Pacific halibut stock assessment development for 2017

PREPARED BY: IPHC SECRETARIAT (I. STEWART; 31 AUGUST 2017)

PURPOSE

To provide the Scientific Review Board (SRB) an overview of data and modelling updates, as well as a preliminary evaluation of the stock assessment ensemble proposed for use in the 2017-2018 annual process.

DATA SOURCE DEVELOPMENT

Measured individual fish weights

The IPHC has relied on a standard length-weight relationship for decades to infer individual fish weight from length measurements (Clark 1991). Recent data collection from the landed catch and sampling at-sea has revealed some systematic differences between observed individual weights and weights predicted from the length-weight relationship (Webster and Erikson 2017, Webster et al. 2016). During 2016, all fish sampled for length and age during IPHC port sampling were also weighed (Erikson and Kong 2017). To improve estimates of average individual weight reported for use in the stock assessment and for catch-weighting of age-frequency data (Stewart 2017), it is proposed that the 2017 stock assessment and supporting analyses use the measured weight for all samples where it is available instead of the predicted weight. This change was made to the data processing scripts by implementing the following steps for each individual biological sample collected:

1) Identify whether a dressed (head-on and gutted) and washed weight ($W_{d,w}$; pounds) was collected. If so, adjust this to net weight (head off; W_n) to be consistent with historically calculated net weights via the equation:

$$W_n = W_{d,w} \cdot 0.88$$

2) In the case that no dressed and washed weight was available, but a dressed and unwashed weight (still including some ice and slime, but not the cavity ice; $W_{d,uw}$) was collected, adjust to net weight via:

$$W_n = W_{d,uw} \cdot 0.98 \cdot 0.88$$

3) In the case that neither head-on weight was collected, adjust to net weight from fork length (L_f; cm) via:

$$W_n = 0.00000692 \cdot {L_f}^{3.24}$$

These weights are collected in port, and are therefore not sex-specific. This means that for analysis purposes, they are important primarily for weighting catch-weighting the age composition samples (catch is estimated in numbers for each Regulatory area via dividing the catch weight by the average individual fish weight) and for presentation purposes of fishery trends. This change was applied to the data collected in 2015, where there were 4,527 unwashed and 861 washed weights, and in 2016 where there were 9,691 unwashed and 980 washed weights.

There were only minor changes (~2.5% lower in aggregate) to the estimated average weight of Pacific halibut landed (Appendix A). When the revised commercial fishery age composition data were included in the stock assessment models, as a test of the importance of this change, the difference in terminal spawning biomass was less than 0.1%.

A similar approach may be possible for fish caught by the IPHC's fishery-independent setline survey if/when weights are collected on a routine basis. A pilot project in ongoing to better understand the quality of and variability in at-sea weights as well as to measure the effects of shrinkage in the hold during the fishing trip through longitudinal sampling of individual fish (Planas 2017). Similar differences observed in the survey samples would be likely to have a larger effect on stock assessment inputs, as those samples are used to define the population weight- and maturity-at-age as well as the sex-specific fishery transition from catch in numbers to catch in weight (Stewart and Martell 2016).

Fishery independent setline survey time-series extension to include 1995-1997.

The extension of the space-time (geospatial) model for setline survey analysis to include 1995-1997 is anticipated to be available for the 2017 stock assessment and stock distribution analyses. Further extension to include additional historical years could be considered, but because of differences in survey standardization (i.e. the bait used) this may require a separate modelling exercise. Therefore, it may be most pragmatic to retain the older historical series in their current form (naïve average catch-rates) and allow variability in survey catchability and selectivity (already included in the assessment models) to capture this source of variability. Focus of modelling efforts could then be placed on including covariates and other improvements to the recent time-series (see paper IPHC-2017-SRB-10-02).

The results of this extension will be available after the production of this document, but in time for inclusion in the 2017 stock assessment.

Continued investigation of historical bycatch estimates and length-frequency data

It has been IPHC practice to correct previous years bycatch mortality estimates, once observed fishery data become available (generally 3-6 months after the end of the year). This process is similar to the calculation of Discard Mortality Rates (DMRs) by the National Marine Fisheries Service (NMFS) for use in the catch-accounting system; however, it uses observed viabilities by fishery from the year being estimated, rather than values predicted for use in in-season management based on previous data. Although the differences in final bycatch estimates are generally small, they can be systematic, as was the case for decreasing DMRs in the freezer-longliner fishery for Pacific cod over the last decade (e.g., Williams 2014, Williams 2016).

A database table has been created at the IPHC to store the catch and biological data obtained annually from the NMFS Observer Program. Use of this table will allow re-evaluation and documentation of the aggregation methods for these analyses as well as automation of the process in the future. It will also allow estimates to be re-created as needed, rather than estimated once and stored only as derived information in perpetuity.

Length-frequency data by Regulatory Area has historically been summarized by the IPHC on an *ad hoc* basis. We currently use a time-series consisting of aggregate estimates with little to no meta-data and likely differing methods of aggregation in different years (e.g., catch weighted, raw length-frequencies, projected values from incomplete data, etc.). These data are currently used in the stock assessment model to inform the selectivity curve describing bycatch removals, but are down-weighted due to these concerns over standardization (Stewart and Martell 2016). Ongoing efforts in 2017 and moving forward will include: 1) identifying raw data sets suitable for inclusion, 2) re-estimating length-frequency distributions using standardized methods for all available years, 3) recording meta-data and results such that they can be recreated if changes to the analysis approach are desired in the future, and 4) updating the stock assessment model inputs to reflect the best available series, potentially increasing the weight on these data, while allowing for an appropriate degree of temporal variability in selectivity to reflect differences among areas, fishing fleets and other factors.

Due to staffing changes within the IPHC secretariat, there has been no additional progress on this effort. Although all series will be updated for the 2017 stock assessment, it is not anticipated that a revised historical series will be available.

Biological data from setline survey expansion stations

It became evident during 2017 that the biological data (ages and lengths) from expansion stations sampled by the IPHC's fishery independent setline survey had not been included in the summaries produced for the annual stock assessment. This was due to the use of alternative values in the database table indicating the purpose of each set as part of the standardized grid, or alternative projects. Although these data represent a relatively small fraction of the total available (approximately 10% over the first few years of expansion), their omission is inconsistent with the use of a trend index from the space-time (geospatial) model that includes predictions at all of the expansion stations.

For the 2017 stock assessment analyses, these data will be added. Since the setline survey selectivity is already parameterized to include a modest amount of temporal variability (Stewart and Martell 2016), no change was necessary to the model structure to accommodate these additional data. These additional age data are included in the stock assessment models in three ways: 1) via the empirical weight-at-age matrices, 2) directly through the survey age frequency distributions, and 3) via the sublegal age frequency distributions used to estimate selectivity of the discarded portion of the commercial catch (Stewart and Martell 2016).

To compare the effects of changes to the data sets to be included in the 2017 stock assessment, the four models were first extended by one year to include the removals from all sources projected for 2017 (Appendix B). This allows for direct comparison of the effect of all changes on the 2018 beginning-of-the-year female spawning biomass, which will be one of the primary outputs for use in the 2017-2018 management process. Building from the models using measured individual Pacific halibut weights from port sampling, as described above, the assessment models were updated with the sublegal and survey age distributions as well as the re-estimated weights at age using these additional biological samples (Stewart and Martell 2016). These data produced slightly differing results from the 2016 assessment, with the largest single-model difference in terminal spawning biomass being +7%. The aggregate change to the ensemble was slightly less, only +3.6%, which has been well within the range of inter-annual variability observed over the last few years (Appendix B).

Future efforts could include application of an approach to standardize the age-compositions in a space-time (geospatial) model, such that abundance weighting would be fully logically consistent with the spatial trend information. This is an avenue of ongoing inquiry that will be addressed in the planned Center for Advancement of Population Assessment Modelling

(CAPAM) meeting in February, 2018. IPHC collaboration on supporting research for this conference is underway.

Effective skate calculations

Recent research comprising a portion of Cole Monnahan's PhD thesis (University of Washington, graduated June, 2017) has re-evaluated the hook-spacing/power relationship used to standardize commercial Pacific halibut fishery logbook records (Monnahan and Stewart 2015). Pending scientific review, a published hook-spacing correction derived from logbook records and compared with original experimental observations (Hamley and Skud 1978) may soon be available. The IPHC may consider updating the formula used in its databases for calculating the number of effective skates for the commercial fishery data (Stewart 2017).

Although this improved estimate is not identical to the *status quo* relationship, the likely effect on current analysis methods is small. Future evaluation, once the approach has been published, will determine exactly how this change could be applied to standard IPHC calculations.

Consideration of model-based Catch-Per-Unit-Effort (CPUE) indices

As has been discussed for several years, the calculation of commercial Pacific halibut fishery CPUE is currently very simple, treating each set with a logbook record as an independent replicate within a Regulatory Area (Stewart 2017). Further, because fishing power is estimated to differ among fixed snap and autoline gear, only fixed gear is included in the standardization for areas 2C-4CDE. Recently completed work by Cole Monnahan (University of Washington, graduated June, 2017), following on an initial analysis in 2015 (Monnahan and Stewart 2015), suggests that after accounting for spatial processes, it may be possible to include all gear types in a model-base standardization. This would allow for the inclusion of approximately 22% of the data that is currently unused in the fixed-gear only calculations. Although preliminary work did not suggest large differences in trends among areas based on the larger data set (Monnahan and Stewart 2015), it would be very desirable from the fishery's perspective to include all possible logbook records in stock assessment analyses. Further, if the Management Strategy Evaluation (MSE) compares the use of commercial CPUE-based harvest control rule or distribution approaches, standardization of these data could be particularly important.

A space-time (geospatial) model could be developed for use in standardizing these data similar to that used for the setline survey. However, an important difference between the datasets is the spatial distribution of the observations, particularly the nonrandom nature of fishery data. Recent work has explored methods to extend space-time (geospatial) models to account for fishery targeting of increased CPUE (Diggle 2010, Thorson et al. 2016).

Another alternative to a fully model-based standardization is to use the coefficients estimated for snap, fixed, and autoline gear to 'calibrate' the data prior to standardization, but retain a simple mean and variance approach within each area. The benefit of this method would be increased simplicity and transparency for the commercial fishery, while still including more of the information. However, the spatial effects would be unaccounted for and could contribute to bias in the time series, and an underestimate of the variance (Cole Monnahan, pending PhD thesis).

Based on discussion during SRB10, the IPHC secretariat has not made any additional change to the commercial Pacific halibut fishery standardization methods for the 2017 stock assessment.

Reporting of 2A CPUE

During the 2016-2017 annual process there was a considerable interest in better understanding the CPUE trend information and estimation approach for Regulatory Area 2A. The IPHC responded with a delineation of tribal and non-tribal catch rates, identifying significantly differing recent trends in each. This exploration revealed the potential importance of how these two pieces of information are weighted, both for presentation and use is assessment-related analyses (See Appendix A for comparisons). Currently, there is no explicit weighting, as the CPUE is calculated via the sum of catch and sum of effort in reported logbooks (Stewart 2017), so if reporting rates differ between the tribal and non-tribal fisheries the trend may be affected.

As discussed during SRB10, there are several potential avenues for improving the use and reporting of these data: 1) continue using the *status quo* approach, but acknowledge the differing sources of trend information, 2) revise the CPUE approach to be catch-weighted rather than area-weighted, which would allow for easy stratification of these (tribal and non-tribal) components (and potentially others in other Regulatory Areas), and 3) move toward a spatial model for CPUE (similar to that used for survey data) and following the methods of Monnahan et al. (In review). The IPHC secretariat has opted for the first of these options for 2017, clearly delineating the differing trends in the tribal and non-tribal catch rates in IPHC Regulatory Area 2A, but not altering the simple method for aggregating CPUE. This leads to a CPUE series for 2A which is globally somewhat consistent between the two fisheries, but has shown markedly different trends in the most recent few years (Appendix C). Based on the request from the SRB in June 2017 (SRB10-Req.01 (para. 18)) an additional plot of tribal vs. non-tribal commercial Pacific halibut fishery CPUE for IPHC Regulatory Area 2A was produced:

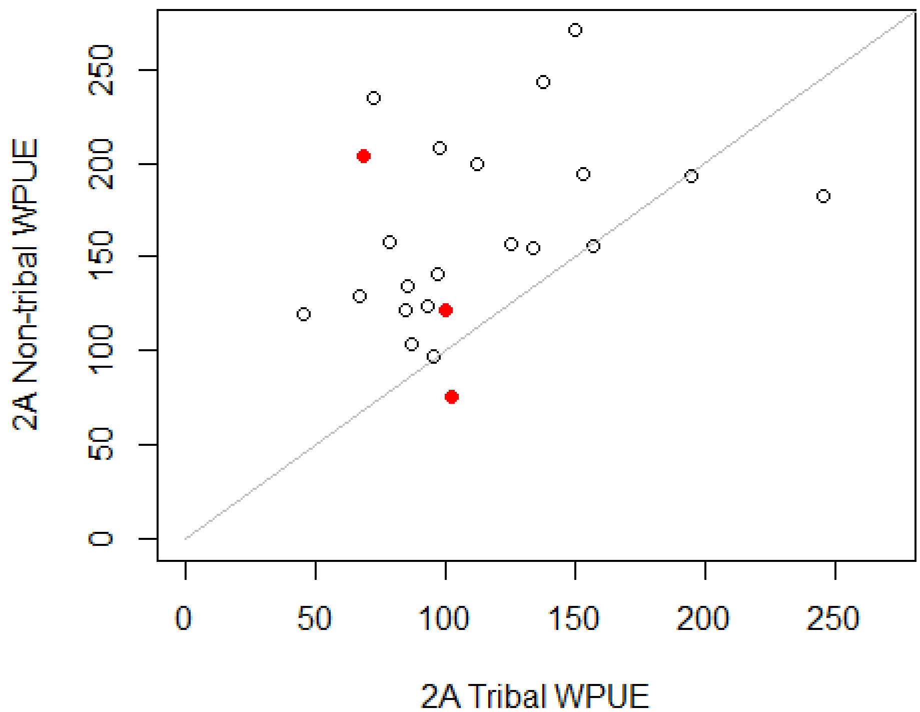

Figure 1. Non-tribal commercial WPUE vs. tribal WPUE (1989-2016) in IPHC Regulatory Area 2A. The most recent three years available (2014-2016) are highlighted in red; the grey line indicates a 1:1 relationship.

Updating the 2016 commercial Pacific halibut fishery CPUE

Each year the IPHC secretariat updates the stock assessment with the most recent year's data, and also updates any data sets from the previous year that were incomplete at the time the assessment was produced (data generally closed on or about 7 November, depending on the ending date of the commercial fishery). Historically, there have sometimes been significant revisions to the commercial fishery CPUE series as additional logbook data from late in the season and remote areas is never included in the current year's assessment data. For this reason, it was decided during previous SRB reviews to double the estimated variance on the terminal year's commercial fishery CPUE, such that incomplete data would not have an undue influence on the assessment results (Stewart and Martell 2016). The pattern for 2016 was similar to previous years, with a small decrease in estimate CPUE in the final data sets relative to those available in November (Appendix C). This led to a revised trend of -2% coastwide, with the largest decrease in a single IPHC Regulatory Area of 9% in IPHC Regulatory Area 2B, and the largest increase of 34% in IPHC Regulatory Area 2A. The 2A trend was actually identified and reconciled prior to the IPHC's 2017 Annual Meeting. Previous comparisons have shown little effect of changes of this magnitude; therefore stock assessment models were not re-run at this time, however the 2017 stock assessment will include the most complete commercial CPUE series available as of November 2017.

Further data updates

There was insufficient time prior to the production of this document to generate preliminary IPHC fishery-independent setline survey results for 2017 and introduce those results to the stock assessment models. However, preliminary evaluation of the survey catch rates suggests trends are likely to be very consistent with projections made using the 2016 stock assessment models. Similarly, final estimated removals for 2016 and preliminary values for 2017 were also not yet available; however, there is no indication to date that removals will change appreciable for 2016 relative to those used in the 2016 stock assessment, nor that 2017 removals will differ greatly from those projected based on adopted IPHC catch limits.

MODEL CODE UPDATE

Recent Pacific halibut stock assessment models have used stock synthesis (Methot Jr and Wetzel 2013) version 3.24u. For the features that are currently included in these models there are no identified bugs that have required updating the IPHC's application to a newer version. However, a more recent release of Stock Synthesis is now available (3.30.07.01 as of 7 August, 2017). This version is the 7th since the formal release. For efficiency, both the MSE and stock assessment development for the 2017-2018 IPHC annual process will be conducted in version 3.24u, with the expectation that version 3.3 will be stable enough to allow both efforts to use the newer code for the 2018-2019 process.

This change will allow exploration of some newer features that may be of particular interest to the Pacific halibut assessment, specifically: alternative error distributions for age-composition data, more flexibility in the treatment of constraints on time-varying processes, age-based discard and retention schedules, as well as others. These changes are likely to be evaluated for the June 2018 SRB meeting.

ADDITIONAL DEVELOPMENT AVENUES

Model weighting methods

Following on several years of analysis of alternative methods for weighting the component models included in the IPHC's stock assessment ensemble, three primary approaches have been identified: 1) weighting based on relative fit to the coastwide setline survey index of abundance, 2) weighting based on the predictive skill of each model relative to the terminal survey index observation, and 3) weighting based on retrospective performance over recent terminal spawning biomass estimates (these were last discussed in detail at the June 2016 SRB-8). Weights based on both of these approaches were compared in June 2016, with neither suggesting appreciably different weighting than the *status quo* approach of equal weighting of each of the four models. At that time, no formal change was made to the weighting approach, but it was recognized that these alternative weights should be routinely compared. A manuscript reporting the approach for both types of weighting is in preparation.

The weighting based on retrospective behavior has suggested the greatest difference from the *status quo* assumption of equal weights in previous analyses. Therefore, this was the focus of updated weights for 2017. Briefly, this calculation is based on the simulation analyses of Hurtado et al. (2014) exploring the probability of generating retrospective patterns by chance (when they were not actually present) across a range of stock assessment configurations. That analysis produced a distribution for rho, the strength of the retrospective pattern, which can be considered a null distribution against which the likelihood of observed values can be compared. Then, based on likelihood theory for model averaging (Burnham and Anderson 2002), the observed values for rho can be converted into model weights (Appendix D). The result of this approach, based on the 2016 stock assessment results, were weights for the four assessment models that ranged from 23.9% to 25.9%. Therefore, no change to equal weighting is proposed for 2017. As long as the results continue to be generally consistent with equal weighting, there is no pressing need for a change in methodology; however, if/when additional models are added to the ensemble and/or performance-based approaches suggest differing model weights this topic may need to be revisited.

Bayesian integration

There have been several recent developments toward alternative Markov Chain Monte-Carlo (MCMC) search algorithms implemented in Automatic Differentiation Model Builder (ADMB; Monnahan et al. 2016), which is the underlying code for the stock synthesis model. In addition to these new tools, research into approaches for regularization, adding informative priors and/or reducing complexity to improve performance, has also moved forward, using the short coastwide model as one of a set of examples (Cole Monnahan, pending PhD Thesis). This work, in concert with some of the options available in the new stock synthesis version (see section above) may provide avenues for improved efficiency in implementing Pacific halibut models in a fully Bayesian framework. Preliminary results suggested similar stock size estimates even after regularizing for more efficient Bayesian integration. This continues to be an avenue for future work, as true probability distributions (rather than asymptotic approximations) would be desirable for calculating probabilistic management results.

Spatial model development

During 2016 there was a substantial amount of review during SRB08 and SRB09 of the initial work toward creating a spatially explicit Pacific halibut model. That effort focused on fitting to various disaggregated datasets at the regional level, including the NMFS trawl survey data from Alaska which has not been included in coastwide models to date. Based on model performance and SRB guidance on those efforts, spatial model development has been refocused toward creating a spatial framework for MSE development that emphasizes hypothesis testing over statistical fitting for tactical management use. Based on the priorities identified by the Management Strategy Advisory Board for 2017 and 2018, further development of the spatially explicit model has been put on hold until coastwide objectives have been addressed by that process.

Ensemble stability

A potential and relatively unexplored benefit of using a stock assessment ensemble is inter-annual stability in stock assessment results. Stability may be created via two avenues: the inclusion of alternative models in each year’s analysis (rather than through periodic changes to a single-base case model), and from the buffering effect of characterizing the central tendency (or distribution) with a set of models. However, this last benefit – temporal stability against the addition of new data – although logically appealing, has not been explored for fisheries stock assessment. The IPHC secretariat has developed a draft manuscript for submission in 2017 that evaluates the results from the International Pacific Halibut Commission’s stock assessment ensemble as an example of the stability created by multiple-models and provides a simple simulation to explore general ensemble behavior. Counter-intuitively, we found little stability benefit in the IPHC’s current ensemble due to high temporal correlations among individual models as annual data are added. However, we also found that even a small number of models with low among-model correlations could have a substantial stability benefit. We suggest that among-model temporal correlations may be a useful diagnostic and consideration for analysts developing ensembles, or those performing sensitivity testing of single-model assessments. The secretariat will present this work more fully to the SRB in future meetings.

SUMMARY

As has been the practice for all recent stock assessments, any further change to data sources and/or model structure that is necessary after the SRB11 meeting in September 2017 will be documented and presented as part of the final stock assessment for 2017. Any questions and/or clarifications will be provided for the SRB during the annual conference call held sometime in December 2017 (after the IPHC's Interim Meeting IM093, and before the IPHC's Annual Meeting AM094).

Much of the assessment focus during 2017 has been on harvest policy related analyses. However, continued refinement of the data sources feeding in to the stock assessment models, the harvest policy analyses, and the management structure remains a priority. A number of improvements have been evaluated, with progress on most and several now completed for use during the 2017-18 management cycle (Table 2).

TABLE 2. Summary of improvements.

Improvement	Rationale	Timeline
TCEY-based management	Requested by commissioners; more transparent comparison of catch limits among Regulatory Areas.	To be discussed at the 2017 September Work Meeting, IM, and 2018 AM.
Measured individual fish weights	Data suggest systematic differences in observed vs. predicted weights from the standard length-weight relationship. Port data now include individual weights for all biological samples.	Included in preliminary 2017 stock assessment analyses.
Setline survey model-based time-series extension	Data from 1995-1997 are available for several Regulatory Areas for use in the space-time model	These data will be included in 2017 stock assessment analyses
Historical bycatch data	Re-analysis of historical data needed to reconcile observed DMRs and ensure length-frequency data have been summarized consistently.	To be included in the 2018 stock assessment analyses.
Biological data from setline survey expansion stations	Additional data from expansion stations are available for use in stock assessment analyses.	Included in preliminary 2017 stock assessment analyses.
Effective skate calculations	A revised hook-power relationship may be published in 2017.	Pending final review of these results, the revised relationship may be included in the IPHC's standard database calculations
Model-based CPUE estimates	Spatial models offer a reasonable approach to improve existing standardization methods and include a greater proportion of logbook data available.	Project on hold based on SRB discussion during SRB10.
Reporting of 2A CPUE	Important differences between trends in tribal and non-tribal catch-rates in Regulatory Area 2A have been identified	Clearer delineation of this information has been developed for the 2017 process.
Model code update	Current stock synthesis version still being tested.	Anticipated update in 2018.
Model weighting methods	Weighting approaches are important in determining ensemble results.	Evaluation of alternative methods will be continued; publication of these approaches is planned.
Bayesian integration	Better represents probability distributions for management use.	Ongoing.
Spatial model development	This level of model complexity will be required in order to evaluate some MSE objectives.	Continued development is planned for 2018.

RECOMMENDATION/S

That the SRB:

1) **NOTE** paper IPHC-2017-SRB11-06 that provided an overview of data and modelling updates, as well as a preliminary evaluation of the stock assessment ensemble proposed for use in the 2017-2018 annual process.
2) **RECOMMEND** any modifications or additions to this proposal for the 2017-2018 annual management process, and any extensions for future processes.

REFERENCES

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd Edition. Springer-Verlag, New York, 488 p.

Clark, W.G. 1991. Validation of the IPHC length-weight relationship for halibut. IPHC Report of Assessment and Research Activities 1990. p. 113-116.

Diggle, P.J. 2010. Geostatistical infrence under preferential sampling. Journal of the Royal Statistical Society Series C Applied Statistics **59**(2): 191-232.

Erikson, L.M., and Kong, T.M. 2017. 2.7 Sampling of commercial landings. IPHC Report of Assessment and Research Activities 2016: 90-100.

Hamley, J.M., and Skud, B.E. 1978. Factors affecting longline catch and effort: II Hook-spacing. International Pacific Halibut Commission Scientific Report No. 62. p. 15-24.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., and Punt, A.E. 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar. Sci. **72**(1): 99-110.

Methot Jr, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**(0): 86-99.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science: Journal du Conseil **56**(4): 473-488.

Monnahan, C.C., and Stewart, I.J. 2015. Evaluation of commercial logbook records: 1991-2013. IPHC Report of Assessment and Research Activities 2014. p. 213-220.

Monnahan, C.C., Thorson, J.T., Branch, T.A., and O'Hara, R.B. 2016. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution **8**(3): 339-348.

Planas, J.V. 2017. 1.1 2017 IPHC biological and ecosystem science research plan IPHC Report of Assessment and Research Activities 2016: 10-24.

Stewart, I.J. 2017. 4.1 Overview of data sources for the Pacific halibut stock assessment and related analyses. IPHC Report of Assessment and Research Activities 2016: 279-364.

Stewart, I.J., and Martell, S.J.D. 2016. Appendix: Development of the 2015 stock assessment. IPHC Report of Assessment and Research Activities 2015. p. A1-A146.

Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., and Winker, H. 2016. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data1. Can. J. Fish. Aquat. Sci.: 1-14.

Webster, R.A., and Erikson, L.M. 2017. 2.8 Analysis of length-weight data from commercial sampling in 2016. IPHC Report of Assessment and Research Activities 2016: 101-109.

Webster, R.A., Erikson, L.M., and MacTavish, K.A. 2016. Analysis of length-weight data from commercial sampling in 2015. IPHC Report of Assessment and Research Activities 2015. p. 77-91.

Williams, G.H. 2014. Incidental catch and mortality of Pacific halibut, 1962-2013, IPHC Report of Assessment and Research Activities 2013. p. 289-310.

Williams, G.H. 2016. Recommendations for Pacific halibut discard mortality rates in the 2016-2018 groundfish fisheries off Alaska. Int. Pac. Halibut Comm. Report of Assessment and Research Activities 2015: XX-XX.

APPENDICES

<u>Appendix A</u>: Figures illustrating the effect of replacing estimated weights with sampled weights in the 2015 and 2016 fishery data.

<u>Appendix B</u>: Comparison of stock assessment model and ensemble results for updated data sources in 2017.

<u>Appendix C</u>: Figures comparing commercial Pacific halibut fishery WPUE.

<u>Appendix D</u>: Description of the model weighting approach based on retrospective analysis using the 2016 stock assessment results.

APPENDIX A

Figures illustrating the effect of replacing estimated weights with sampled weights in the 2015 and 2016 fishery data.

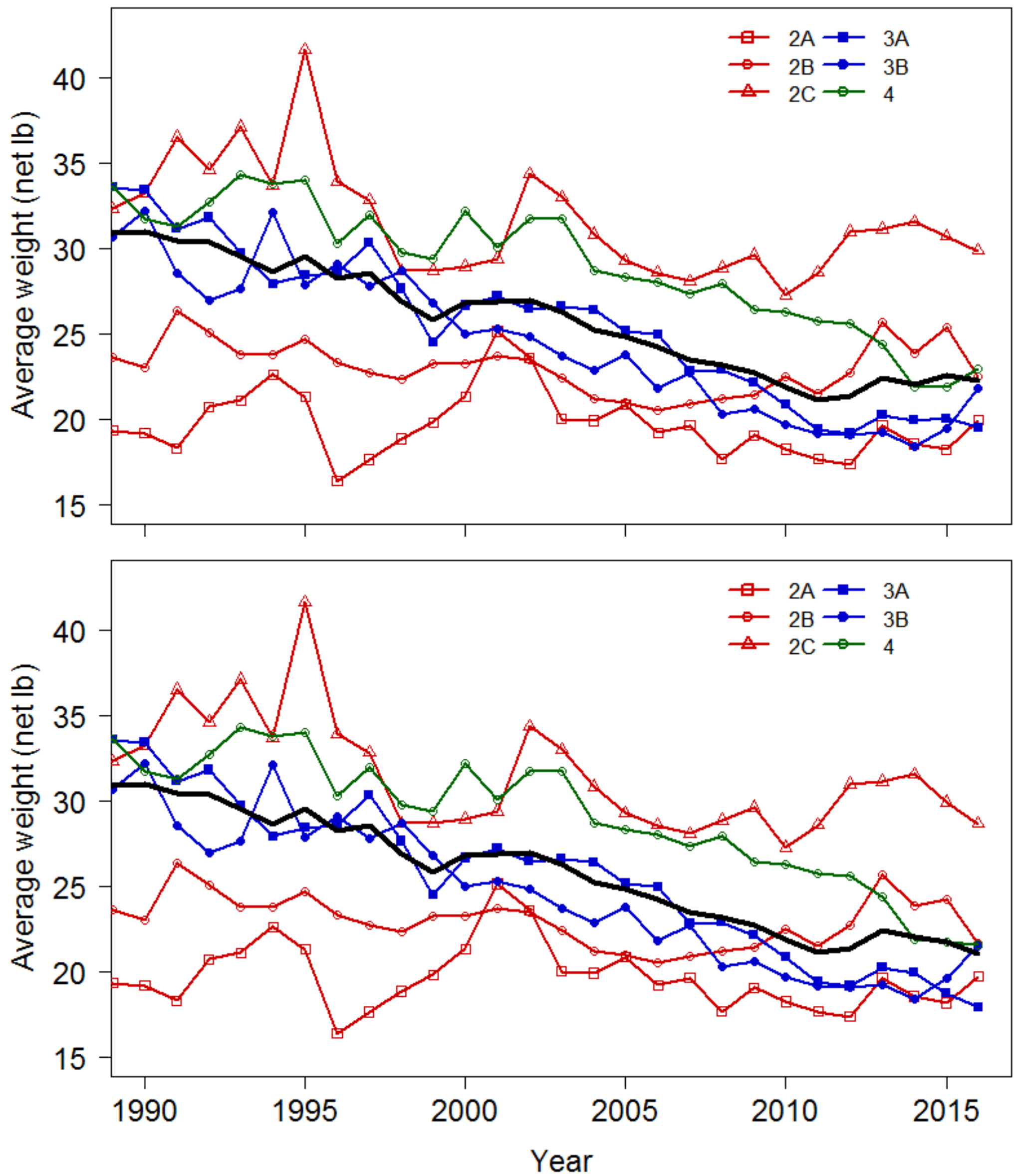

Figure A1. Commercial average landed fish weight by IPHC Regulatory Area based on estimated (upper panel) and sampled (lower panel) individual weights.

APPENDIX B

Comparison of stock assessment model and ensemble results for updated data sources in 2017.

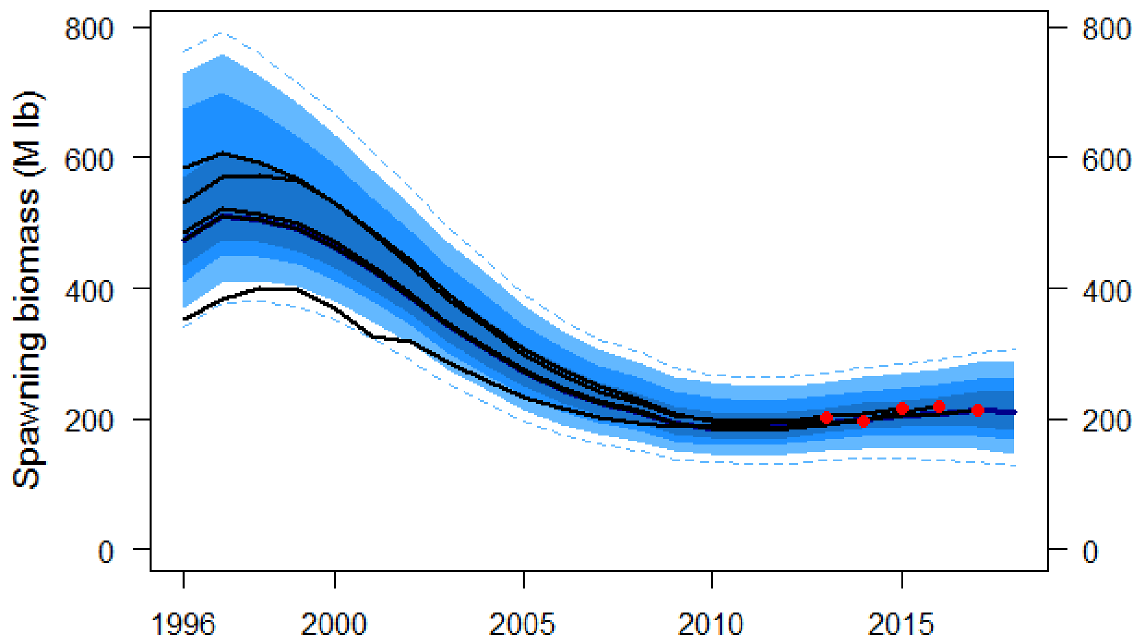

Figure B1. The 2017 stock assessment ensemble results with an additional year added to the time-series, but no change to the data sources. Catches for 2017 were assumed to be exactly those predicted for the adopted catch limits. The black lines denote point estimates from previous assessments conducted in 2012-2016, with the red markers indicating the terminal biomass estimate. The shaded area represents the approximate probability distribution from the 2017 ensemble.

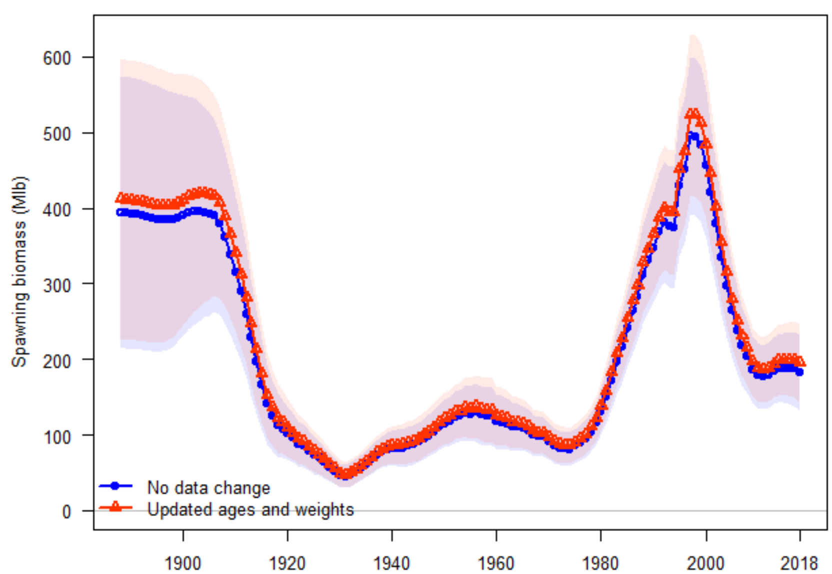

Figure B2. Cumulative change in the coastwide long time-series model due to the addition of biological data from survey expansion stations.

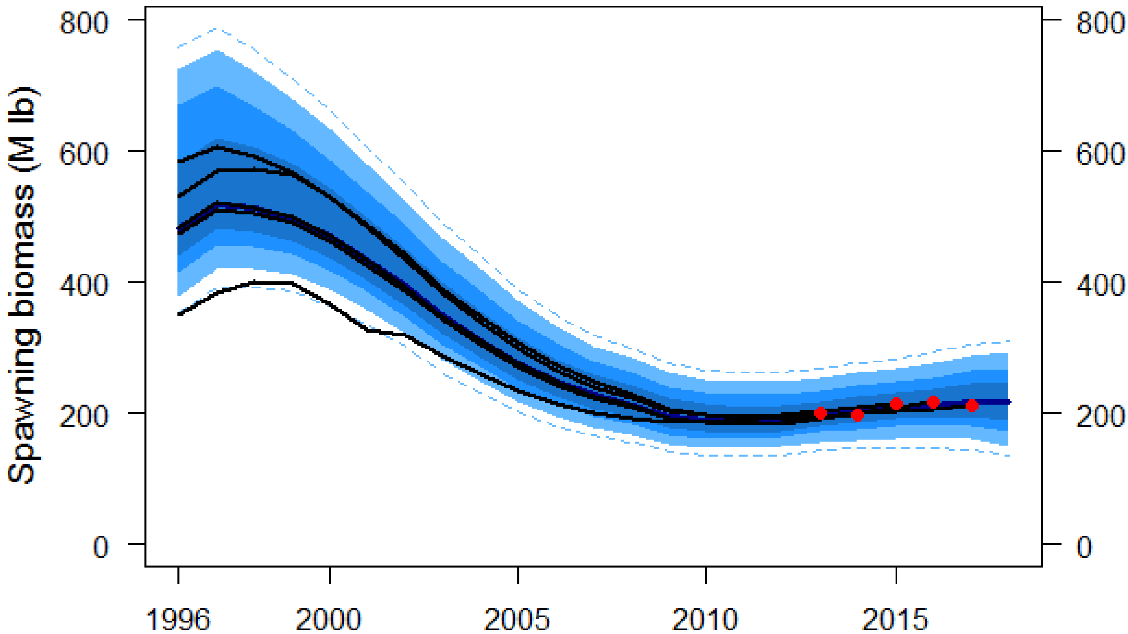

Figure B3. Cumulative change in the ensemble due to the addition of biological data from survey expansion stations (compare to figure B1).

APPENDIX C

Figures comparing commercial Pacific halibut fishery WPUE.

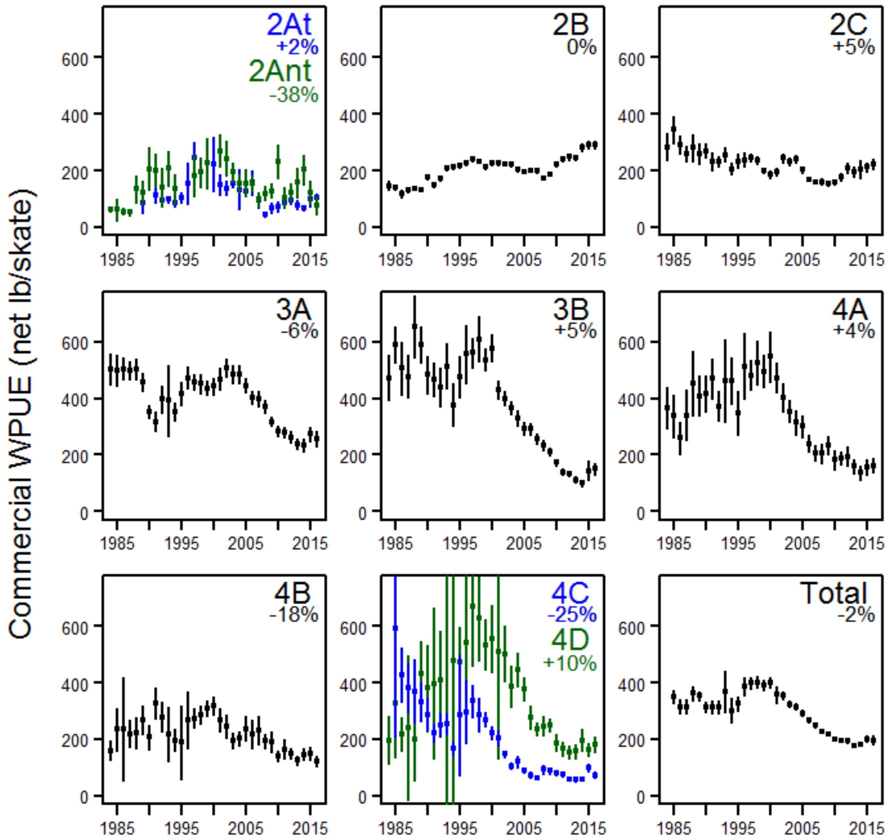

Figure C1. Commercial WPUE separated tribal (2At, blue) and non-tribal (2Ant, green) time series in IPHC Regulatory Area 2A. Percentages indicate the change from 2015-2016; vertical bars an approximate 95% confidence interval based only on between-set variability. Other IPHC Regulatory Areas are presented for comparison of the level of inter-annual and within-Area variability.

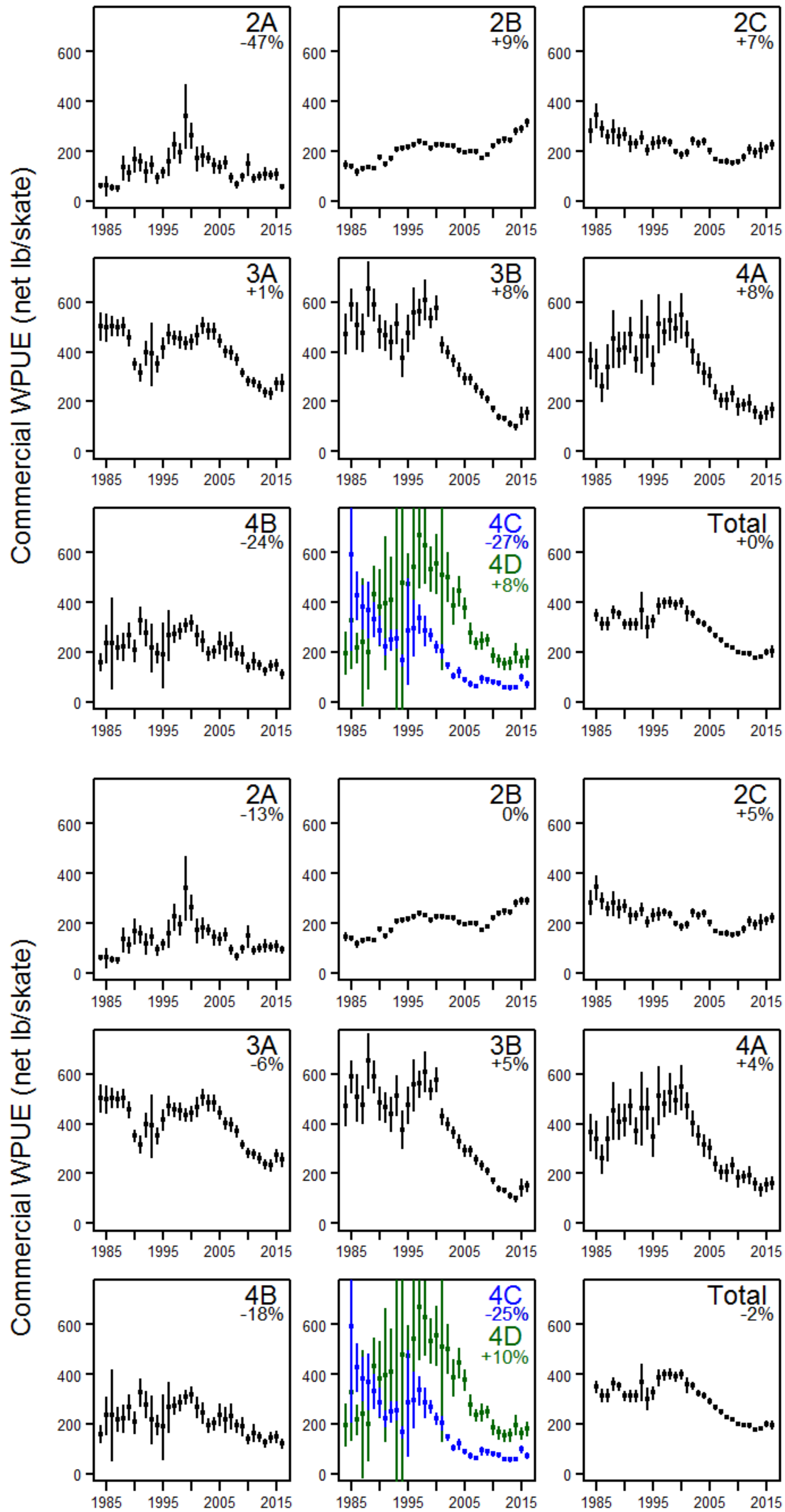

Figure C2. Commercial WPUE calculated for the 2016 stock assessment (upper panels), and updated for the 2017 stock assessment (lower panels). Percentages indicate the change from 2015-2016; vertical bars an approximate 95% confidence interval based only on between-set variability. Changes reflect additional data entered and verified after completion of the 2016 stock assessment.

APPENDIX D

Description of the model weighting approach based on retrospective analysis using the 2016 stock assessment results.

The method for model weighting based on retrospective performance was described during SRB08 in June 2016. The approach is based on Mohn's rho (ρ), which is primary metric for evaluating the degree of pattern observed in retrospective analyses: sequentially removing annual data from the terminal year of a stock assessment and re-estimating the time series (Mohn 1999). A slightly modified version of Mohn's rho (to make the number of years unimportant to the absolute value of the metric) was introduced by Hurtado et al. (2014). That analysis suggested the values approximately > 0.2 and < -0.2 were unlikely to be generated by random chance across a range of stock assessment configurations. This information was used to generate a null likelihood distribution for rho ($\tilde{\rho}$) against which to compare the observed values for rho for each of the four assessment models (*i*) averaged over the most recent three years:

$$NLL_i = \frac{(\tilde{\rho} - \rho_i)^2}{2{\sigma_{\tilde{\rho}}}^2} + \ln(\sigma_{\tilde{\rho}})$$

Each model's negative log likelihood (*NLL*) was then compared to the best model (with the lowest average rho):

$$\Delta_i = 2 \cdot NLL_i - 2 \cdot NLL_{min}$$

Finally, these relative differences were turned into model weights (w_i) normalized to 100%:

$$w_i \propto P(\rho_i | M_i) = e^{(-0.5\Delta_i)}$$

All four stock assessment models had an average rho over 2013-2015 retrospective comparisons very close to zero (Figure D1).

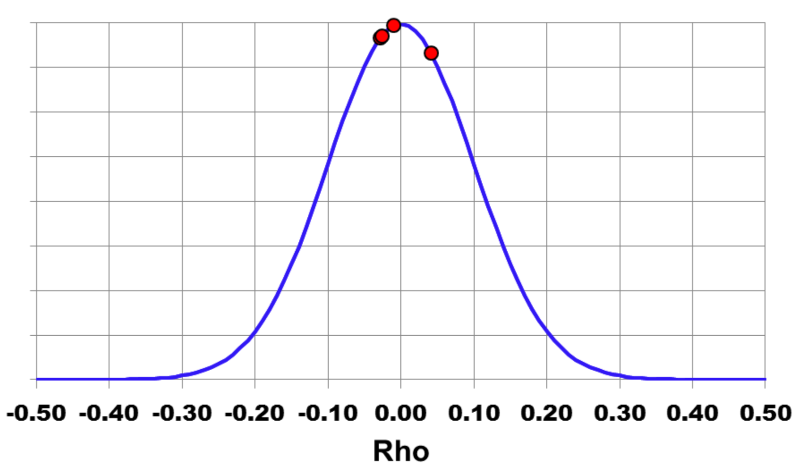

Figure D1. Approximate distribution of simulated value for rho (describing the degree of retrospective pattern in recent stock assessment results) from Hurtado et al. (2014). Points indicate the four values for rho estimated from each of the 2016 stock assessment models contributing to the ensemble.

Based on the small estimated values for rho for all four models, the implied weights are almost identical to the *status quo* (equal weighting, or 25% for each of four models) used in recent ensembles (Figure D2).

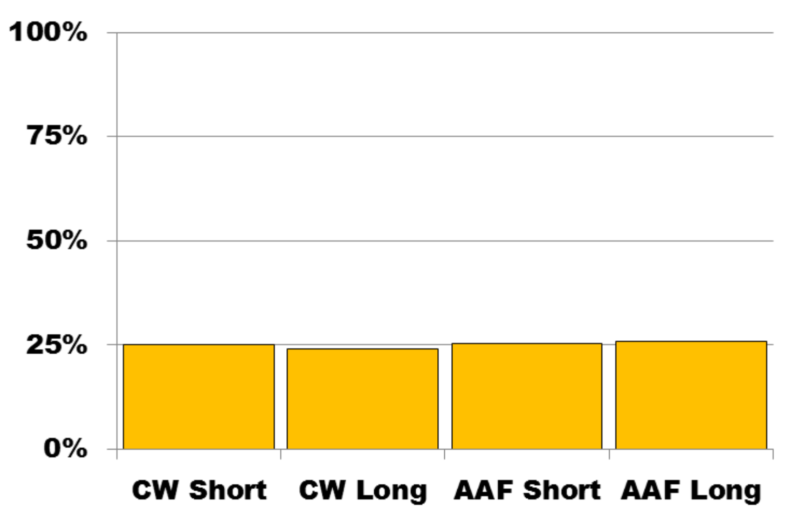

Figure D2. Model weights implied by the estimate rho values from the 2016 stock assessment.