Update on the development of the 2022 stock assessment

PREPARED BY: IPHC SECRETARIAT (I. STEWART & A. HICKS; 17 AUGUST 2022)

PURPOSE

To provide the IPHC's Scientific Review Board (SRB) a response to requests from SRB020 (IPHC-2022-SRB020-R) and to provide the Commission with an update on the development of the 2022 assessment.

INTRODUCTION

This document provides an update on stock assessment development progress since SRB020. The 2022 stock assessment represents a full analysis, following updates in 2020 and 2021 of the 2019 full assessment. The preliminary analysis presented for SRB020 (IPHC-2022-SRB020-7) included extensive detail on improvements to data sources, software, structural modelling choices, bridging analyses from the 2021 assessment, as well as the introduction of a new method for weighting models within the ensemble. This document includes a response to recommendations from SRB020, a description of three minor updates to the modelling and/or data, as well as a brief summary of data that will be included in the final assessment for 2022.

SRB RECOMMENDATIONS, REQUESTS AND RESULTS

The SRB made the following assessment recommendations and requests during SRB020:

1) SRB020-Rec.02 (para. 23):

   "The SRB NOTED that most models within the ensemble produced reasonable and well-constrained estimates of natural mortality (M) and RECOMMENDED that estimation of M should be adopted in the short AAF assessment model with consideration in other models as part of the stock assessment research program."

2) SRB020-Rec.03 (para. 24):

   "The SRB NOTED that the bootstrapping approach to determining maximum samples sizes for age-composition data improved assessment model performance and stability and, therefore, RECOMMENDED that the bootstrapping approach be adopted for data-weighting in future assessments."

3) SRB020-Rec.04 (para. 25):

   "The SRB NOTED apparent discrepancies in marine mammal prevalence among anecdotal reports, FISS observations, and preliminary evaluation of logbook data, and therefore RECOMMENDED further investigation of methods to better estimate marine mammal prevalence and impacts on the fishery."

4) SRB020-Req.06 (para. 26):

   "The SRB NOTED the proposed new ensemble model weighting scheme using the MASE criterion and REQUESTED investigation of predictive skill on additional quantities such as fishery CPUE and mean age in FISS samples."

A response to each of these requests is provided below.

1. Estimation of natural mortality

As in the preliminary assessment, the final 2022 stock assessment is planned to include a short Areas-As-Fleets (AAF) model that includes estimation of natural mortality (*M*) constrained by a relatively diffuse prior based on maximum observed age. Future modelling exploration for the coastwide (CW) short model will focus on parameters and processes that may be correlated and/or confounded with *M* and a search for a set of model structural assumptions that would allow estimation of *M* in this model as well.

2. Bootstrapped sample sizes

Also consistent with the preliminary assessment presented at SRB020, the final 2022 stock assessment will begin internal data weighting from bootstrapped sample sizes following the methods described in Stewart and Hamel (2014) and Stewart and Hicks (2022). The addition of sex-ratio at age data from 2021 (described below) provided a test of the technical feasibility of adding an additional processing step to the normal work-flow of data input to the models included in the ensemble. Because the bootstrapping code has been integrated into the R code for age composition generation, this additional processing step does not appreciably affect the efficiency of model updating.

3. Marine mammal depredation

Following SRB020 the secretariat has continued to explore avenues for better understanding the prevalence of marine mammal depredation in the directed commercial Pacific halibut fishery. Following the end of the 2022 fishing season, the logbook fields being recorded, the way in which IPHC field specialists collect the information if the harvester has not filled out the fields and an improved outreach program will be evaluated. An update of the existing analysis of depredation is underway for the sablefish (*Anoplopoma fimbria*) fishery (Goethel et al. 2021). The IPHC secretariat will coordinate with this effort, as the logbooks are collected by IPHC field specialists and the harvesters overlap substantially between the two fisheries. Finally, the IPHC has an ongoing research project to test catch protection methods to reduce depredation in the directed commercial fishery using either a 'shuttle' to collect the fish from the hooks underwater or a 'shroud' to cover fish captured on branchline-rigged demersal longlines. These devices are being adapted from technology developed in the southern ocean and applied to toothfish fisheries. An update on all of these efforts will be provided at SRB022 in June 2023.

4. Model weighting

Based on the initial investigation of model weighting presented at SRB020, the secretariat has continued evaluation of Mean Absolute Standardized Error (MASE; Hyndman and Koehler 2006) as a tool for measuring the skill of each model in predicting one year ahead observations. The MASE statistic is calculated as

$$MASE = \frac{\frac{1}{n}\sum_{t=1}^{n} \left| \frac{O_t - E_t}{\sigma_t} \right|}{\frac{1}{n}\sum_{t=1}^{n} \left| \frac{O_t - O_{t-1}}{\sigma_t} \right|}$$

Where *O* indicates the observation at time *t*, *E* the prediction (or expected value) and σ_t is the standard deviation of the observation. The calculation can be averaged over any number of years or lags relevant to the predictive problem. As defined, MASE estimates must be positive, and the range of values is interpreted as:

>1: model predictive skill is worse than the naïve prediction (last year's index) – model not worth pursuing further

1: model predictive skill is exactly equal to the naïve prediction

<1: model predictive skill exceeds that of the naïve prediction

0: model predictions perfectly match subsequent observations

In order to turn the MASE statistic into a model weight we need to specify the scale of the weighting and the behavior at the end-points. In this case, for model (*m*) within the set of models (*M*; limited to those models with MASE values <1):

$$MASE\ weight_m = \frac{1 - MASE_m}{\sum_{m=1}^{M} 1 - MASE_m}$$

This approach ensures that a model that does not outperform the naïve prediction (MASE >= 1) will get zero weight, and that a set of models all perfectly predicting the next observation will receive equal weights.

Initial application of this approach to the Fishery Independent Setline Survey (FISS) index of abundance presented at SRB020 resulted in all models performing better than the naïve predictor when averaging over the most recent 1, 2 3, or 4 years (Table 1, Figure 1). Therefore, MASE weights averaging over the same time-periods were relatively stable (Table 2). The secretariat therefore preliminarily recommended that MASE weights be calculated based on the most recent year's performance only, in order to tie the weights to the most relevant performance observed in the time series.

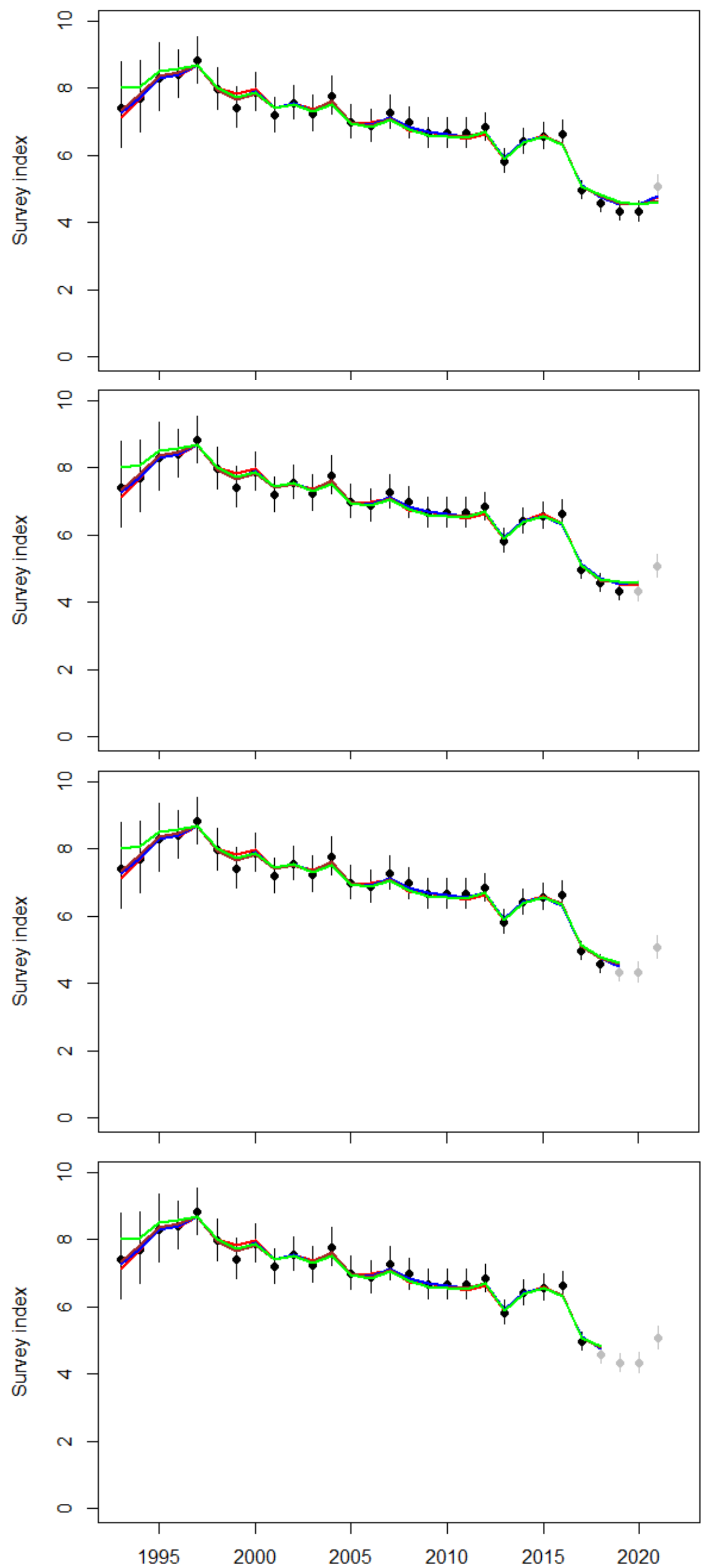


Figure 1. Predictions from each of the four models (colored lines) for the 2021 to 2018 (top to bottom panels) FISS observations (grey dots and CIs) using data through 2020 to 2017 (black dots and CI).

Table 1. One-year ahead standardized MASE estimates for each of the four stock assessment models averaged over the most recent 1, 2, 3, and 4 years.

	Model			
Years included	CW short	CW long	AAF short	AAF long
4	0.70	0.65	0.82	0.72
3	0.83	0.75	0.94	0.83
2	0.86	0.76	0.88	0.78
1	0.59	0.46	0.52	0.44

Table 2. One-year ahead standardized MASE weights for each of the four stock assessment models averaged over the most recent 1, 2, 3, and 4 years.

	Model			
Years included	CW short	CW long	AAF short	AAF long
4	27.5%	31.3%	15.8%	25.4%
3	26.0%	38.0%	9.3%	26.8%
2	19.1%	33.9%	16.4%	30.6%
1	20.5%	27.2%	24.0%	28.3%
Status quo weights	25.0%	25.0%	25.0%	25.0%

At the request of SRB020, this calculation was extended to include coastwide commercial fishery Weight-Per-Unit-Effort (WPUE). In contrast to the FISS index, the fishery WPUE index has been nearly flat over the most recent four years (Figure 2). With very little contrast, the naïve predictor (the previous index value) equaled or exceeded model predictions, and led to weighting that varied from 0.0 to 100% and never included more than 2 models in any of the averaged periods (Table 3), except in 2021 when all four models performed more poorly than the naïve predictor and were assigned equal weights. This difference from the FISS-based weights is probably also enhanced by the use of time-vary catchability for the fishery WPUE, which provides for relatively good fits to the data, but relatively poor forward predictions.

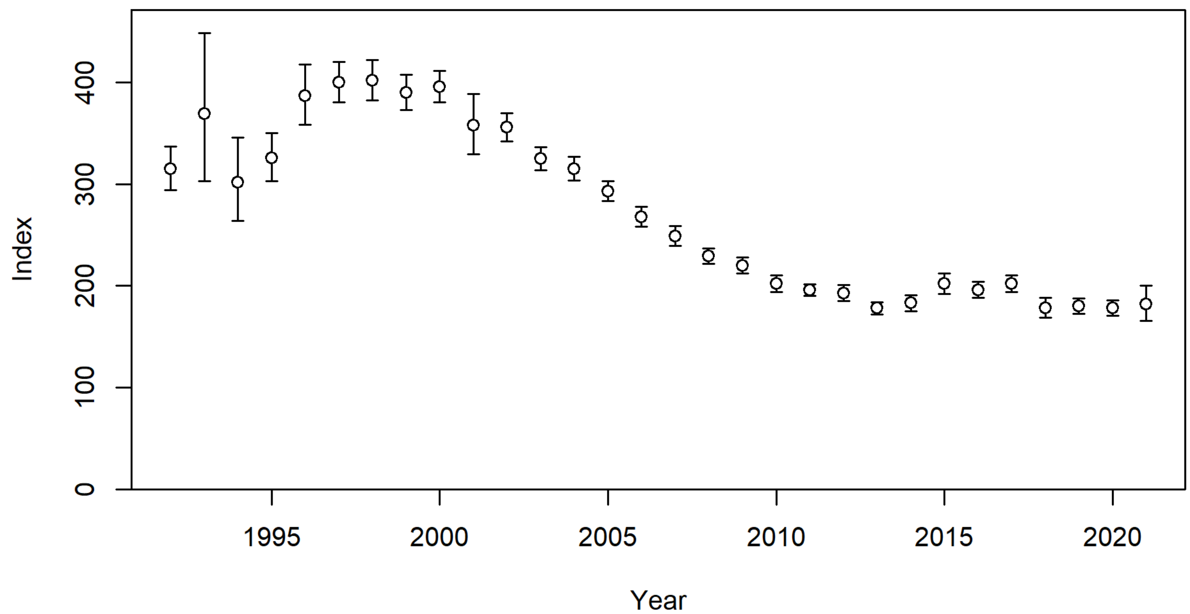


Figure 2. Time series of directed commercial fishery WPUE observations (grey dots and approximate 95% CIs), 1992-2021.

Table 3. One-year ahead standardized MASE weights based on commercial fishery WPUE for each of the four stock assessment models averaged over the most recent 1, 2, 3, and 4 years. Note that in 2021, all four models had MASE values >1.

	Model			
Years included	CW short	CW long	AAF short	AAF long
4	0.0%	53.1%	0.0%	46.9%
3	0.0%	27.8%	0.0%	72.2%
2	0.0%	0.0%	0.0%	100.0%
1	25.0%	25.0%	25.0%	25.0%
Status quo weights	25.0%	25.0%	25.0%	25.0%

Based on the variability observed for the commercial fishery WPUE, the MASE weights based on FISS predictions were further evaluated to better understand the role of contrast in the underlying data to the calculated weights. In the case of the FISS, the 2021 observation represented a sharp increase from earlier observations (Figure 1, top panel), which was predicted well by all four models. This meant that as 2-, 3- and 4-year averages were calculated, the 2021 prediction maintained the stability in the weighting. When single-year MASE weights were calculated for each of the four years alone, they were more variable ranging from 0.0 to 45.5% (Figure 3). In 2020, all four models performed more poorly than the naïve predictor, and so were assigned equal weights.

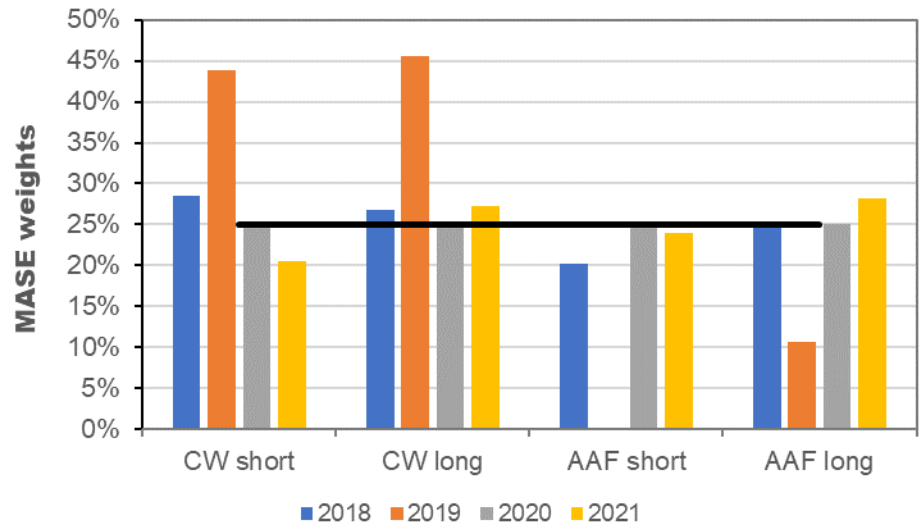


Figure 3. Comparison of one-year MASE weights for each of the four models calculated in each of the most recent four years. Horizontal line indicates the *status quo* equal weighting (25%).

With only four years of data to work with (2018-2021) the 2- 3- and 4-year averages calculated in each of the terminal years have fewer replicates to compare (Figure 4). These comparisons show that the ability of the MASE statistic to rank and weight models depends heavily on the contrast in the underlying observations. It seems desirable to have at least one year included in the calculation that has the 'power' to detect model skill. This is also consistent with the concept that management quantities will be most affected by rapid changes in the index (either up or down) and therefore the model predictive skill when the stock is changing is most relevant.

Based on this extended evaluation, the secretariat recommends moving forward with 4-year average MASE weights based on FISS predictions for the 2022 assessment (Table 2, top row; Figure 4, bottom panel). Looking forward, a 4-year moving average will continue to include the 2021 prediction through the next full stock assessment planned for 2025, and therefore have a reduced risk of large and/or abrupt changes in model weighting during the updated assessments conducted in 2023 and 2024. However, it will still provide for an updating of model weights as individual model performance evolves. In 2025, with several additional years of data and weighting available a more informed evaluation of the stability and performance of MASE weights can be undertaken.

Figure 4. Comparison of 2-year (top panel), 3-year (middle panel) and 4 year (lower panel) MASE weights for each of the four models by terminal year. Horizontal lines indicate the *status quo* equal weighting (25%). Note that the scale of the y-axis differs among panels.

ADDITIONAL STOCK ASSESSMENT DEVELOPMENT IN 2022

In addition to the request and recommendations made during SRB020, the secretariat has also continued with minor updates and improvements to each of the models and data sets. These included: 1) an investigation of the use of small constants in the population size-at-age calculations within in the Stock Synthesis modelling software, and 2) the effect of sparse weight-at-age data observed in 2021 for the oldest ages observed in IPHC Regulatory Area 3A.

During ongoing development of the Stock Synthesis software a potential convergence issue was identified relating to the small constant added to the internal age-length key used to convert numbers of fish to size and biomass. Previously, a small constant (0.0001) was recommended to be added to the calculations (page 21, Methot Jr et al. 2021). In cases where the growth curve was internally estimated, this small constant was found to occasionally cause convergence issues. Because the Pacific halibut model uses empirical weight-at-age it was largely unaffected by this issue; model runs removing this constant did not differ after routine rounding for any management quantities.

During development of the Operating Models for the 2022 Management Strategy Evaluation it was discovered that very sparse data for the oldest ages in Biological Region 3 had led to negative estimated raw weight-at-age for female Pacific halibut. This was caused by the extrapolation of trend from the last two ages with data, which showed a negative trend. The issue only arose for the raw observations used in fitting the FISS index, as all population matrices are smoothed to reduce the effects of observation error (Stewart and Webster 2022). Because there were only trivial numbers of fish in these ages, when the weight-at-age was forced to remain constant across missing ages there was no change in model fit or estimated quantities.

The sex-ratio-at-age based on genetic analyses for the biological samples collected during the 2021 fishery were made available by the Biological and Ecosystem Sciences Branch in time for inclusion in a set of preliminary model runs. The sex-ratio-at-age information from 2021 was largely consistent with model estimates and resulted in only minor changes (<2%) to the spawning biomass of any of the four models. However, these data continue to improve our understanding of recent population dynamics. This information has been critically important to accurately estimating the spawning biomass and the effect of fishing on the lifetime reproductive output of the stock beginning with the 2019 stock assessment. As the time-series grows longer, it is now possible to evaluate better how the sex ratios are changing over time, and to better delineate trends from interannual variability. Generally, the observed sex-ratio is closely correlated to the average age in the landings: younger fish are a higher proportion female than older fish (Figure 5). Thus, as the 2005 year-class has aged the proportion female has generally decreased, and the pattern was mixed among Biological Regions in 2021, reflecting the uneven contribution of the younger 2012 year-class to those landings (Table 4). The contributions of the 2011 and 2012 year-classes can be clearly seen in the female age information from 2021, while male landings continue to be largely comprised of the 2005 and older year-classes (Figure 6). Some of the observed variability in the sex-ratio information is likely due to sampling variability; in particular Biological Region 4B has included only 10-17 fishery deliveries sampled over this period, and Region 4 only 47 and 43 deliveries in 2020 and 2021, down from over 100 in earlier years. With five years of sex-ratio information now available it may be timely to consider whether annual processing of the genetic samples is optimal, given the cost and trade-off with other potential research.

Figure 5. Relationship between percent female and average age in the directed commercial Pacific halibut fishery by Biological Region. Each point represents one year (2017-2021).

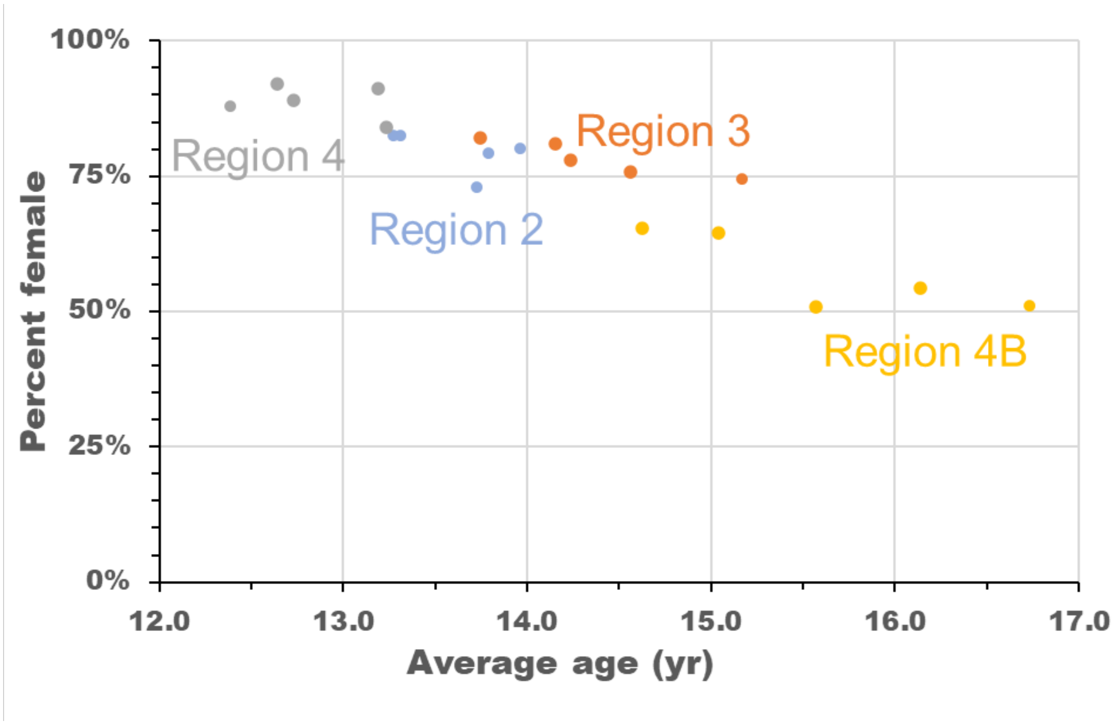
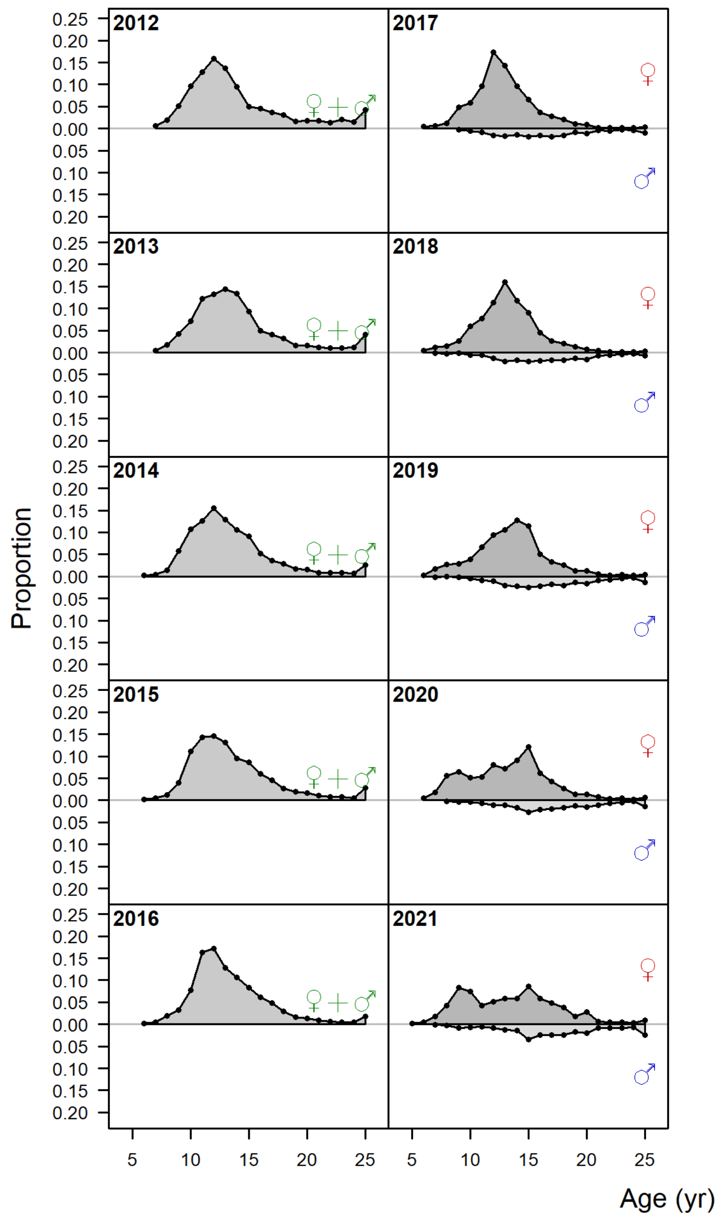


Table 4. Percent of the directed commercial fishery landings comprised of female Pacific halibut.

Year	Coastwide	Biological Region 2	Biological Region 3	Biological Region 4	Biological Region 4B
2017	82%	82%	82%	92%	65%
2018	80%	82%	78%	91%	65%
2019	78%	80%	76%	89%	51%
2020	80%	79%	81%	84%	54%
2021	74%	73%	74%	88%	51%

Figure 6. Recent age compositions from the directed commercial landings.



PRELIMINARY DATA UPDATES

No preliminary data was available from 2022 in time for this document. Standard data sources that will be included in the final 2022 stock assessment include:

1) New modelled trend information from the 2022 FISS for all IPHC Regulatory Areas.
2) Age, length, individual weight, and average weight-at-age estimates from the 2022 FISS.
3) Directed commercial fishery logbook trend information from 2022 (and any earlier logs that were not available for the 2021 assessment) for all IPHC Regulatory Areas.
4) Directed commercial fishery biological sampling from 2022 (age, length, individual weight, and average weight-at-age) from all IPHC Regulatory Areas.

5) Biological information (lengths and/or ages) from non-directed discards (all IPHC Regulatory Areas) and the recreational fishery (IPHC Regulatory Area 3A only) from 2021. The availability of these data routinely lags one year.
6) Updated mortality estimates from all sources for 2021 (where preliminary values were used) and estimates for all sources in 2022.

RECOMMENDATION/S

That the SRB:

a) **NOTE** paper IPHC-2022-SRB021-08 which provides a response to requests from SRB020, and an update on model development for 2022.
b) **RECOMMEND** any changes to be included in the final 2022 stock assessment to be completed for presentation at IM098.
c) **REQUEST** any further analyses to be provided at SRB022, June 2023.

REFERENCES

Goethel, D.R., Hanselman, D.H., Rodgveller, C.J., Echave, K.B., Williams, B.C., Shotwell, S.K., Sullivan, J.Y., Hulson, P.F., Malecha, P.W., Siwicke, K.A., and Lunsford, C.R. 2021. Assessment of the sablefish stock in Alaska. NPFMC Bering Sea, Aleutian Islands and Gulf of Alaska SAFE. Chapter 3. 347 p.

Hyndman, R.J., and Koehler, A.B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting **22**(4): 679-688. doi:10.1016/j.ijforecast.2006.03.001.

IPHC. 2021. Report of the 18th session of the IPHC Scientific Review Board (SRB018). IPHC-2021-SRB018-R. 21 p.

IPHC. 2022. Report of the 20th session of the IPHC Scientific Review Board (SRB020). Meeting held electronically, 14-16 June 2022. IPHC-2022-SRB020-R. 19 p.

Methot Jr, R.D., Wetzel, C.R., Taylor, I.G., Doering, K.L., and Johnson, K.F. 2021. Stock synthesis user manual version 3.30.18. NOAA Fisheries. Seattle, WA. 234 p.

Stewart, I., and Hicks, A. 2022. Development of the 2022 Pacific halibut (*Hippoglossus stenolepis*) stock assessment. IPHC-2022-SRB020-07. 128 p.

Stewart, I., and Webster, R. 2022. Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. IPHC-2022-SA-02. 56 p.

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