions have been aggregated to the same level as the coastwide model results.

An additional consideration relevant to the evaluation of steepness and the stock-recruitment relationship in general is the potential for maternal effects. Maternal effects have been identified in several fish species, and are represented by increased egg-output/survival per body mass based on the age (and/or size) of the fish. There are currently no data that support a fecundity or survival that increases with increasing body mass for Pacific halibut; however, it is a topic of frequent questions and represents an important aspect of the current IPHC research program along with basic maturity characterization. The effect of an unidentified maternal effects relationship would be most likely to manifest in increased σ_r, as the stock-recruitment relationship should appear weaker (and thus have additional variability) if the x-axis (spawning biomass) was systematically incorrect due to maternal effects.

In order to investigate whether the estimated time-series' for Pacific halibut might be better explained in the presence of maternal effects, two alternative configurations were run for each of the four individual stock assessment models. Specifically, maturity and weight-at-age were maintained at estimated time-series values, while allowing egg output per body weight to increase from a value of 1.0 at age-0 (no fish less than age 8 are mature) to a value of 1.25 (25% higher) or 1.50 (50% higher) at age-30. These alternative reproductive output arrays were then applied in each stock assessment model, and the resulting σ_r values were compared.

All models produced levels of recruitment variability that were either equal to or higher than estimated in the absence of maternal effects. This result is consistent with other studies showing that the effects of maternal effects on population demographics may be relatively small (Andersen et al. 2019). As expected, the absolute value for 'spawning biomass' is larger for these alternative model runs, as it is now egg output and not directly comparable to previous biomass estimates. The relative spawning biomass estimates did not change appreciably under the 25% maternal effects alternative and were only slightly lower (~2%) for the ensemble results under the 50% maternal effects alternative. Although this is a very weak 'test', there appears to be no evidence in the current data that the addition of a simple age-based maternal effects relationship improves the ability of the current stock assessment models to explain the time-series of estimated recruitments.

Request b – fleet impacts

The Secretariat has frequently been asked to evaluate alternative historical scenarios in which different management procedures or decisions were made and the likely cumulative effects of those decisions over time. On a technical level, such revised histories are relatively easy to produce; however, their clear interpretation may be extremely challenging as subsequent management decisions necessarily depend on all previous decisions, the status of the resource and other information available at the time. This topic was discussed at length during SRB012, with the report (IPHC-2018-SRB12-R) ultimately including:

> SRB012 (para. 23): *"NOTING the request for "replay" analyses, the SRB AGREED that "what if" questions about past behaviour are not appropriate for stock assessment models because those analyses do not adequately reflect the information available at the time or information feedbacks to future decision over time. An MSE analysis, on the other hand is specifically designed to answer "what if" questions under particular future scenarios while properly accounting for stock assessment errors in response to changing information."*

Regardless of these caveats it may be informative to compare the aggregate effect of 'impact' of each source of mortality on the Pacific halibut spawning biomass over the estimated time-series' in order to better understand the relative effects. Although other stock assessments have likely undertaken similar exercises, the steps used were similar to those reported by Wang et al. (2009). These consisted of the following:

1) Begin with the estimated time-series of spawning biomass (*SB*) and the estimated time-series of spawning biomass in the absence of fishing mortality (dynamic SB_0) from the stock assessment ensemble, representing the integrated results from each individual model.

2) Re-calculate (but do not re-estimate) each of the four individual models by fixing all parameter values at their maximum likelihood estimates, removing each source of mortality sequentially. In these models mortality is divided into 5 sources: directed commercial fishing, directed commercial discard mortality (due to the minimum size limit, regulation and lost fishing gear), non-directed commercial fishing (all Pacific halibut are required to be discarded), recreational mortality, and subsistence mortality.
3) Integrate the four individual model results for the ensemble.
4) Compare the relative change in *SB* with each source of mortality removed.

Because this analysis is based on calculated and not estimated time-series, there are no variance estimates and therefore no consistent method for integrating the results for the ensemble of models. For simplicity, a raw average of the four model *SB* time-series was used. This is similar to the weighting used in the ensemble but ignores the difference in variance within and among the time-series, which leads to some differences with the integrated ensemble SB and SB0 time-series', particularly at the beginning of the time-series.

Wang et al. take a further step and use the relative change in the *SB* series to proportionally assign the estimated difference between SB_0 and *SB* in each year to a specific fleet. That approach does not recognize that the mortality from each of these sources of mortality is simultaneous, and therefore not proportional at differing levels of overall fishing intensity, nor a simple summation across fleets. Such a deconstruction of relative simultaneous effects requires a much more complex calculation sometimes referred to as the '*fisheries footprint'* (Martell et al. 2016).

The results of this analysis are very consistent with the relative cumulative mortality from each source over the estimated time-series. The largest component of mortality has been the directed commercial fishery, comprising approximately two-thirds of the total in recent decades and showing the largest 'impact' in this analysis (Figure 11). Non-directed discard mortality, followed by recreational mortality comprise the next largest impacts on the *SB* time-series, with directed discards and subsistence having negligible effects. The simple averaging of individual models, instead of including each variance estimate in the integration leads to some spurious effects at the beginning of the time series, and for subsistence and commercial discards, which appear to be slightly below the estimated time series in some years.

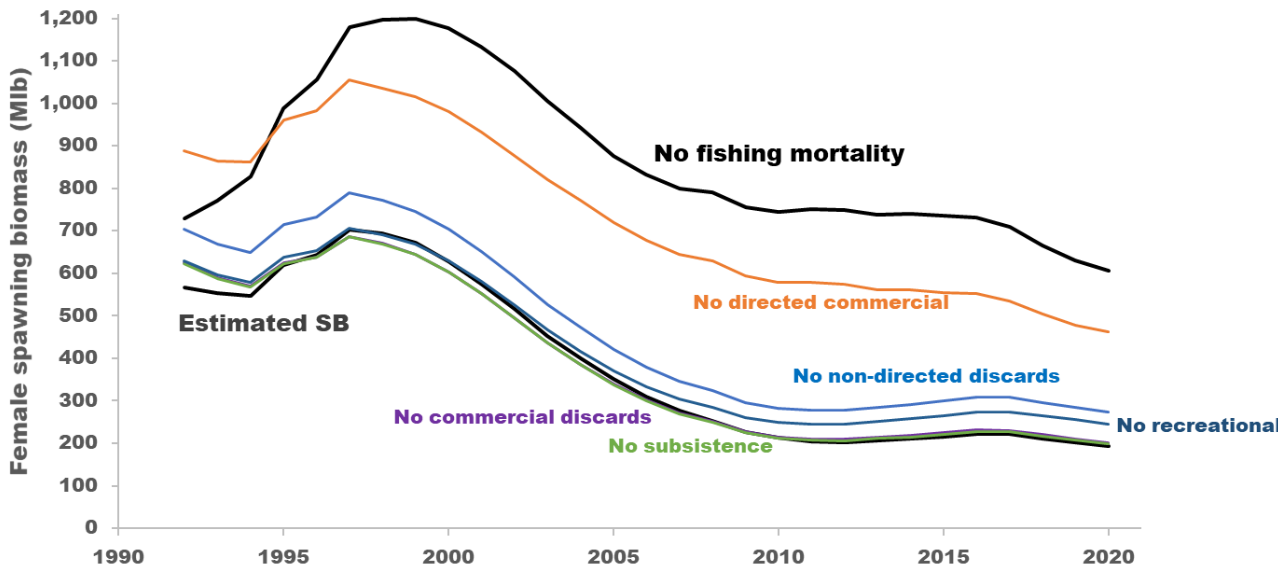

Figure 11. Time series of female spawning biomass (millions pounds) estimated from the 2019 stock assessment (lower black line), and estimated in the absence of any fishing mortality (*SB_0*; upper black line). Colored series represent 'impact analyses' sequentially removing the fishing mortality associated with each source of fishing mortality: directed commercial (orange), non-directed discards (blue), recreational (grey), commercial discards (purple) and subsistence (green).

Recommendation 4, part a – update data weighting

Data weighting was updated for the final 2019 stock assessment. There were relatively small changes to all components except the coastwide survey age data (for coastwide models) and the Biological Region 3 survey age data (for AAF models). For these fleets, the survey age data were down-weighted substantially in the final assessment relative to the preliminary assessment (columns 1-2, Table 1). This represented a much larger change, and was not a result of the additional year of commercial sex-ratio information as anticipated. The IPHC's Fishery Independent Setline Survey (FISS) completed the 6th and final year of planned expansions into previously unsampled areas in 2019, with Biological Region 3 fully sampled (155 additional stations) for the first time (IPHC-2020-AM096-06). This new information for historically unsampled stations had a very large effect on the variance of the survey time-series for Region 3 and coastwide (Figure 12; IPHC-2020-AM096-07). The result of the stock assessment models fitting the survey information better was a reduction in fit to the survey age data and therefore a reduction in the iteratively weighted sample sizes.

Table 1. Comparison of data weighting implied by the Francis method (iterated input sample sizes) for age composition data from the preliminary and final 2019 assessment (larger changes indicated by bold text) and a two-year retrospective analysis of the final 2019 assessment.

	Preliminary 2019	Final 2019	Data through 2018	Data through 2017
Coastwide short				
Fishery	37	38	36	39
Discards[1]	9	9	9	9
Bycatch[1]	5	5	5	5
Sport[1]	5	5	5	5
Survey	372	**263**	265	275
Coastwide long				
Fishery	140	136	136	149
Discards[1]	6	6	6	6
Bycatch[1]	2.5	2.5	2.5	2.5
Sport[1]	2.5	2.5	2.5	2.5
Survey	125	**65**	65	72
AAF short				
Region 2 Fishery[2]	545	538	545	553
Region 3 Fishery[2]	281	278	282	286
Region 4 Fishery	29	26	27	32
Region 4B Fishery[2]	23	22	23	23
Discards[1]	6	6	6	6
Bycatch[1]	5	5	5	5
Sport[1]	5	5	5	5
Region 2 Survey	9	7	8	8
Region 3 Survey	221	**22**	27	37
Region 4 Survey	72	88	84	79
Region 4B Survey	31	42	39	35
AAF long				
Region 2 Fishery[2]	270	271	270	269
Region 3 Fishery[2]	167	167	167	167
Region 4 Fishery	30	30	29	32
Region 4B Fishery[2]	22	22	22	23
Discards[1]	6	6	6	6
Bycatch[1]	2.5	2.5	2.5	2.5
Sport[1]	5	5	5	5
Region 2 Survey	9	8	8	8
Region 3 Survey	43	**15**	16	18
Region 4 Survey	82	97	96	89
Region 4B Survey	40	54	50	44

[1]Inputs downweighted, and not iteratively reweighted (Stewart and Hicks 2019b).

[2]Sample size equal to maximum (input based on number of samples).

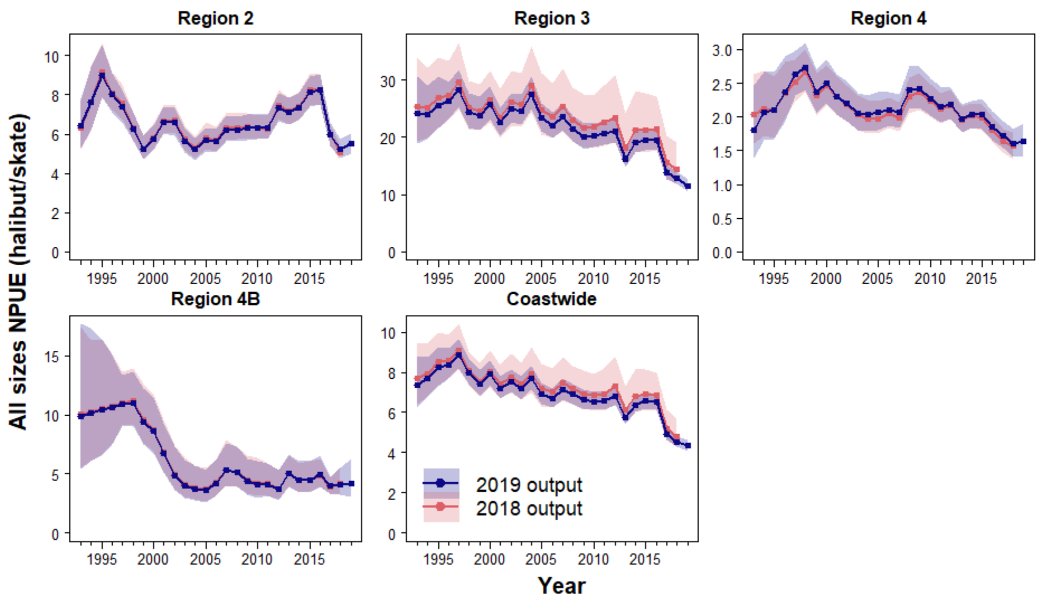

Figure 12. Time series of modelled survey results for all sizes NPUE (halibut per skate) showing the change from the 2018 to the post-expansion results in 2019. Shaded regions indicate approximate 95% credible intervals. Figure from Webster (2020).

Recommendation 4, part b, section i – evaluate types of data weighting

Data weighting represents a challenging but necessary step in fisheries stock assessment models in order to create internal consistency between input and output error distribution assumptions and results as well as address conflicts among data sources. A CAPAM (Center for the Advancement of Population Assessment Methodology) workshop on data weighting was attended by IPHC Secretariat staff 19-23 October 2015 (See full special issue of Fisheries Research; Maunder et al. 2017). Although a wide range of analyses and approaches were presented and discussed, no clear consensus on a single approach for weighting compositional data was reached. Many methods remain in common use, including nominal sample sizes based on fish, samples or trips, the harmonic mean (McAllister and Ianelli 1997), the average age (Francis 2011; Francis 2017), and others, including the Dirichlet-multinomial (Thorson et al. 2017; Xu et al. 2020).

In the 2019 Pacific halibut stock assessment the initial input sample sizes were derived from the number of survey sets and fishery trips (and not the number of individual fish measured, which would be much larger). These nominal input values, as in most assessments that are relatively 'data-rich' were considerably larger than commonly applied weighting methods for stock assessment models would suggest. These values were iteratively reduced based on evaluation of three considerations: the relative magnitude of the standardized residuals, comparison of the input value for each fleet with the harmonic mean effective sample size which is an unbiased estimator for a set of independent multinomial samples (Stewart and Hamel 2014), and the scaling suggested by the Francis (2011) method (as implemented in the r4ss package). After initial process error tuning (selectivity deviations), input sample sizes were adjusted downward until the weights suggested by the fit to the mean age over the time series were approximately equivalent to the input values (the "Francis method'; Francis 2011). For almost all fleets and all models, this approach led to a substantial reduction from initial sample sizes. In no cases were the input values allowed to increase from those derived from the number of trips or stations represented in the data.

Although the approach taken for the Pacific halibut stock assessment is consistent with the basis for several of the commonly applied methods, one potential alternative to an iterative approach to data weighting is the Dirichlet-multinomial (DM; Thorson et al. 2017) The DM estimates one additional parameter (θ), that serves as a scalar on the input sample size and can be estimated simultaneously with other model parameters. The use of the DM has several appealing properties, including propagation of variance associated with the data-weighting and elimination of the need for iteration/updating of data weights along with process error and other components needed to maintain internal model consistency (Thorson 2019). However, the DM also has some undesirable properties including that when nominal input sample sizes do not require reduction then the maximum likelihood estimate for θ will naturally lie at whatever upper bound has been specified, thereby requiring that this parameter is fixed or allowed to remain on the bound risking reduced convergence properties for AD Model Builder (Fournier et al. 2012) based implementations. The DM does not function in the same manner as a simple iteratively-tuned multiplier on the input sample size. Specifically, the degree of reduction to the input sample size is nonlinear and depends on the sample size through:

$$effective\ multinomial\ sample\ size = \left(\frac{1}{1+\theta}\right) + n\left(\frac{\theta}{1+\theta}\right)$$

Where *n* is the input sample size, and θ is the transform of the estimated parameter ln($\tilde{\theta}$). This leads to the counter-intuitive result that when θ is small, the effective weight per nominal sample size goes up – more weight is being given to the worst samples in a time-series relative to the best. For example for θ = 0.05, a sample size of 50 receives almost 50% more weight per sample than a sample size of 500-1000 (Figure 13). This also means that the use of the DM is not scale-independent: manually adjusting the input sample sizes and then introducing them to the assessment model will result in differing weights than applying the DM to the original values, irrespective of whether the proportionality is maintained.

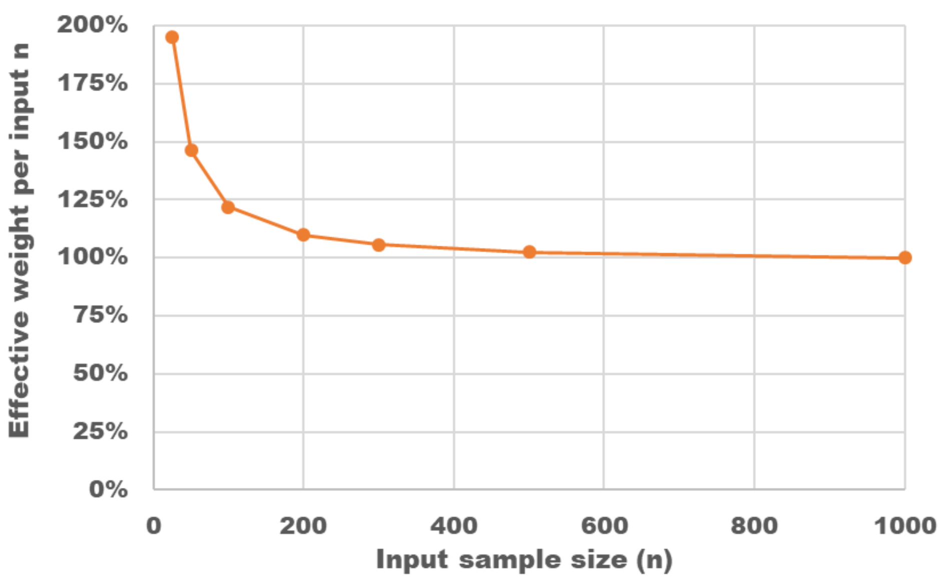

Figure 13. Illustration of the relative weighting per input sample size for a Dirichlet-multinomial heavily down-weighting from the nominal sample sizes (θ of 0.05).

In order to better evaluate the DM for use in Pacific halibut models, the coastwide short time-series model (which is the simplest of the four and has the shortest time for convergence) was used as a test case. Specifically, model-stability, effect on point estimates and effect on uncertainty for both time-series and management-related quantities was evaluated. Three test cases were explored: allowing the weighting of the commercial fishery age data to be estimated via the DM, allowing the survey age data weighting to be estimated, and allowing both weightings to be estimated simultaneously.

The DM resulted in much higher weighting of the commercial fishery age composition data; approximately 77% of the nominal sample size rather than 4% as iteratively weighted (in this case the input sample size represents the number of sampled fishing trips). The DM also resulted in much higher weighting of the FISS age composition data with the maximum likelihood estimate for θ occurring at the upper bound, implying a sample size of 100% of the nominal sample size compared to 27% using the iterative approach (for the FISS age composition data the input sample size represents the number of sampled stations). The model did not produce a positive definite Hessian when θ was estimated at a bound (as in the survey case), so the value had to be fixed, thereby losing any propagation of the variance associated with the scaling. The model also did not converge when both θ parameters were estimated simultaneously, or when the survey θ was fixed at a large value (effective scaling near 100%), and the fishery θ estimated freely. Maximum gradients tended to be much larger (1-3 orders of magnitude) when either of the DM parameters were estimated.

As would be expected, the higher weighting of the compositional data through the use of the DM did improve the residual patterns in the age data; however, it also produced very large residuals (both the extreme values and on average) resulting in a lack of fit that was highly implausible for both fishery (Figure 14) and survey (Figure 15) age composition data. The distribution of Pearson residuals should be approximately standard normal, and therefore values larger than 2 should be relatively rare, and values greater than 3-4 implausible). In contrast, the Francis method employed in recent assessments, tends to down-weight the compositional data such that maximum Pearson residuals tend to be around 2 and only occasionally larger.

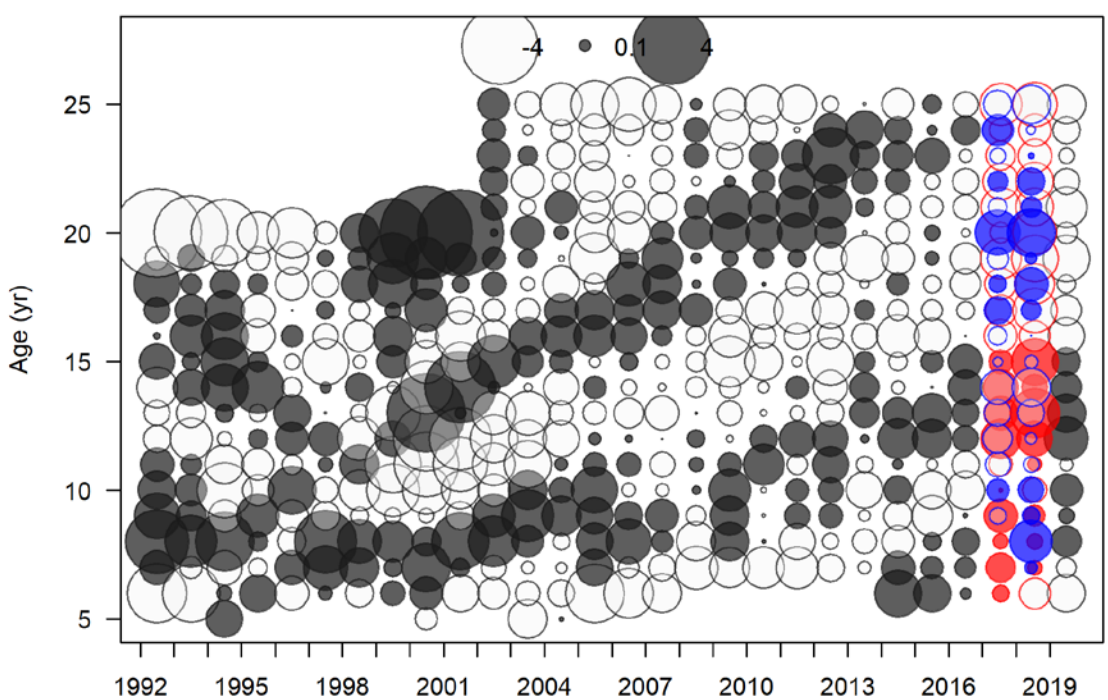

Figure 14. Commercial fishery Pearson residuals (observed minus expected) from the coastwide short model with approximately 77% of the nominal weighting for the input fishery age data (Dirichlet-multinomial θ estimated). Maximum residual in the figure is a value of 5.8.

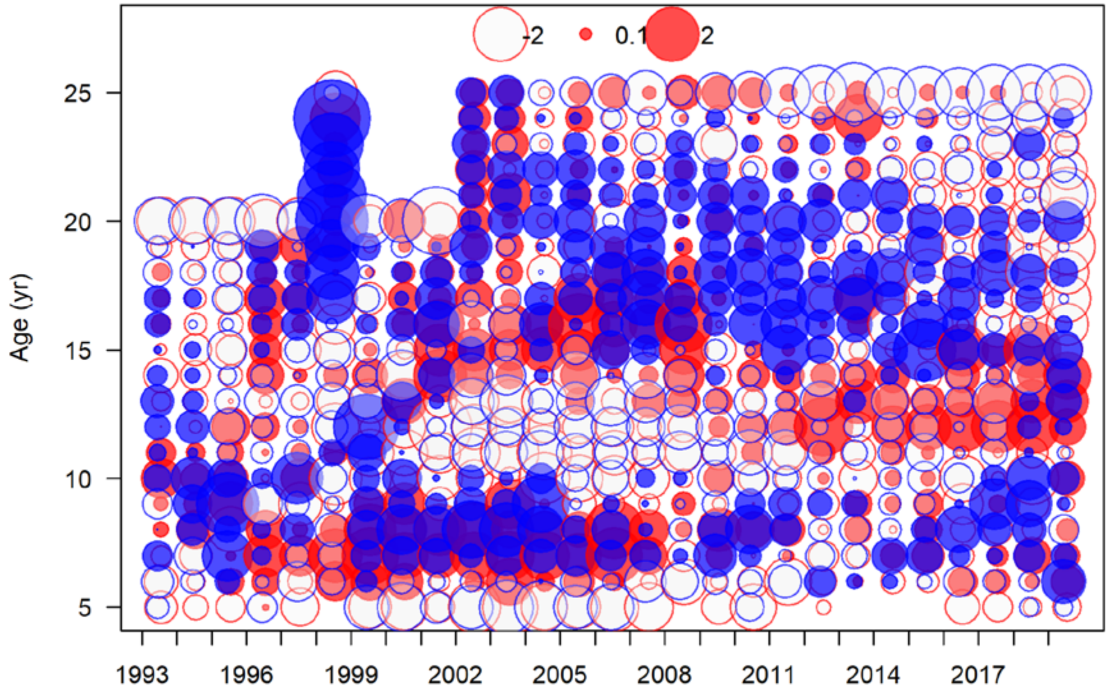

Figure 15. FISS Pearson residuals (observed minus expected) from the coastwide short model with 100% of the nominal weighting of the input survey age data (implied by estimating and then fixing the DM θ at a large value). Maximum residual in the figure is a value of 4.0.

Although the spawning biomass and recruitment time-series' were relatively similar for these two approaches to weighting, the DM did produce a larger variance estimate on the spawning biomass in the early years of the series when θ was estimated for the commercial fishery age composition data (Figure 16). At the end of the time-series there was little difference in the credible interval for spawning biomass, and the effect on the variance of recruitment was mixed.

In aggregate, it is not clear that the use of the DM would improve the Pacific halibut assessment models, and would appear to result in greater instability as well as a reduction in internal consistency between input variance and sample sizes, error distribution assumptions and output statistics of model fit. Data weighting remains an important avenue for further investigation, and additional tools and approaches may be available for future analyses, including error distributions that include a correlation structure that better matches that of actual compositional information(Albertsen et al. 2017).

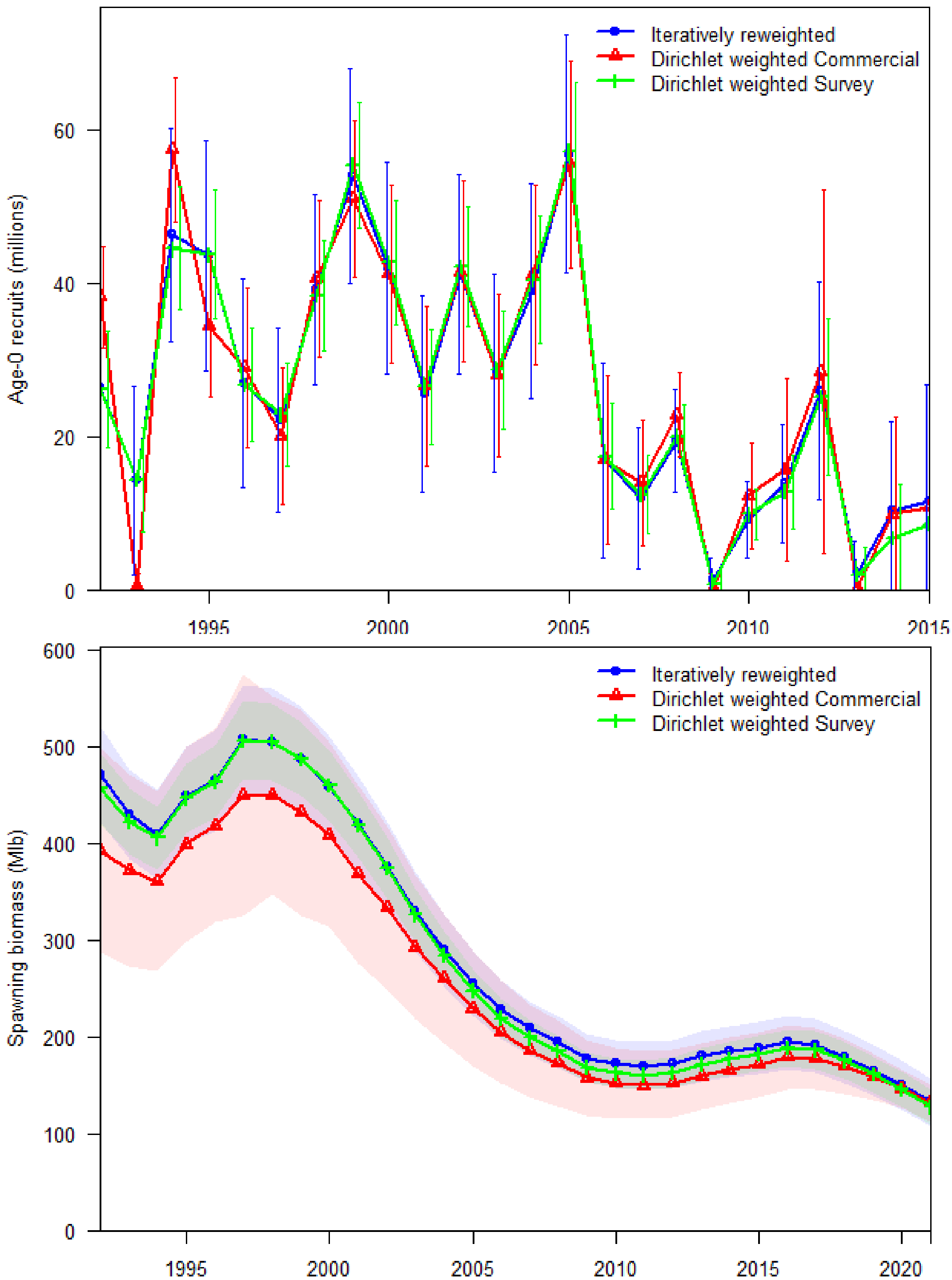

Figure 16. Time series of recruitment (upper panel) and spawning biomass (lower panel) based on Dirichlet-multinomial weighting the fishery and survey (separately) compared to the results of the iterative reweighting based on the Francis method.

Recommendation 4, part b, section ii – advise on the impact of data reweighting

One of the questions posed for the SRB015 review was whether to annually update the data-weighting within each stock assessment model during years in which an update was being conducted. The expectation was that new data, particularly the sex-specific age composition data from the commercial fishery could have large enough effects on the relative weighting to necessitate re-weighting even for an updated stock assessment. The 2019 stock assessment was a somewhat unique case, with the final assessment relying on substantially updated modelled survey results as well as an additional year of sex-specific commercial age data (see 4, a above). In order to evaluate the likely degree of change in data weighting as additional years of data are added to the final 2019 stock assessment, a retrospective approach was employed. First, the iterated input age composition data sample sizes for each model from the final 2019 stock assessment were summarized by model fleet (column 2, Table 1). Next, two years of data (2019, 2018) were sequentially removed from the assessment model, and the age composition data re-weighted at each step (columns 3-4, Table 1). Evaluated based on data through 2017, 2018 and 2019, this exercise mimics what might be seen over a 2-3 year update period between full assessments.

The results show that the iteratively re-weighted sample sizes for the composition data did not change nearly as much with each year of additional data as was observed for the final 2019 stock assessment. However, in a few cases (e.g., Region 3 survey age data in the AAF short model) changes in weighting were likely large enough to warrant re-weighting. It seems most logically appealing to update the data weighting annually, both to retain the desirable internal model consistency as well as to allow for cases when a time-series may change (such as 2019) necessitating a larger change in weighting.

Recommendation 4, part b, section iii – review and report on software developments

The IPHC has relied on a variety of model platforms for implementing its stock assessment, many of which have been developed specifically for Pacific halibut (e.g., Clark and Hare 2006; Deriso et al. 1985; Quinn et al. 1990). From 2012 to 2014, the IPHC transitioned from a single stock assesment model to an ensemble of models including alternative structural assumptions. At the same time, the software platform was also transitioned from the previous halibut-specific model implemented directly in ADMB to models using stock synthesis, a generalized analysis platform capable of a wide range of model structural configurations and providing consistently formatted output (Methot and Wetzel 2013a; Methot and Wetzel 2013b). This transition was made in order to speed the evaluation of a wide range of alternative models, facilitate quantitative summary of multiple models, reduce the potential for undiagnosed coding errors, and provide for more tranparent review. The benefits of using a generalized platform for the Pacific halibut stock assessment come with costs, including lack of some parameterizations specific to the needs of the analysis, delayed development of new approaches, and in some cases run times that are inflated due to unused model features. These pros and cons have been discussed the the SRB and were noted in the 2019 external review (Stokes 2019).

IPHC Secretariat staff attended and presented at the CAPAM workshop on the creation of frameworks for the next generation general stock assessment models 4-8 November 2019. That workshop covered widely ranging topics from programming languages and technical considerations, to high-level usage needs including simulation-testing and Management Strategy Evaluation (MSE). The workshop identifed a range of existing generalized softare platforms (e.g., stock synthesis, CASAL, SAM, MULTIFAN, etc.) that have many shared capabilites, but also many unique features that could make reconsiliation into one 'super-model' extremely challenging (Punt et al. 2020).

This workshop represented an important opportunity for the IPHC to participate in the planning for future generalized platforms, and a venue to survey the likely toptions and timeline for such tools. Although there were few specific conclusions, it was very clear that next generation generalized stock assesement software will require a long (multi-year) time-frame and commitments from one or more large national fisheries agencies/organizations.

From the IPHC's perspective, stock synthesis currently meets the assesment modelling needs, albiet with some constraints on specific features (e.g., random effects, flexible tagging parameterizations, etc.). As noted in the 2019 external review, it may be desirable to minimize large changes in the stock assesment in the short-term in order to best facilitate implementation of a management procedure resulting from the current MSE (Stokes 2019). Further, the development of the operating model for use in the MSE (largely based on the framework of the current stock assesment but programmed independently) is and will continue to refine the Secretariat's understanding of key biological processes and technical modelling needs that may feed back to the stock assessment. Additionally, the MSE framework will be useful for testing the stock assessment behavior under various assumptions through simulation.

Ultimately, the choice of a medium- to long-term assessment platform may depend on the type of management procedure selected. The current compressed stock assesment analysis conducted each fall in order to provide annual management information is based on the current year's data and must be stable and simple enough to be completed in less than two weeks. If a management procedure based on modelled survey trends, or a multi-year procedure is adopted, it may be uneccesary to conduct annual stock assessments. That type of procedure and timeline could allow for the development of more complex stock assessment ensembles/models (including fully Bayesian analyses), given extended development time between assessments. Therefore, the MSE, adoption of a management procedure by the IPHC and strategic planning for the stock assessment modelling platform should be considered together and the long-term focus should be on selecting the most efficient tools to meet management needs as they continue to evolve.

NEW DATA AVAILABLE

Since 2014, the stock assessment has included age composition data collected from the recreational fishery (charter and private sectors combined) in IPHC Regulatory 3A. These data are only generally representative of the coastwide recreational mortality, but comprise the only source of age information from this sector. Therefore, these data are included in the models in order to estimate a time-invariant selectivity curve for recreational removals, but are substantially down-weighted to avoid creating spurious information on recruitment (Stewart and Hicks 2019b). For 2020, Alaska Department of Fish and Game (ADFG) staff were able to re-process the subset of the available ages that included sex-specific information to produce age compositions for the entire time series (1994-2018; Sarah Webster, ADFG, personal communication; Figure 17). These data indicated that the recreational fishery has been harvesting an average of 72% female Pacific halibut in recent decades. The stock assessment models were updated to include this new information, and selectivity for male Pacific halibut was allowed to differ from that of females as for other fleets in the models. The previous model configuration assumed equal selectivity at age, but estimated a dome-shaped relationship to account for the lack of older individuals. By allowing relative male selectivity to decrease, the effect on model results was to increase both the estimated female spawning biomass and relative fishing intensity slightly. Unlike the change to commercial fishery sex-ratios (which relate to roughly 2/3rds of the total mortality) observed in the 2019 stock assessment, the recreational sex-ratio information resulted in only a minor change to the scale of the assessment: a 2% increase in female spawning biomass at the end of the time-series and a 1% change in fishing intensity. These data will also be included in the final step-wise bridging analysis provided with the final 2020 stock assessment, so that their effect on stock estimates can be compared with other data updates.

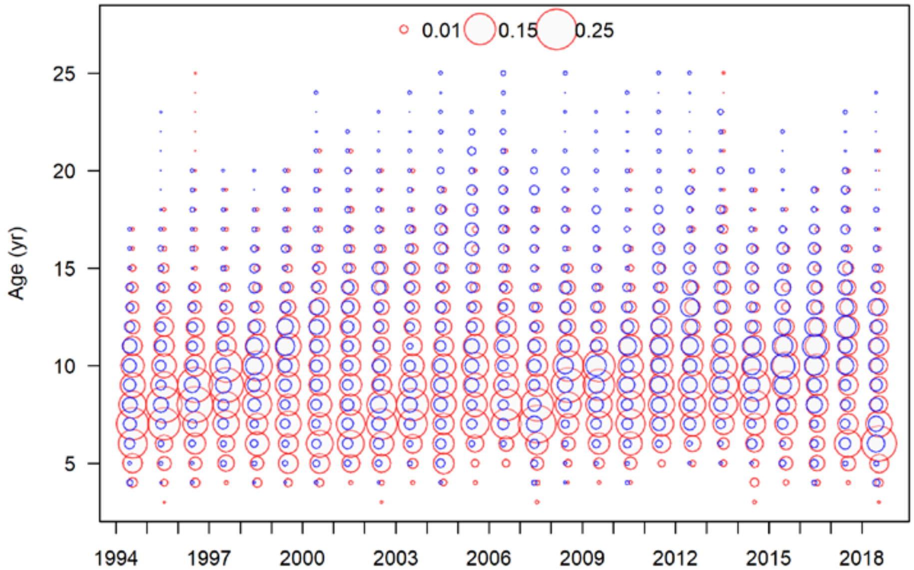

Figure 17. Time series of proportions at age for females (red circles) and males (blue circles) from the IPHC Regulatory Area 3A recreational fishery. Circle size is proportional to the relative number of fish in each category, proportions sum to 1.0 in each year.

DISCUSSION AND ADDITIONAL 2020 DEVELOPMENT

The results of the review and analyses reported here do not suggest that specific changes to the methods or basic assessment structure are necessary for 2020. Looking forward, it is anticipated that an additional year (2019) of sex-specific age composition data from the directed commercial fishery landings may be available for inclusion into preliminary models to be developed for SRB017. Additional data collected during 2020, including updates to existing data series will not be available until late October for inclusion in the final 2020 stock assessment. At the time this document was produced, a greatly restricted 2020 FISS was planned, to include only the center of the stock distribution in IPHC Regulatory Areas 2B-3B (see IPHC-2020-SRB016-05). This limited survey will provide some information to update the index, even for unsampled Areas via the space-time model, and age composition information in the stock assessment. It is also possible that some sectors may not harvest their full allocation in 2020, which may lead to a revised projection when actual mortality is included in the assessment. However, the situation remains highly uncertain and the 2020 assessment may therefore represent much more of a minor update to existing projections for 2021+ than a more standard analysis.

RECOMMENDATION/S

That the SRB:

a) **NOTE** paper IPHC-2020-SRB016-07 which provides a response to requests from SRB015 and an update on model development for 2020.
b) **REQUEST** any further analyses to be provided at SRB017, September 2020.

REFERENCES

Albertsen, C.M., Nielsen, A., and Thygesen, U.H. 2017. Choosing the observational likelihood in state-space stock assessment models. Can. J. Fish. Aquat. Sci. **74**(5): 779-789. doi:10.1139/cjfas-2015-0532.

Andersen, K.H., Jacobsen, N.S., and van Denderen, P.D. 2019. Limited impact of big fish mothers for population replenishment. Can. J. Fish. Aquat. Sci. **76**(3): 347-349. doi:10.1139/cjfas-2018-0354.

Berkeley, S.A., Chapman, C., and Sogard, S.M. 2004. Maternal age as a determinant of larval growth and survival in a marine fish, *Sebastes melanops*. Ecology **85**(5): 1258-1264.

Clark, W.G., and Hare, S.R. 2006. Assessment and management of Pacific halibut: data, methods, and policy. International Pacific Halibut Commission Scientific Report No. 83, Seattle, Washington. 104 p.

Deriso, R.B., Quinn, T.J., II, and Neal, P.R. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. **42**: 815-824.

Erikson, L. 2020. Fishery-independent setline survey (FISS) design and implementation in 2019. IPHC-2020-AM096-06. 11 p.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software **27**(2): 233-249.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. **68**: 1124-1138.

Francis, R.I.C.C. 2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. **192**(5-15). doi:10.1016/j.fishres.2016.06.006.

IPHC. 2018. Report of the 12th Session of the IPHC Scientific Review Board (SRB012). Seattle, Washington, U.S.A., 19-21 June 2018. IPHC-2018-SRB012-R. 17 p.

IPHC. 2019a. Report of the 15th session of the IPHC Scientific Review Board (SRB015). Seattle, Washington, U.S.A., 24-26 September 20-19. IPHC-2019-SRB015-R, 18 p.

IPHC. 2019b. Report of the 14th session of the IPHC Scientific Review Board (SRB014). Seattle, Washington, U.S.A., 26-28 June 2019. IPHC-2019-SRB014-R. 16 p.

Lee, H.-H., Maunder, M.N., Piner, K.R., and Methot, R.D. 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fish. Res. **125–126**(0): 254-261. doi:10.1016/j.fishres.2012.03.001.

Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R., Pourzand, R., and Ralston, S. 2013. A perspective on steepness, reference points, and stock assessment. Can. J. Fish. Aquat. Sci. **70**(6): 930-940.

Martell, S., Stewart, I., and Wor, C. 2016. Exploring index-based PSC limits for Pacific halibut. IPHC Report of Assessment and Research Activities 2015. p. 238-285.

Maunder, M.N., Crone, P.R., Punt, A.E., Valero, J.L., and Semmens, B.X. 2017. Data conflict and weighting, likelihood functions and process error. Fish. Res. **192**: 1-4. doi:10.1016/j.fishres.2017.03.006.

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling - importance resampling algorithm. Can. J. Fish. Aquat. Sci. **54**: 284-300.

Methot Jr, R.D., Wetzel, C.R., Taylor, I.G., and Doering, K. 2019. Stock Synthesis User Manual Version 3.30.14. NOAA Fisheries. Seattle, WA. 217 p.

Methot, R.D., and Wetzel, C.R. 2013a. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Appendix A: Technical description of the Stock Synthesis assessment program. Fish. Res. **142**: 26 p.

Methot, R.D., and Wetzel, C.R. 2013b. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**(0): 86-99. doi:http://dx.doi.org/10.1016/j.fishres.2012.10.012.

Punt, A.E., Dunn, A., Elvarsson, B.Þ., Hampton, J., Hoyle, S.D., Maunder, M.N., Methot, R.D., and Nielsen, A. 2020. Essential features of the next-generation integrated fisheries stock assessment package: A perspective. Fish. Res. **229**: 105617. doi:10.1016/j.fishres.2020.105617.

Quinn, T.J.I., Deriso, R.B., and Neal, P.R. 1990. Migratory catch-age analysis. Can. J. Fish. Aquat. Sci. **47**: 2315-2327.

Shelton, A.O., Hutchings, J.A., Waples, R.S., Keith, D.M., Akçakaya, H.R., and Dulvy, N.K. 2015. Maternal age effects on Atlantic cod recruitment and implications for future population trajectories. ICES Journal of Marine Science: Journal du Conseil **72**(6): 1769-1778. doi:10.1093/icesjms/fsv058.

Stewart, I., and Hicks, A. 2019a. Updates on the development of the 2019 stock assessment. IPHC-2019-SRB015-07. 10 p.

Stewart, I., and Hicks, A. 2019b. 2019 Pacific halibut (*Hippoglossus stenolepis*) stock assessment: development. IPHC-2019-SRB014-07. 100 p.

Stewart, I., and Hicks, A. 2020. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2019. IPHC-2020-SA-01. 32 p.

Stewart, I., and Webster, R. 2020. Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. IPHC-2020-SA-02. 53 p.

Stewart, I., Hicks, A., Webster, R., and Wilson, D. 2020. Summary of the data, stock assessment, and harvest decision table for Pacific halibut (*Hippoglossus stenolepis*) at the end of 2019. IPHC-2020-AM096-09 Rev_2. 26 p.

Stewart, I.J., and Hamel, O.S. 2014. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. **71**(4): 581-588. doi:10.1139/cjfas-2013-0289.

Stokes, K. 2019. Independent peer review for the 2019 IPHC stock assessment. August 2019. 31 p. https://www.iphc.int/uploads/pdf/sa/2019/stokes_2019-independent_peer_review_for_the_2019_iphc_stock_assessment.pdf.

Thorson, J.T. 2019. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. Fish. Res. ***217***: 133-139.

Thorson, J.T., Dorn, M.W., and Hamel, O.S. 2018. Steepness for West Coast rockfishes: Results from a twelve-year experiment in iterative regional meta-analysis. Fish. Res. doi:10.1016/j.fishres.2018.03.014.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. **192**: 84-93. doi:10.1016/j.fishres.2016.06.005.

Wang, S.-P., Maunder, M.N., Aires-da-Silva, A., and Bayliff, W.H. 2009. Evaluating fishery impacts: Application to bigeye tuna (Thunnus obesus) in the eastern Pacific Ocean. Fish. Res. **99**(2): 106-111. doi:10.1016/j.fishres.2009.05.010.

Webster, R. 2020. Space-time modelling of IPHC Fishery-Independent Setline Survey (FISS) data. IPHC-2020-AM096-07. 32 p.

Xu, H., Thorson, J.T., and Methot, R.D. 2020. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Can. J. Fish. Aquat. Sci. **77**(2): 247-263. doi:10.1139/cjfas-2019-0107.