

INTERNATIONAL PACIFIC



HALIBUT COMMISSION

# Development of the 2022 stock assessment

Agenda item: 6

IPHC-2022-SRB020-07

(I. Stewart)



# Outline (1)

- IPHC process and recent SRB requests
- Data
  - Data sources included
  - External information on M
  - Age data
  - Marine mammal depredation
- Modelling
  - Multi-model approach
  - Structural assumptions
  - Technical configuration
  - Changes from 2021
  - Diagnostics and results



# Outline (2)

- Evaluation of uncertainty
  - Sensitivity analyses
  - Likelihood profiles
  - Retrospective analyses
  - Other considerations
- Ensemble
  - Methods
  - Weighting based on predictive skill
  - Preliminary results for 2022
- Research priorities and future development



# IPHC Assessment and review process

- Annual assessments
  - Full analyses ~ every 3 years (2015, 2019, 2022)
  - Updates in between
    - Include all new data available and limited model changes
- Annual improvements reviewed by the SRB in June
- Refined and finalized for the September SRB
- Final data added in early November
- Results presented at the Interim Meeting

[Stock assessment web-page](#)



# Documentation

- The assessment document ([IPHC-2022-SRB020-07](#))
- Additional files (Appendix A):
  - Input files for each model
  - Output files and graphics for each model
  - Software documentation
  - Recent data overview and stock assessments
  - Relevant manuscripts
  - Full history of assessment and review ([assessment web site](#))



# SRB requests (SRB019)

Para. 31:

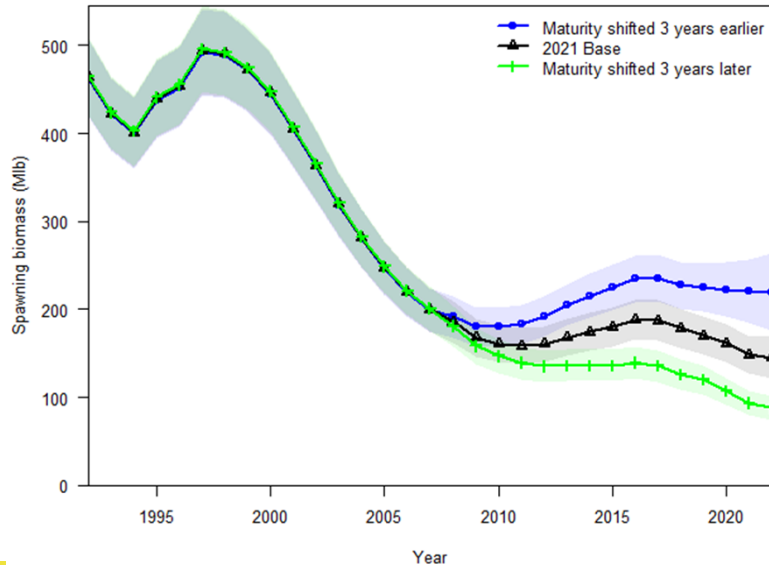
*The SRB **REQUESTED** that the IPHC Secretariat consider the following topics for inclusion in the 2022 full stock assessment and presentation for SRB evaluation at SRB020 in June 2022:*

- a) Sensitivity analysis of the assessment to processes being investigated by the Biological and Ecosystem Research Program, e.g. spatiotemporal differences in maturity schedules, discard mortality, and length-weight relationships;
- b) Continued exploration of data weighting;
- c) Evaluation of treatment of commercial sex ratio;
- d) Use of the Pacific Decadal Oscillation (PDO) and other environmental covariates to predict recruitment;
- e) Estimation of whale depredation mortality for potential explicit inclusion in the assessment model; and
- f) Other factors discussed since the last stock assessment.



# SRB requests (SRB019)

- a) Sensitivity analysis of the assessment to processes being investigated by the Biological and Ecosystem Research Program, e.g. spatiotemporal differences in maturity schedules, discard mortality, and length-weight relationships;
- Sensitivity to depredation mortality, temporal change in maturity, PDO relationship included in 2021 assessment ([IPHC-2022-SA-01](#))



# SRB requests (SRB019)

b) Continued exploration of data weighting;

- Bootstrapping of the raw age data to better inform the relative data weights among sources and across years





# SRB requests (SRB019)

## c) Evaluation of treatment of commercial sex ratio;

With 4 years of sex-specific age compositions (5 will be available for the final 2022 assessment), sex-ratio is allowed to vary over time.

- 'Disconnects' recent dynamics from historical period
- Improved retrospective analyses
- Does not require additional model changes to track sex-ratio in the future



# SRB requests (SRB019)

d) Use of the Pacific Decadal Oscillation (PDO) and other environmental covariates to predict recruitment;

- Investigation of 5 methods for using the PDO as a recruitment covariate

(*Status quo* performed the best)



# SRB requests (SRB019)

- e) Estimation of whale depredation mortality for potential explicit inclusion in the assessment model; and
- Exploratory data analysis suggested relatively low additional mortality
  - Additional work to be done on methods for data collection, creating incentives for reporting



# SRB requests (SRB019)

f) Other factors discussed since the last stock assessment.

- As described throughout;
- Natural mortality was identified as a critical area of uncertainty in the 2019 full assessment; therefore, this was the focus of most additional exploration for 2022

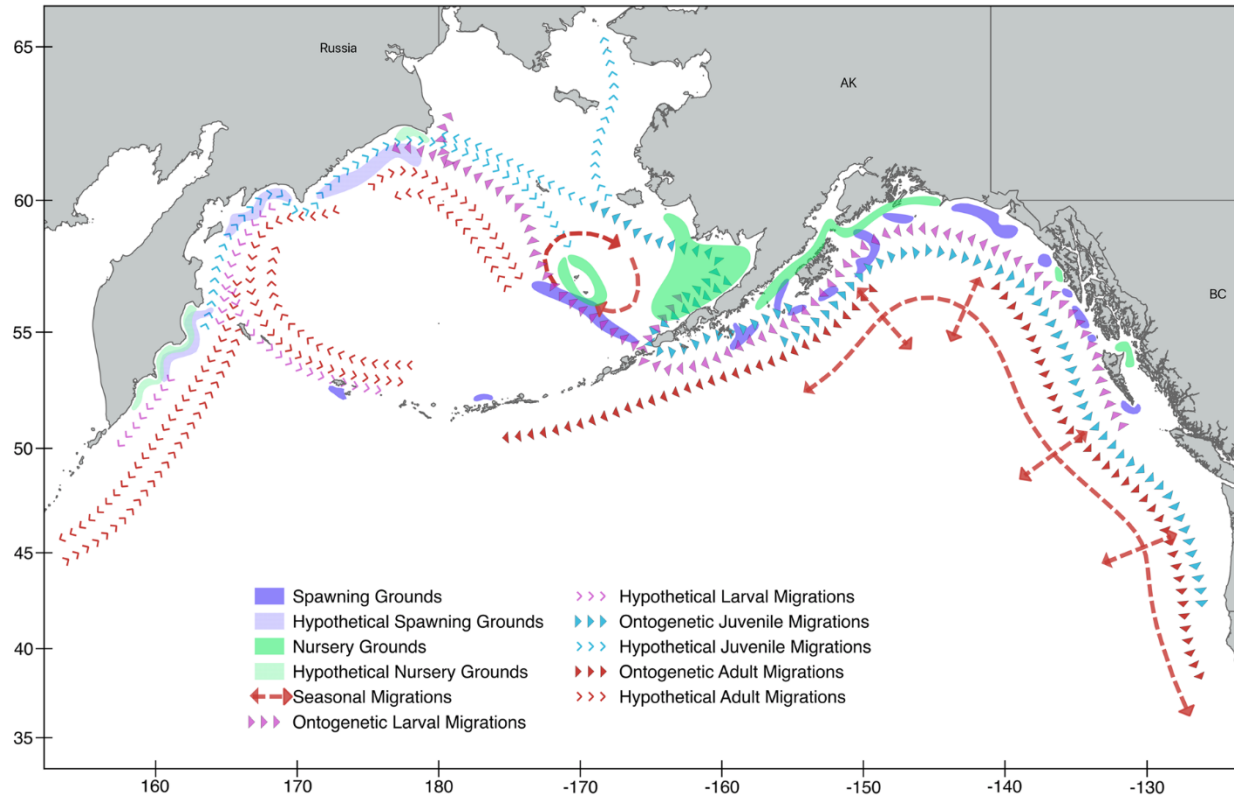


# Outline (1)

- IPHC process and recent SRB requests
- Data
  - Data sources included
  - External information on M
  - Age data
  - Marine mammal depredation
- Modelling
  - Multi-model approach
  - Structural assumptions
  - Technical configuration
  - Changes from 2021
  - Diagnostics and results



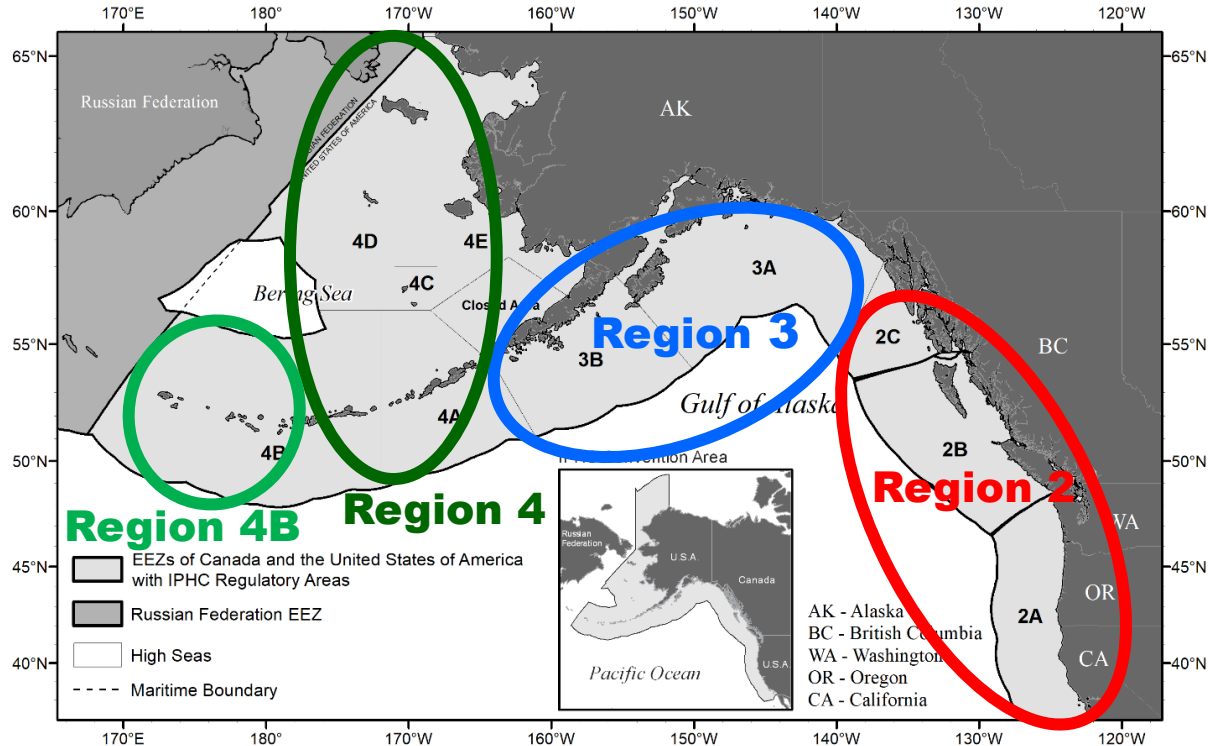
# Basic life history



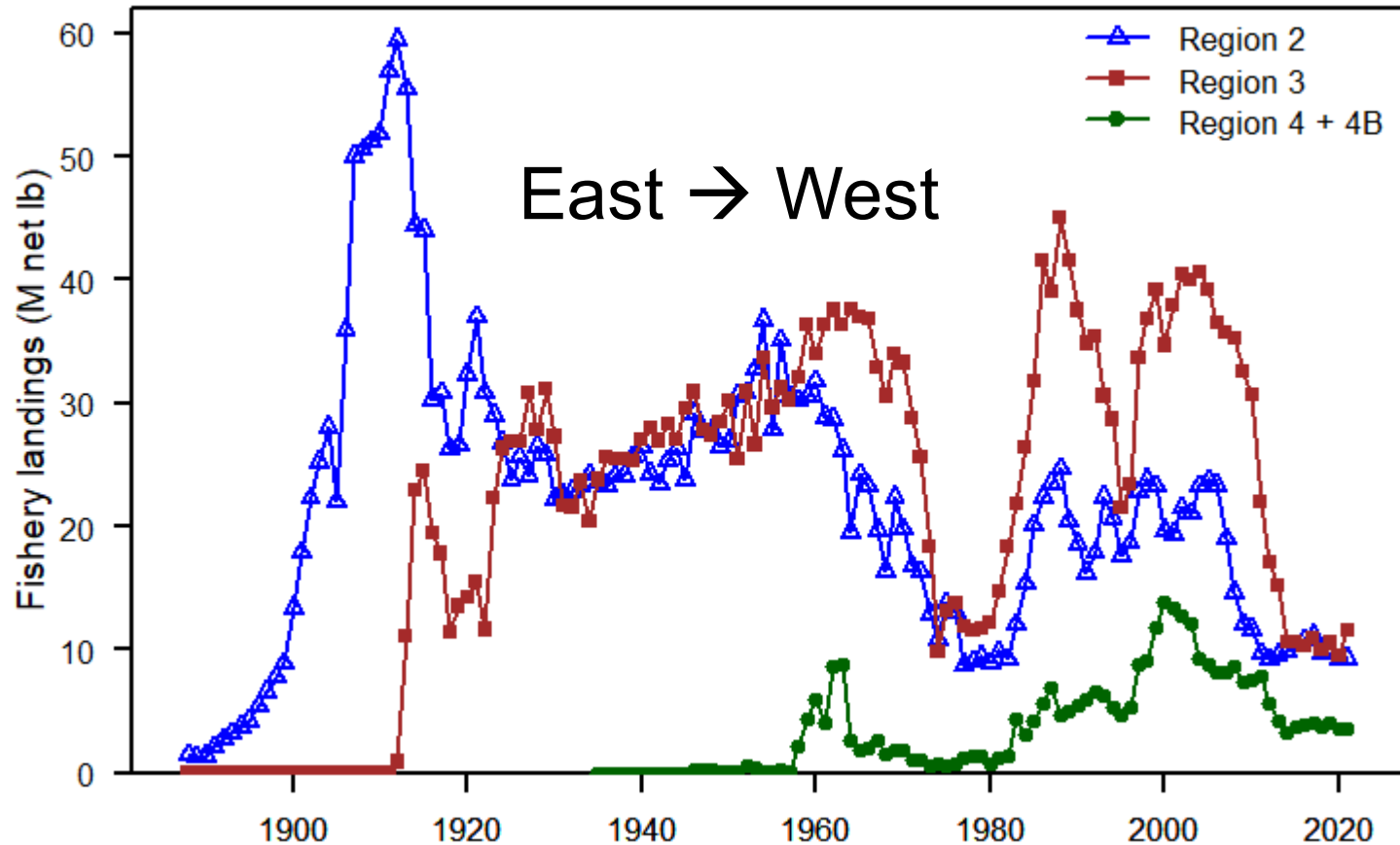
Carpi, P., Loher, T., Sadorus, L.L., Forsberg, J.E., Webster, R.A., Planas, J.V., Jasonowicz, A., Stewart, I.J., and Hicks, A.C. 2021. Ontogenetic and spawning migration of Pacific halibut: a review. *Reviews in Fish Biology and Fisheries*. doi:10.1007/s11160-021-09672-w.



# One stock, four Biological Regions

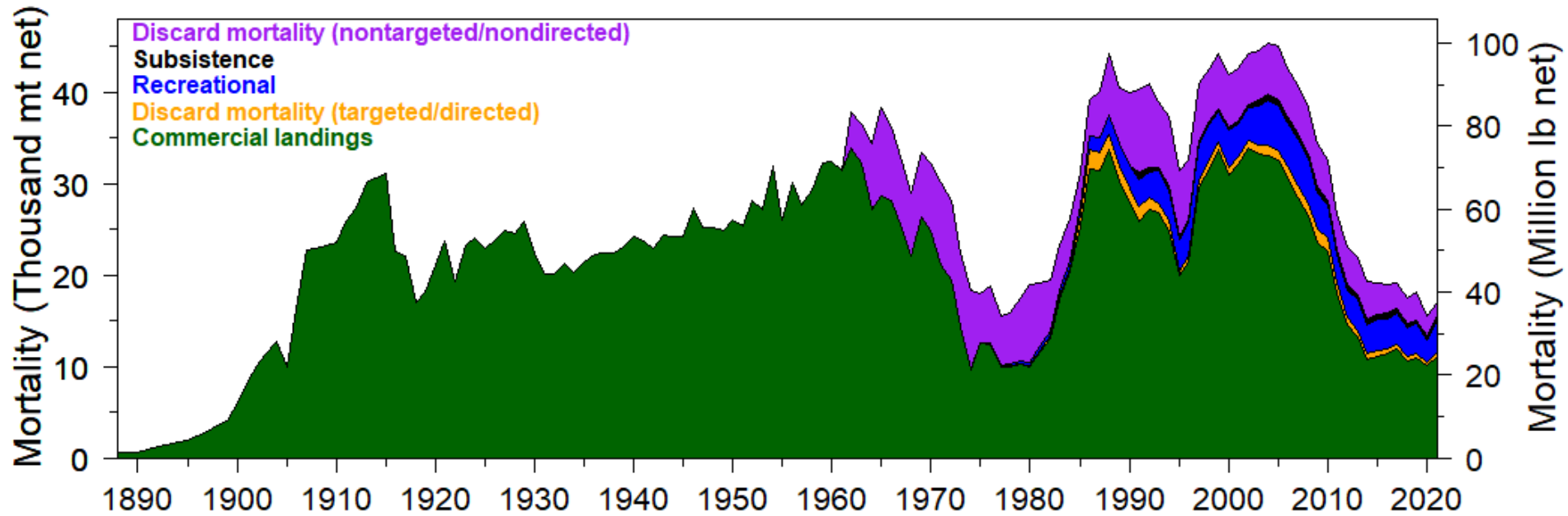


# Historical directed commercial fishery

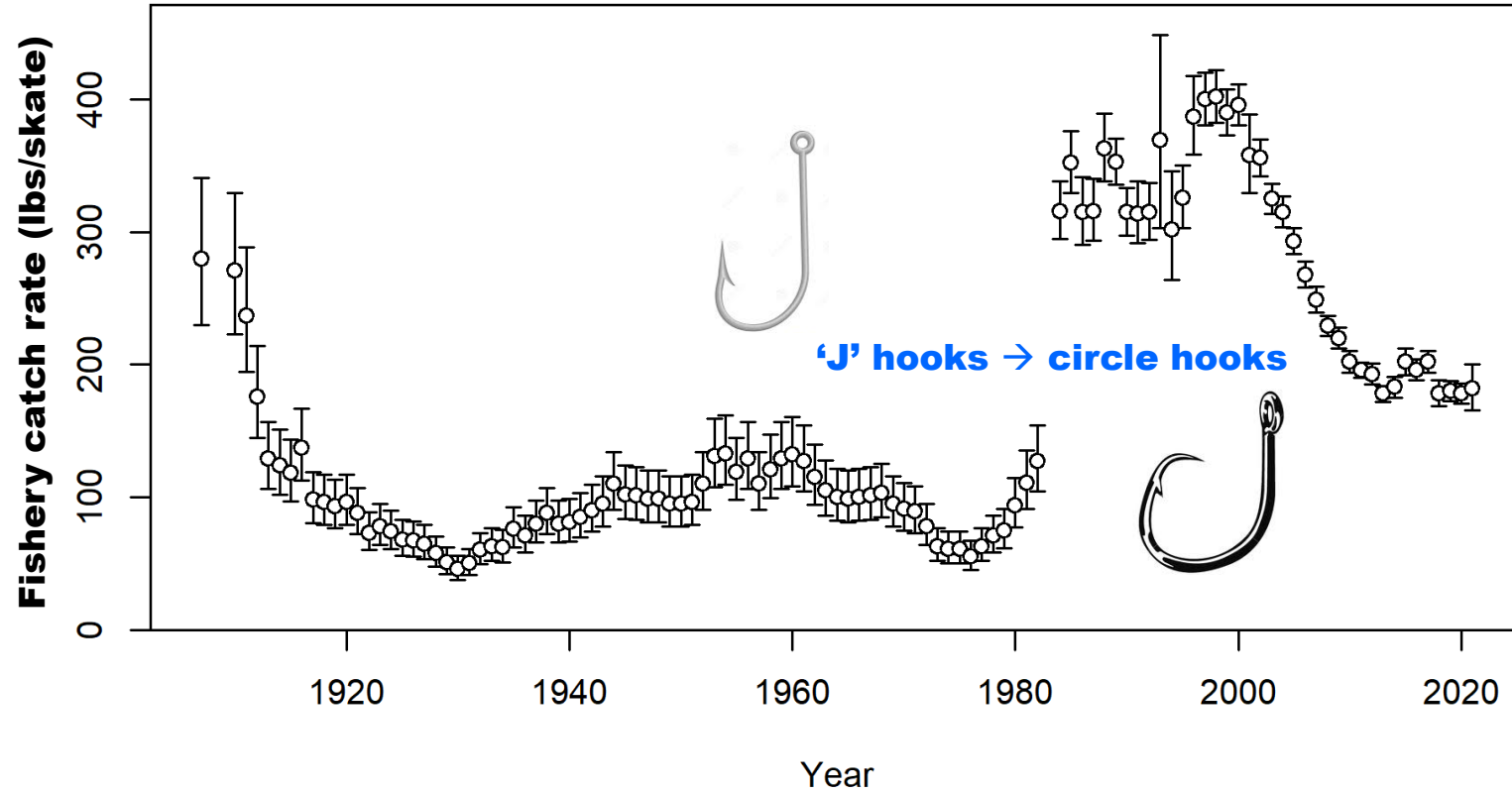




# Historical mortality by sector



# Historical CPUE



# Management notes

- 32” Commercial minimum size limit
- Commercial seasons: March-December
  - FISS: June-August
- IFQ/ITQ in AK and BC (Derby in WA/OR/CA)
- Longline and pot gear legal
  - Trawl gear must discard all halibut
- Recreational, personal use/subsistence managed differently by IPHC Regulatory Area
  - Size, bag, temporal and possession limits
- Discard mortality rates vary from 4-100% by fishery



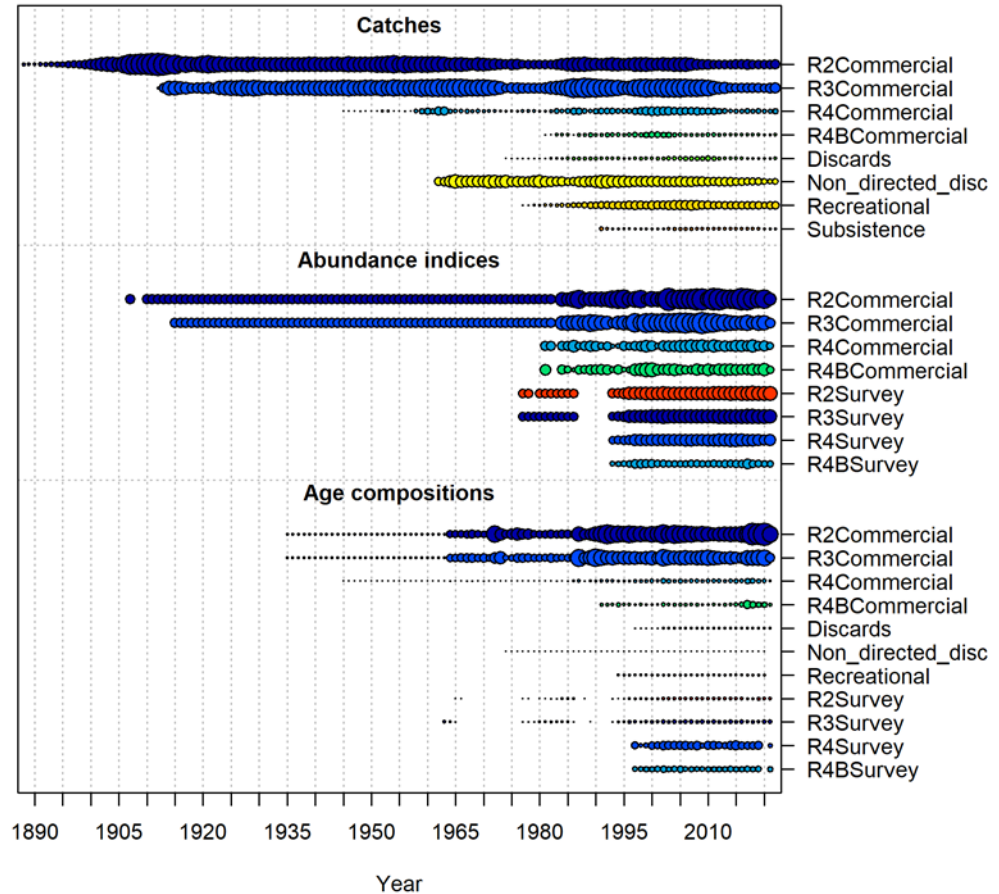
# Basic data sources

Mortality time-series' very complete

Indices: fishery only before 1992

Excellent age data: FISS + commercial

Poor information for directed discards, recreational, non-directed discards, subsistence



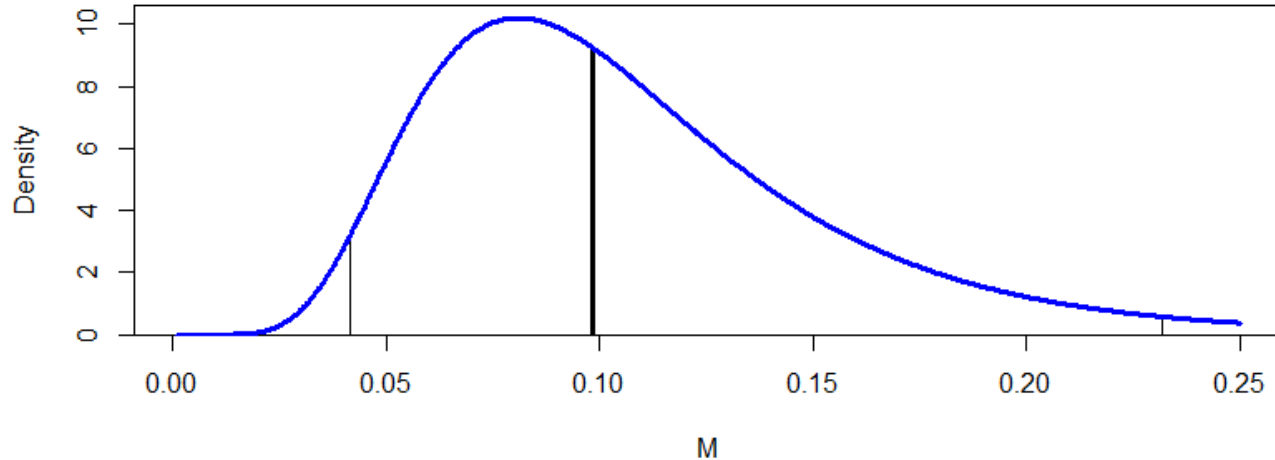
# Other data inputs (Table 8)

- PDO index
- Maturity ogive
- Fecundity information
- Weight-at-age
- Length-weight relationship
- Ageing error (bias and imprecision)
- 'Priors' on bycatch, discard and recreational selectivity



# External information on M

- Generalized prior for marine fish (Hamel 2014)



- Size-dependent information via 26 flatfish stocks in the NE Pacific (Table 9)

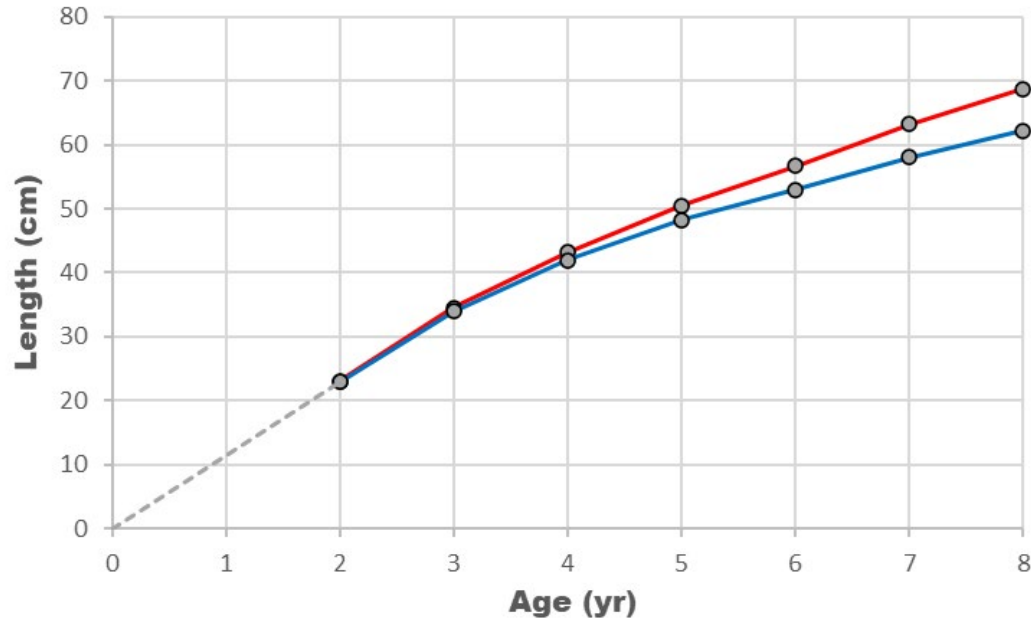


# Size/age dependent M

- CAPAM: we should expect size/age dependent M from first principles
  - What ages/sizes should this apply to?
  - How extreme?
  - Why does it matter if there is little data on these ages?



# Size/age dependent M



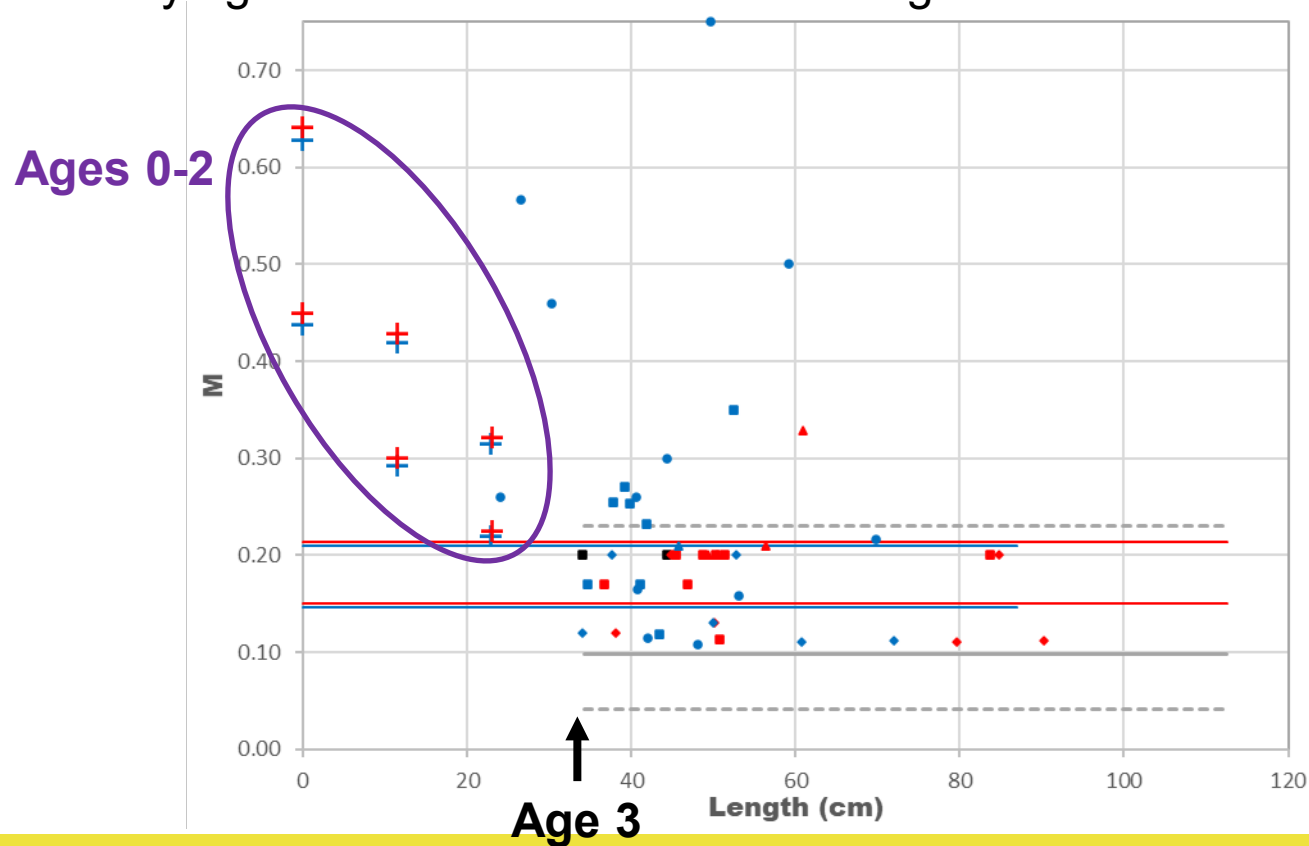
Size-at-age from trawl survey data: dimorphic growth not important at younger ages





# Ages/sizes at which M may be higher

By ages 3-5 Pacific halibut are as large as most flatfish at  $L_\infty$

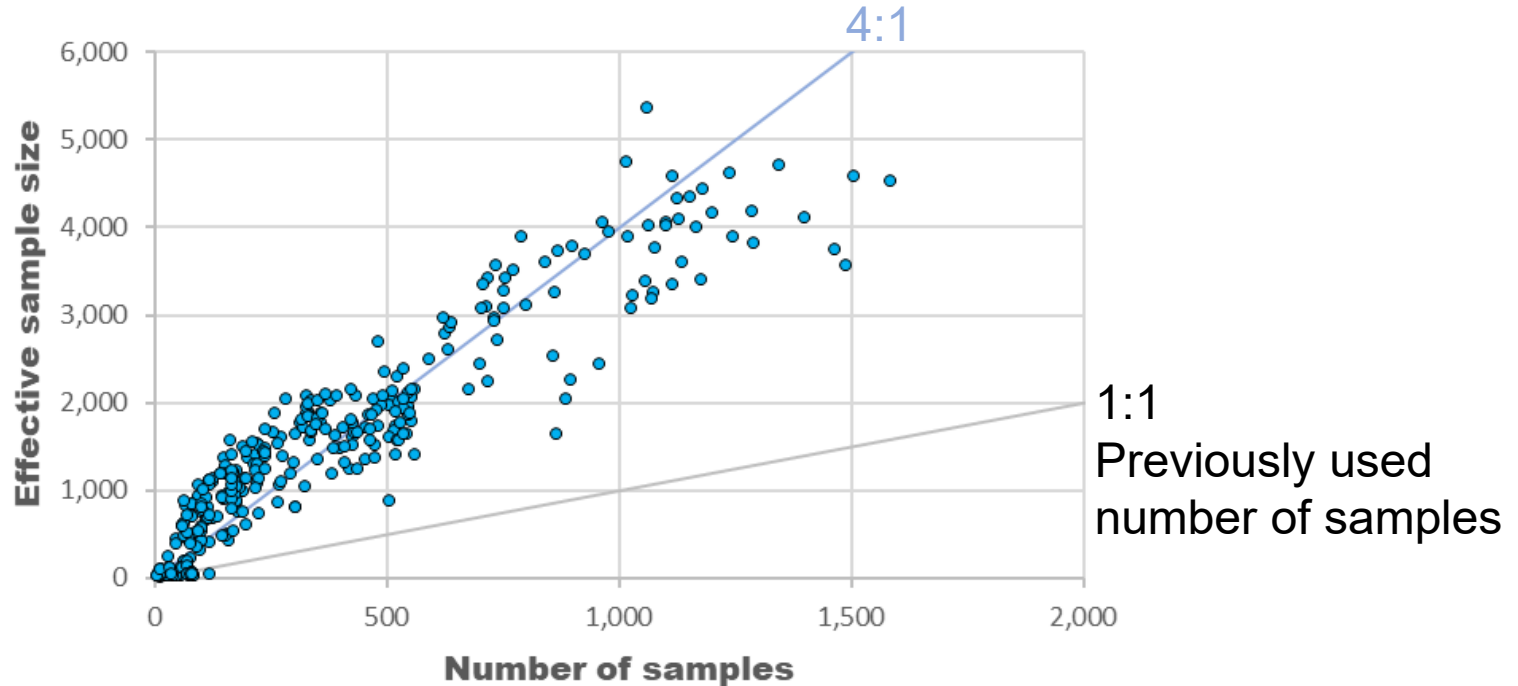


# Bootstrapping age composition sample sizes

- A quantitative estimate of the minimum variance associated with each year's specific age composition
- Does not include variance associated with missing areas
- Better than the previously-used raw number of samples (fishery trips or survey stations)



# Bootstrapping age composition sample sizes



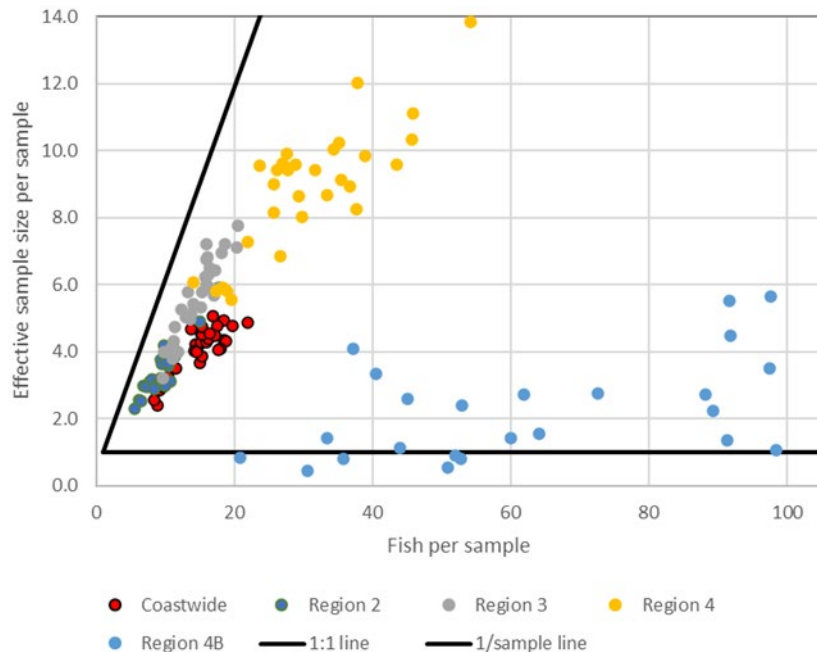
(Averages by data type provided in Table 10)



# Bootstrapping age composition sample sizes

- Generally increased maximum value used for tuning in the models (Tables 11-13)
- Also provides a basis for refinement of sampling targets for trips/stations vs. # of fish

4B stands out as an area with a high degree of clustering among samples  
→ more fish are not as informative as more samples.



# Marine mammal depredation

- Known source of ‘fishing related’ mortality
- Difficult to quantify
- Ideal solution:
  - Use survey catch rates in the space time model to estimate a ‘depredation effect’ when present  
([IPHC-2021-SRB019-05](#))
  - Use fishery data to determine the frequency of occurrence

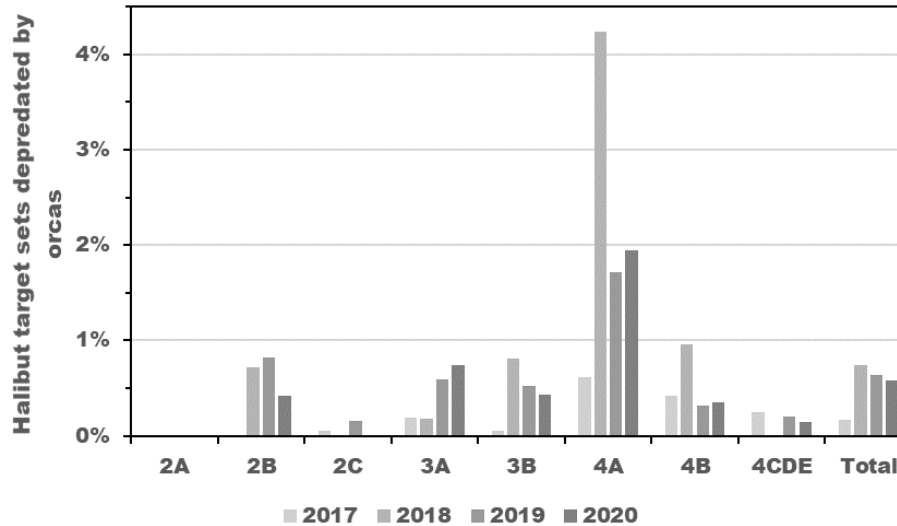


# Marine mammal depredation

- Logbook fields added in 2017
- Ongoing ‘slow rollout’ as old logs are replaced
  - Canadian logs still do not contain this information
- Participation and clear data collection protocols are still improving
- Therefore, this is a very preliminary evaluation



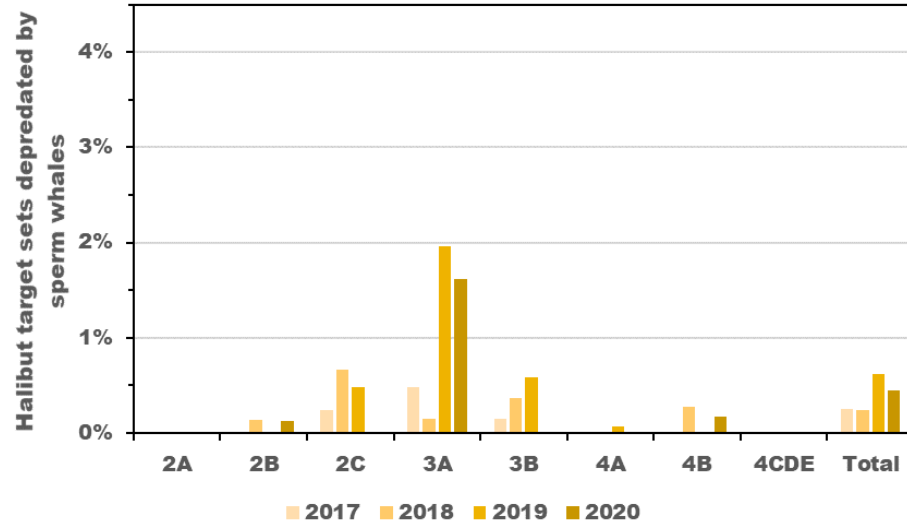
# Marine mammal depredation



Orca activity in 4A consistent with FISS observations  
Very low rates of observed whales and gear/catch damage  
Raw effect on average catch-rates similar to FISS



# Marine mammal depredation

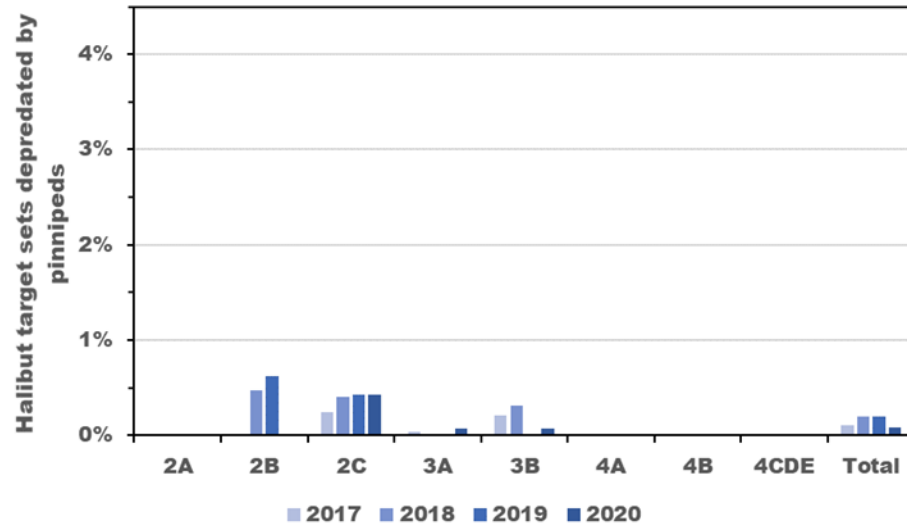


Sperm whale activity in 3A also consistent with FISS observations  
Very low rates of observed whales and gear/catch damage  
Raw effect on catch-rates highly variable.





# Marine mammal depredation



Very low rates of observed pinniped activity and gear/catch damage

Raw effect on catch-rates highly variable.



# Marine mammal depredation

- A depredation mortality estimate not yet included in the assessment
- Next steps:
  - Refine codes and collection procedures to improve accuracy and clarity of records
  - Outreach program to encourage the fleet to report should include specific ways in which the data will and will not be used
- Marine mammal interactions remain a sensitive topic for many fisheries

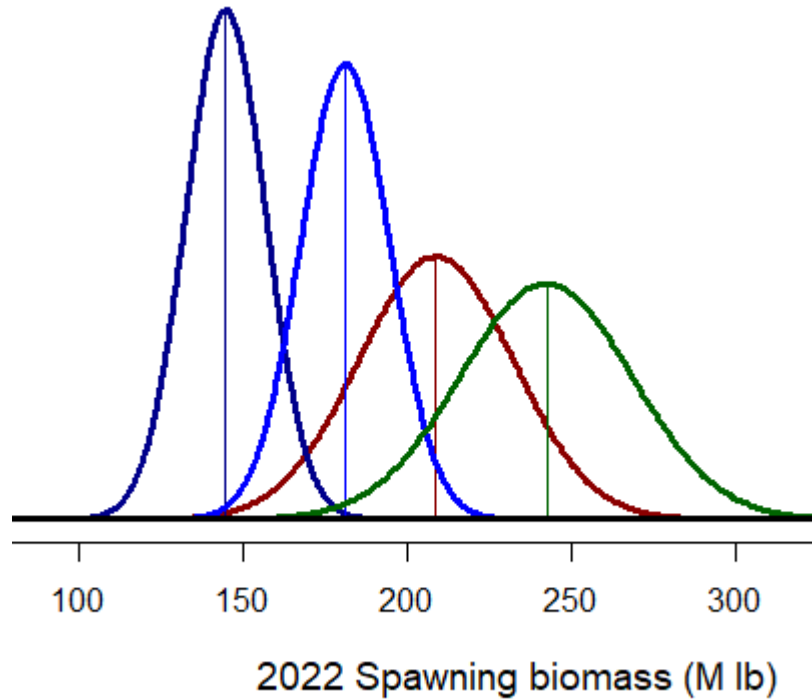


# Outline (1)

- IPHC process and recent SRB requests
- Data
  - Data sources included
  - External information on M
  - Age data
  - Marine mammal depredation
- Modelling
  - Multimodel approach
  - Structural assumptions
  - Technical configuration
  - Changes from 2021
  - Diagnostics and results



# Ensemble approach: four individual models



- Four ways to aggregate the data
- Respond differently to trend and age data by Region
- Provides stability from year to year as individual model results change



# Ensemble approach: four individual models

- 2 x 2 cross of Coastwide (CW) vs. Areas-As-Fleets (AAF) and long (1888+) vs. short (1992+) time-series
- Spans the basic types of models commonly used for NE Pacific groundfish stocks
- Each model could be a stand-alone assessment



# Basic structural assumptions (Table 14)

- Age-based models with empirical weight-at-age
- Pope's approximation for fishing mortality
- Beverton-Holt S-R relationship, with tuned annual deviations, fixed steepness (0.75), initial offset (short models) and PDO coefficient (long models)



# Basic structural assumptions (Table 14)

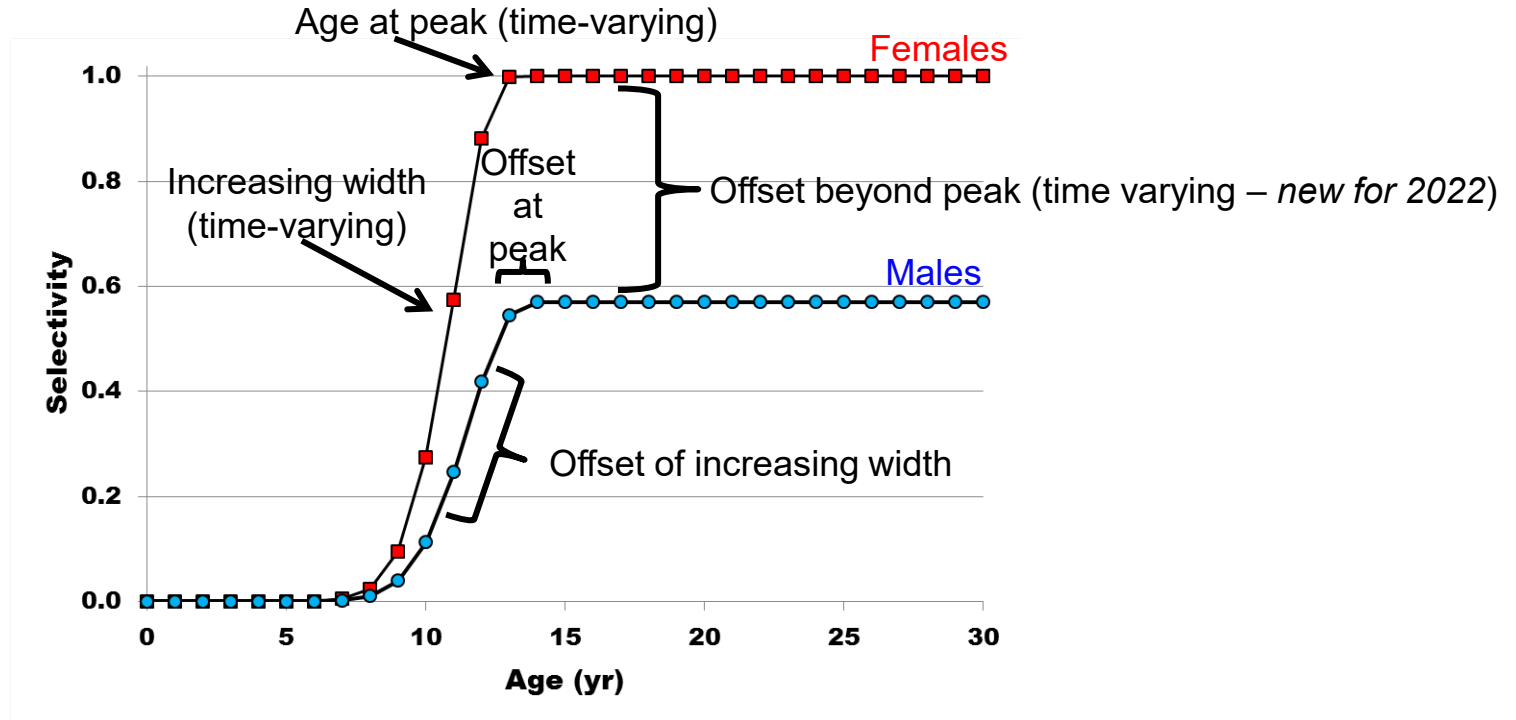
- Asymptotic (coastwide models) or domed (AAF models) selectivity
- Time-varying ascending limb, peak and scale (for males) for commercial fishery
- All process deviations tuned to reflect process and estimation variance:

$$\sigma_{tuned} \sim \sqrt{SE_{devs}^2 + \bar{\sigma}_{dev}^2}$$

- Results in internal consistency and unbiased variance (e.g., Methot and Taylor 2011)



# Treatment of time-varying selectivity





# Commercial sex-ratios

|      | Coastwide<br>% female | Region<br>2 | Region<br>3 | Region<br>4 | Region<br>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 genetic assay data will be available for the final 2022 assessment  
Rate and frequency of application to be reviewed based on 5 years completed



# Treatment of the PDO

- Estimated link coefficient ( $\beta$ ) adjusting the scale of the S-R relationship:

$$R_0' = R_0 * e^{\beta * PDO_{regime}}$$

$$R_y = f(SB_y, R_0', SB_0, h) * e^{r_y - \frac{\sigma^2}{2}}$$



# ‘Priors’ on selectivity

- Down-weighted age data used as a prior for recreational/subsistence, directed discards, and non-directed discards
  - Non-directed discards – lengths converted to ages based on trawl survey age-length keys; incomplete coverage and weighting among sectors
  - Recreational – age-data from IPHC Regulatory Area 3A only
  - Directed discards – ages from FISS sublegal catch used as a proxy for discards (comparison only available in 2B)

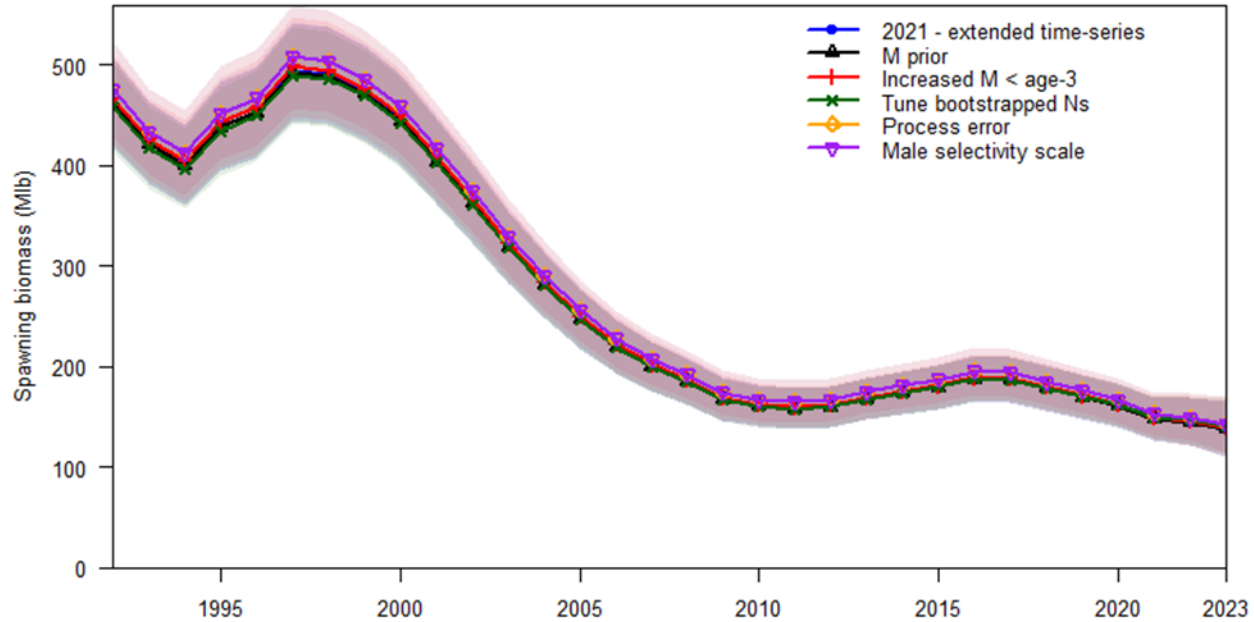


# Bridge of changes from 2021

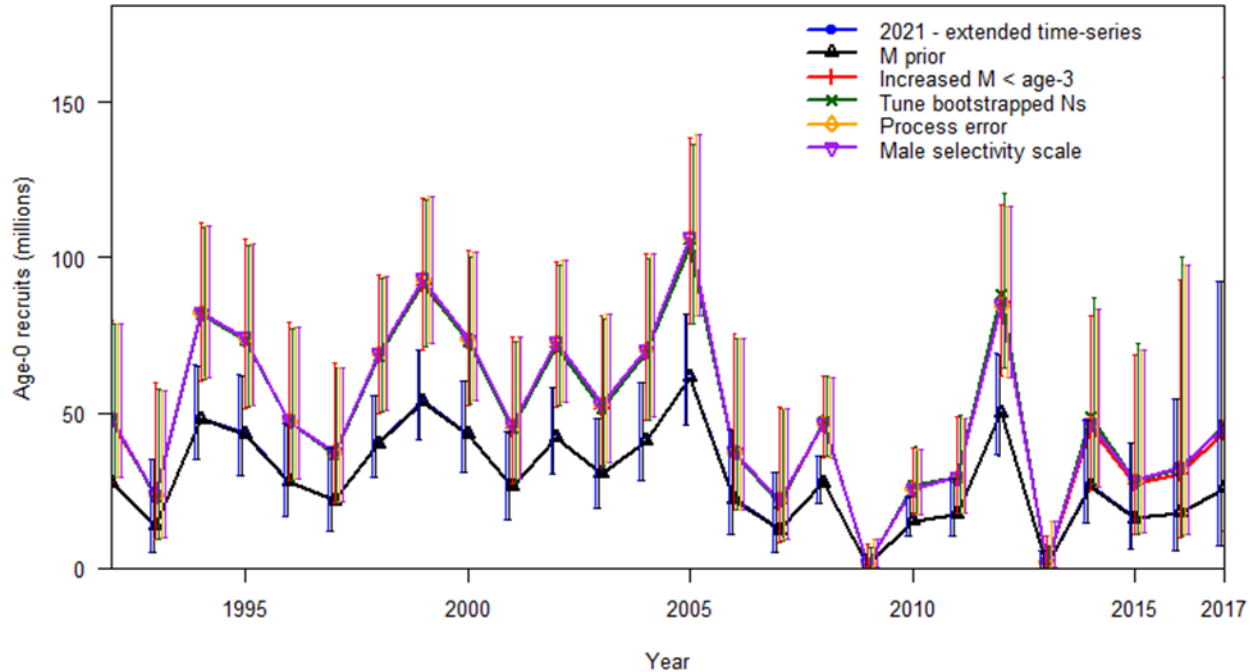
- 1) Extend the time series using projected mortality for 2022 (does not change the historical estimates)
- 2) Update the stock synthesis software (identical results)
- 3) Add prior on  $M$  and increased  $M$  at ages 0-2
- 4) Add bootstrapped age composition samples sizes
- 5) Retune process and observation error variances
- 6) Allow time-varying fishery sex-ratio (scale of male selectivity)
- 7) Test estimation of  $M$  in short models (relative to historically fixed value of 0.15)



# Coastwide short



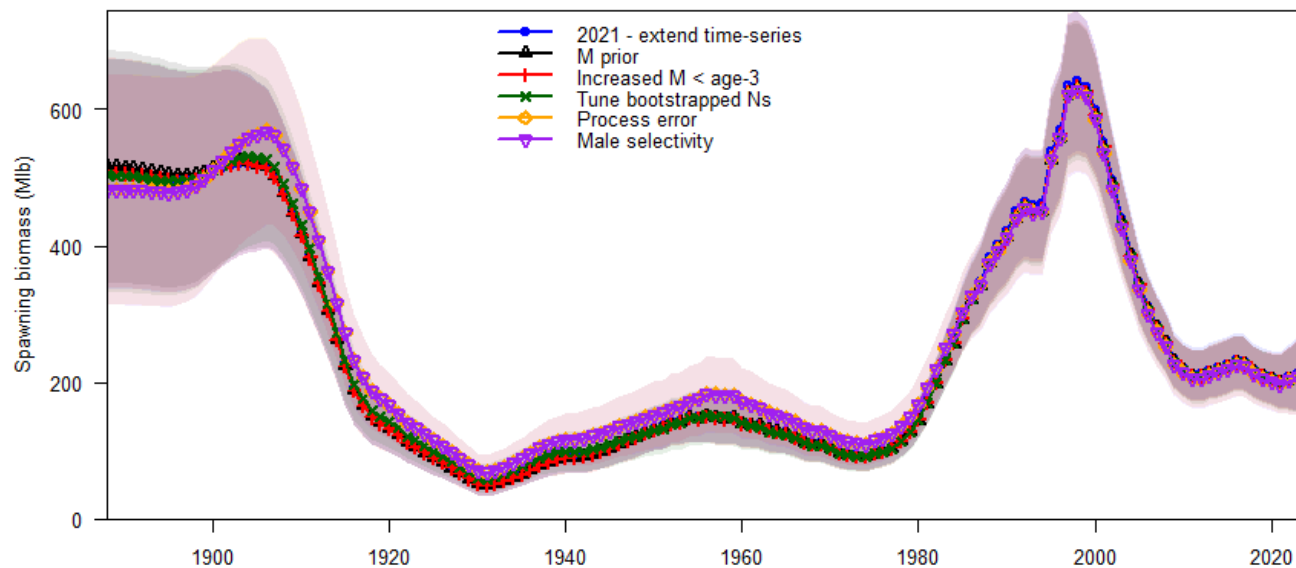
# Coastwide short



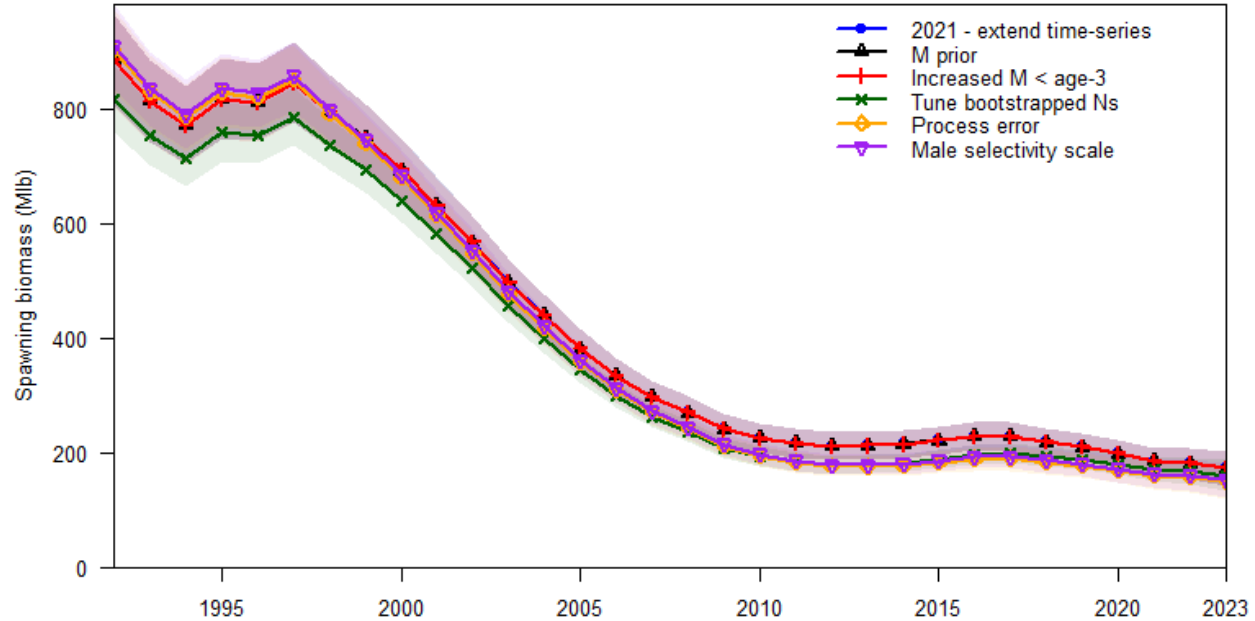
Increased M at ages 0-2 changes the scale,  
but not the relative recruitment strengths at age-0



# Coastwide long

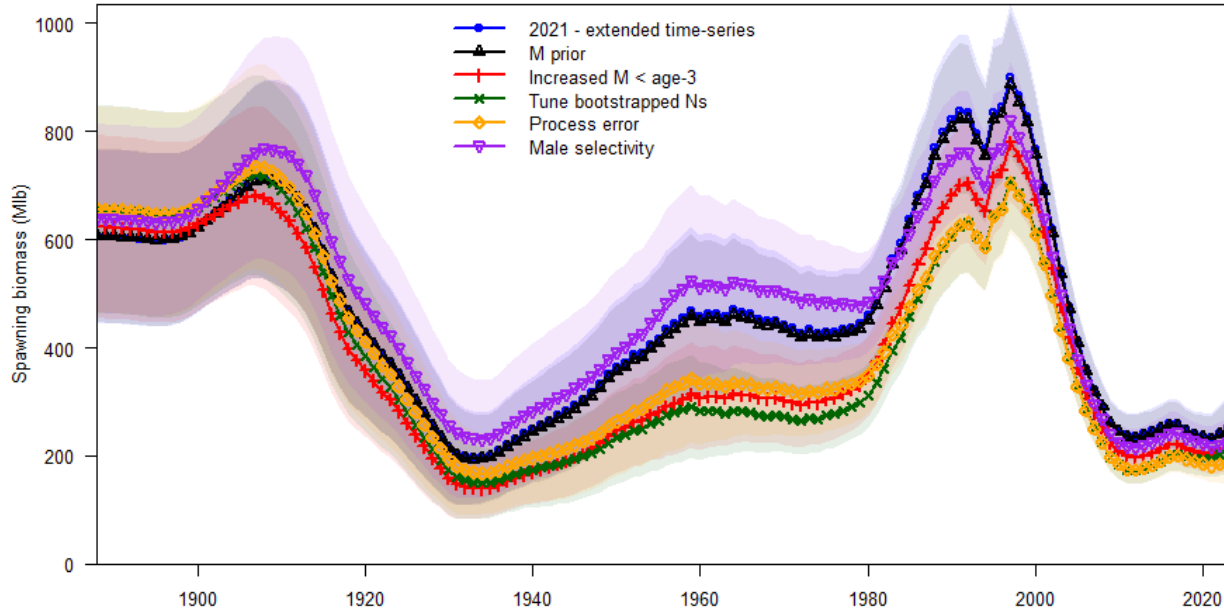


# AAF short





# AAF long



Most sensitive of the 4 models, primarily in the historical period

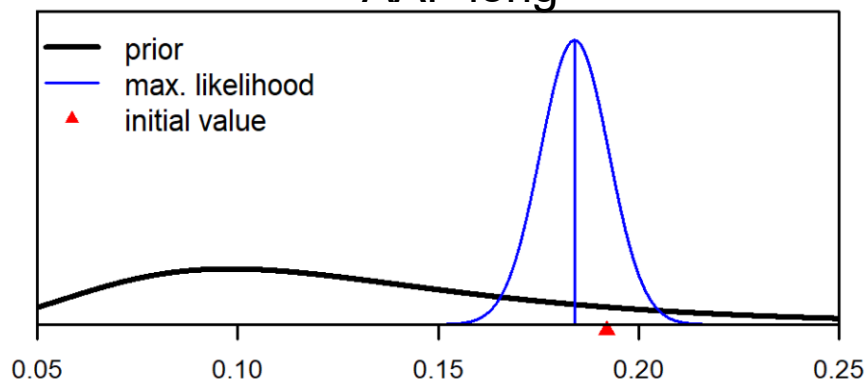
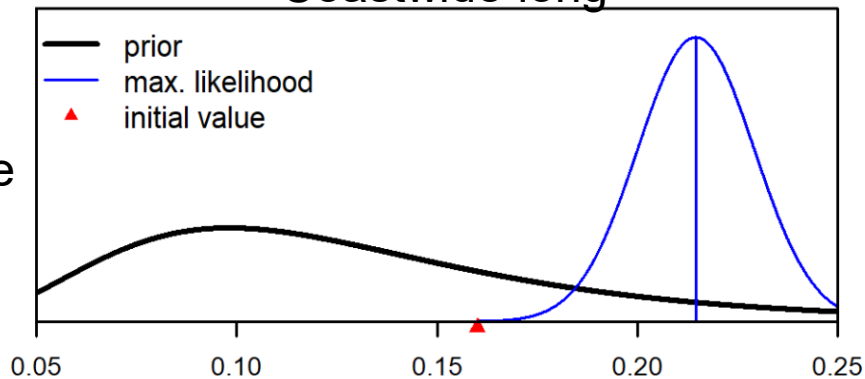


# Natural mortality - long models

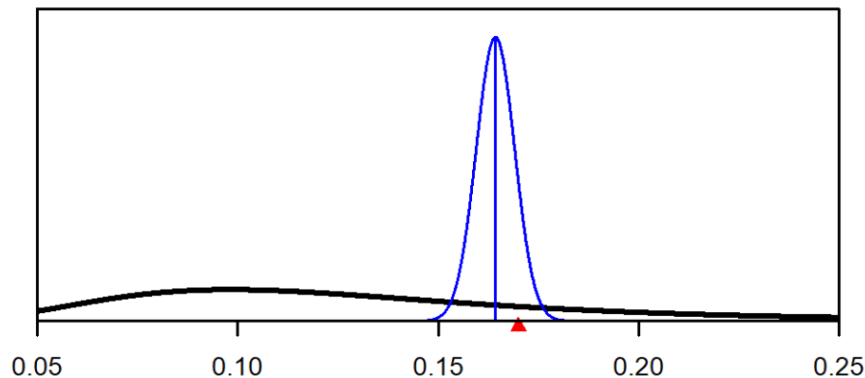
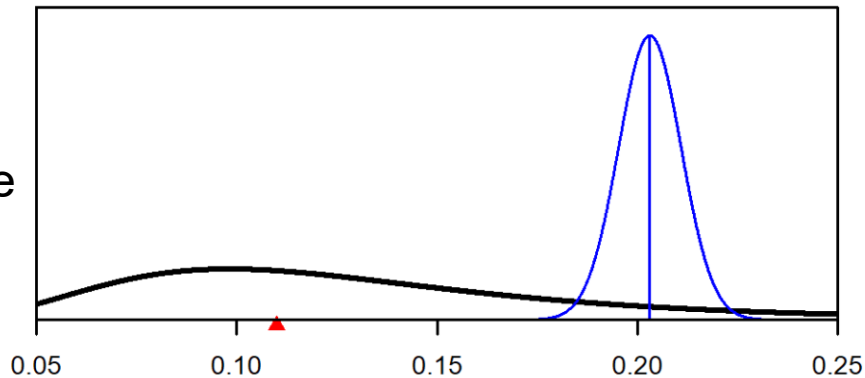
Coastwide long

AAF long

Female



Male



# Where did $M = 0.15$ come from?

- The 1998 assessment (Table 1 from Stewart and Martell 2014):

A history of stock assessments, methods employed, and major issues identified at the time.

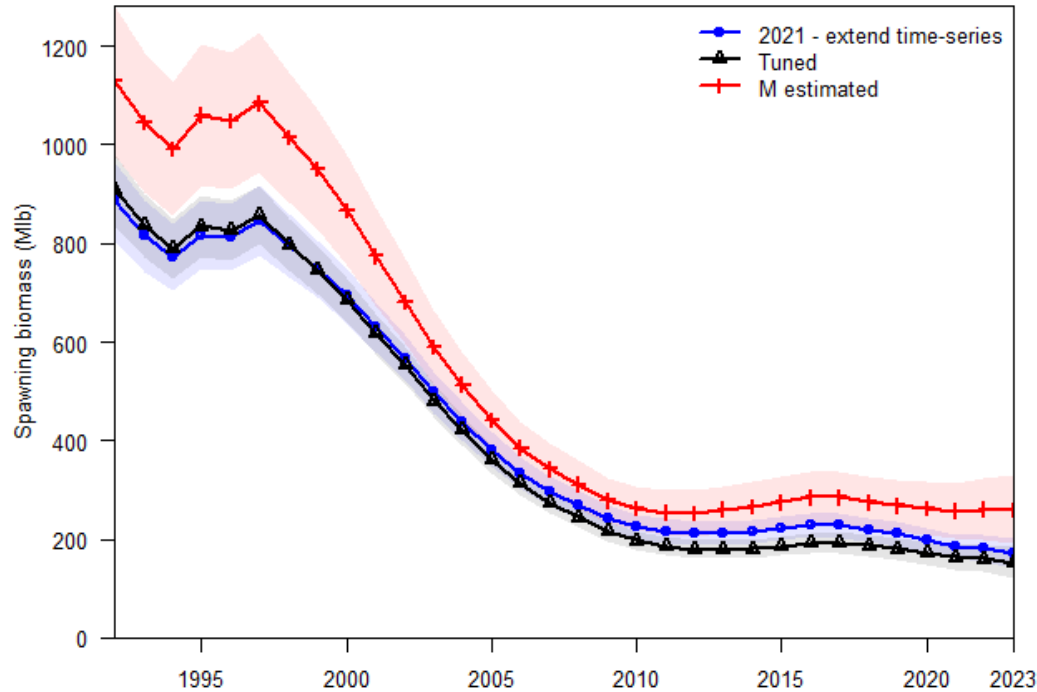
| Years                  | Model structure  |   |                           |                      | Noted issues                         |
|------------------------|--|---|---------------------------|----------------------|--------------------------------------|
|                        | Model(s)   | Spatial treatment                       | Natural mortality ( $M$ ) | Selectivity approach |                                      |
| ≤1977 <sup>a</sup>     | Yield, yield-per-recruit, simple stock-production models | Coastwide (core areas only)             | Various                   | Simple               | No growth or recruitment dynamics    |
| 1978–1981 <sup>b</sup> | Cohort analysis  | Coastwide (core areas only)             | Fixed at 0.2              | $F$ -at-age by year  | Unstable estimates                   |
| 1982–1983 <sup>c</sup> | Catch-AGE-ANalysis (CAGEAN)                              | Coastwide (core areas only)             | Fixed at 0.2              | Age-based            | Migratory dynamics not accounted for |
| 1984–1988 <sup>d</sup> | CAGEAN   | Coastwide and areas linked by migration | Fixed at 0.2              | Age-based            | “Trends differ by area”              |
| 1989–1994 <sup>e</sup> | CAGEAN   | Area-specific                           | Fixed at 0.2              | Age-based            | Retrospective pattern                |
| 1995–1997 <sup>f</sup> | Statistical Catch-Age (SCA)                              | Area-specific                           | Fixed at 0.2              | Length-based         | “ $M$ estimate imprecise”            |
| 1998–1999 <sup>g</sup> | SCA  | Area-specific                           | Fixed at 0.15             | Length-based         | “Poor fit to the data”               |
| 2000–2002 <sup>h</sup> | Revised SCA  | Area-specific                           | Fixed at 0.15             | Age-based            | Retrospective pattern                |
| 2003–2005 <sup>i</sup> | SCA  | Area-specific                           | Fixed at 0.15             | Length-based         | “Migratory dynamics created bias”    |
| 2006–2011 <sup>j</sup> | SCA  | Coastwide (all areas)                   | Fixed at 0.15             | Length-based         | Retrospective pattern                |

## Clark and Parma 1999:

- “Until 1998 the estimate of  $M = 0.20$  had been used in all assessments. This estimate is quite imprecise, and analysis done by the staff during the year suggested that a lower working value would be appropriate. The value of  $M = 0.15$  was chosen and used as a standard, which lowered abundance estimates by about 30%.”
- “Analysis done during the year by the staff showed that in the short term an overestimate of natural mortality could lead to a substantial overestimate of stock size when past fishing mortality rates were low, as they have been for Pacific halibut. On the other side, the consequences of an underestimate of natural mortality are less serious.”



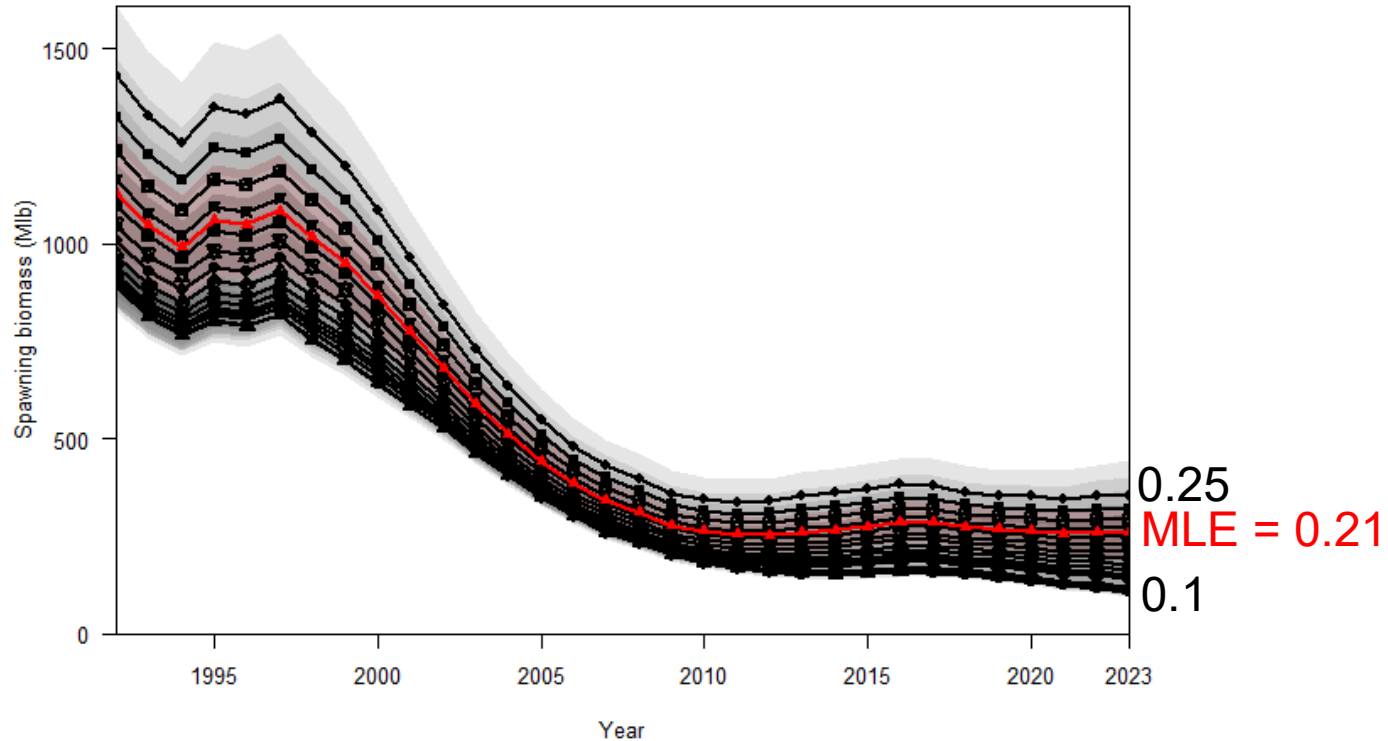
# Natural mortality – short AAF model



Modest effect on central tendency, large effect on uncertainty



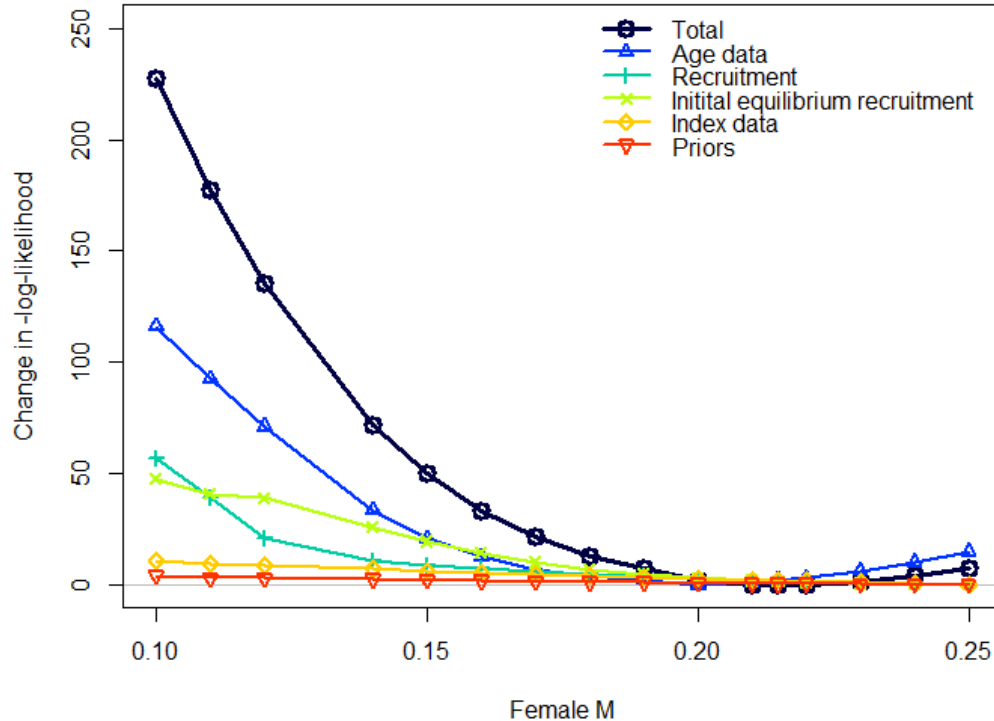
# Natural mortality – short AAF model



Scale of the estimated SB is closely correlated with M



# Natural mortality – short AAF model



All data sources show improved fit at values of female M higher than 0.15



# Model diagnostics and results

- Convergence criteria – no issues identified in AAF short, CW short, or CW long
- AAF long jitter analysis showed a challenging likelihood surface requiring good starting values



# Model diagnostics and results

- All models generally fit the FISS and fishery indices well
- Starting from bootstrapped sample sizes only 4B age data in the AAF short and long models tuned to the input (maximum value)





# Model diagnostics and results (Table 16)

|  | Average iterated input | Harmonic mean effective | Francis weight effective | Maximum Pearson residual |
|--|------------------------|-------------------------|--------------------------|--------------------------|
| <b>Coastwide short</b>                   |                        |                         |                          |                          |
| <b>Fishery</b>                           | 62                     | 294                     | 62                       | 2.45                     |
| <b>Discards<sup>1</sup></b>              | 13                     | 270                     | 49                       | 0.98                     |
| <b>Non-directed discards<sup>1</sup></b> | 5                      | 47                      | 39                       | 2.25                     |
| <b>Recreational<sup>1</sup></b>          | 5                      | 114                     | 27                       | 0.88                     |
| <b>FISS</b>                              | 242                    | 668                     | 242                      | 2.06                     |
| <b>Coastwide long</b>                    |                        |                         |                          |                          |
| <b>Fishery</b>                           | 112                    | 289                     | 122                      | 4.09                     |
| <b>Discards<sup>1</sup></b>              | 6                      | 210                     | 90                       | 0.78                     |
| <b>Non-directed discards<sup>1</sup></b> | 3                      | 37                      | 7                        | 1.33                     |
| <b>Recreational<sup>1</sup></b>          | 3                      | 145                     | 31                       | 0.51                     |
| <b>FISS</b>                              | 82                     | 194                     | 83                       | 2.88                     |

<sup>1</sup>Inputs down-weighted, and not iteratively reweighted.



# Model diagnostics and results (Table 16)

|                                    | Average iterated input | Harmonic mean effective | Francis weight effective | Maximum Pearson residual |
|------------------------------------|------------------------|-------------------------|--------------------------|--------------------------|
| <b>AAF short</b>                   |                        |                         |                          |                          |
| Region 2 fishery                   | 723                    | 676                     | 1,078                    | 4.47                     |
| Region 3 fishery                   | 808                    | 699                     | 951                      | 3.85                     |
| Region 4 fishery                   | 23                     | 78                      | 36                       | 3.54                     |
| Region 4B fishery <sup>2</sup>     | 36                     | 138                     | 81                       | 1.82                     |
| Discards <sup>1</sup>              | 13                     | 219                     | 73                       | 1.21                     |
| Non-directed discards <sup>1</sup> | 5                      | 58                      | 22                       | 1.12                     |
| Recreational <sup>1</sup>          | 5                      | 143                     | 20                       | 0.85                     |
| Region 2 FISS                      | 7                      | 86                      | 7                        | 1.04                     |
| Region 3 FISS                      | 18                     | 262                     | 18                       | 1.25                     |
| Region 4 FISS                      | 66                     | 181                     | 63                       | 3.95                     |
| Region 4B FISS <sup>2</sup>        | 41                     | 185                     | 50                       | 1.83                     |

<sup>1</sup>Inputs down-weighted, and not iteratively reweighted.

<sup>2</sup>Iterated sample size equal to maximum (bootstrapped input).



# Model diagnostics and results (Table 16)

|                                    | Average iterated input | Harmonic mean effective | Francis weight effective | Maximum Pearson residual |
|------------------------------------|------------------------|-------------------------|--------------------------|--------------------------|
| <b>AAF long</b>                    |                        |                         |                          |                          |
| Region 2 fishery                   | 322                    | 304                     | 651                      | 4.31                     |
| Region 3 fishery                   | 266                    | 309                     | 544                      | 3.78                     |
| Region 4 fishery                   | 18                     | 60                      | 28                       | 4.36                     |
| Region 4B fishery <sup>2</sup>     | 37                     | 129                     | 80                       | 1.90                     |
| Discards <sup>1</sup>              | 6                      | 189                     | 84                       | 1.56                     |
| Non-directed discards <sup>1</sup> | 3                      | 43                      | 8                        | 1.12                     |
| Recreational <sup>1</sup>          | 8                      | 151                     | 23                       | 0.91                     |
| Region 2 FISS                      | 7                      | 78                      | 8                        | 1.39                     |
| Region 3 FISS                      | 12                     | 101                     | 13                       | 1.26                     |
| Region 4 FISS                      | 72                     | 182                     | 68                       | 3.53                     |
| Region 4B FISS <sup>2</sup>        | 41                     | 185                     | 45                       | 1.93                     |

<sup>1</sup>Inputs down-weighted, and not iteratively reweighted.

<sup>2</sup>Iterated sample size equal to maximum (bootstrapped input).



# Model diagnostics and results (Table 17)

|   | Model                     |                           |                           |   |
|---|---------------------------|---------------------------|---------------------------|---|
|   | Coastwide Short           | Coastwide Long            | AAF Short                 | AAF Long  |
| Female M  | 0.150 (Fixed)             | 0.215 (0.186-0.243)       | 0.211 (0.195-0.227)       | 0.184 (0.167-0.200)   |
| Male M  | 0.149 (0.138-0.159)       | 0.203 (0.188-0.218)       | 0.177 (0.167-0.187)       | 0.164 (0.154-0.173)   |
| Log( $R_0$ )  | 11.375<br>(11.167-11.582) | 11.857<br>(11.546-12.168) | 12.347<br>(12.115-12.579) | 11.545<br>(11.262-11.829)   |
| Initial log( $R_0$ ) offset                                     | -1.469 (-1.685--1.253)    | NA                        | -0.368 (-0.596-0.140)     | NA  |
| Environmental Link ( $\beta$ )                                  | NA                        | 0.372 (0.144-0.600)       | NA                        | 0.349 (0.129-0.569)   |
| Survey Log( $q$ ) $\Delta$ 1984<br>(transition to circle hooks) | NA                        | 0.945 (0.592-1.299)       | NA                        | R2: 1.222 (0.844-1.600)<br>R3: 1.822 (1.553-2.092)  |
| Fishery Log( $q$ ) $\Delta$ 1984                                | NA                        | 0.718 (0.541-0.895)       | NA                        | R2: 0.586 (0.402-0.769)<br>R3: 0.920 (0.724-1.115)<br>R4: 0.858 (0.663-1.053)<br>R4B: 0.529 (0.347-0.712) |
| 2012 Recruitment (Millions)                                     | 85 (58-112)               | 283 (127-439)             | 278 (163-393)             | 195 (119-270)   |
| 2022 SB (Million lb)  | 150 (126-173)             | 202 (155-250)             | 259 (199-320)             | 218 (178-260)   |



# Model diagnostics and results

- Switch to document or files as needed



# Model strengths and weaknesses

- Varied
- AAF long is the most challenging technically
- Each model is internally consistent, but differs importantly from the others
- No single model clearly exceeds the others across all aspects of model fit and performance



# Outline (2)

- Evaluation of uncertainty
  - Sensitivity analyses
  - Likelihood profiles
  - Retrospective analyses
  - Other considerations
- Ensemble
  - Methods
  - Weighting based on predictive skill
  - Preliminary results for 2022
- Research priorities and future development



# Sensitivity analyses

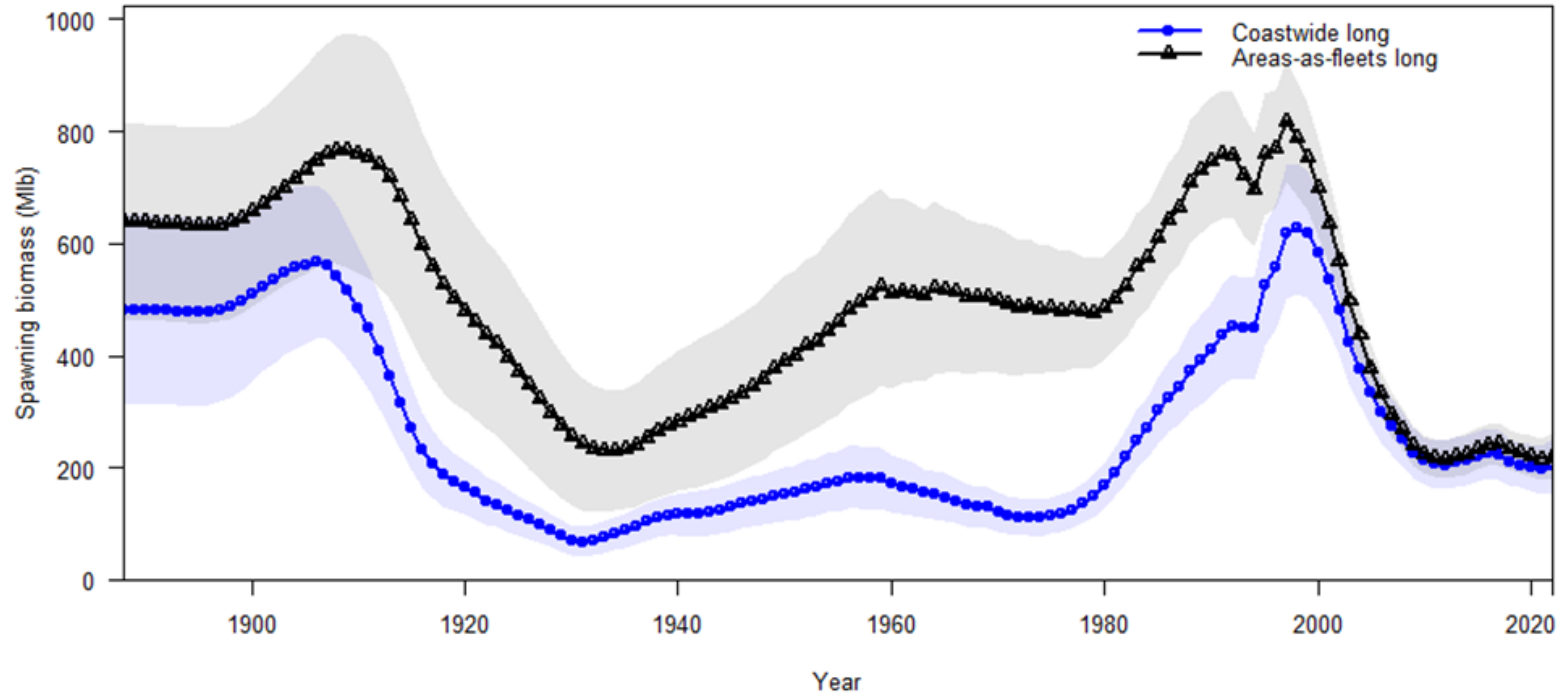
- 2015: *Ensemble*, fishery  $q$ , M:F selectivity,  $M$ ,  $h$ , historical selectivity
- 2016: *Ensemble*, maturity, M:F selectivity, directed fishery DMRs
- 2017: *Ensemble*, maturity, M:F selectivity, unobserved mortality (e.g., depredation)
- 2018: *Ensemble*, maturity, M:F selectivity, unobserved mortality 2019: *Ensemble*,  $M$ ,  $h$ , data weighting
- 2020: *Ensemble*, bridging
- 2021: *Ensemble*, unobserved mortality, PDO, maturity
- 2022: *Ensemble*, PDO,  $M$





# Sensitivity analyses

- Historical connectivity: long CW vs. long AAF



# Treatment of the PDO

Small set of hypotheses about how recruitment is related to environmental conditions:

1. Status quo
2. Annual deviations
3. Moving average
4. Extreme values
5. Excluding the PDO



# Treatment of the PDO

Test: Reduction in the Root-Mean-Squared-Error (RMSE) of the estimated recruitment deviations

→ Have we exchanged process error for modelled process?



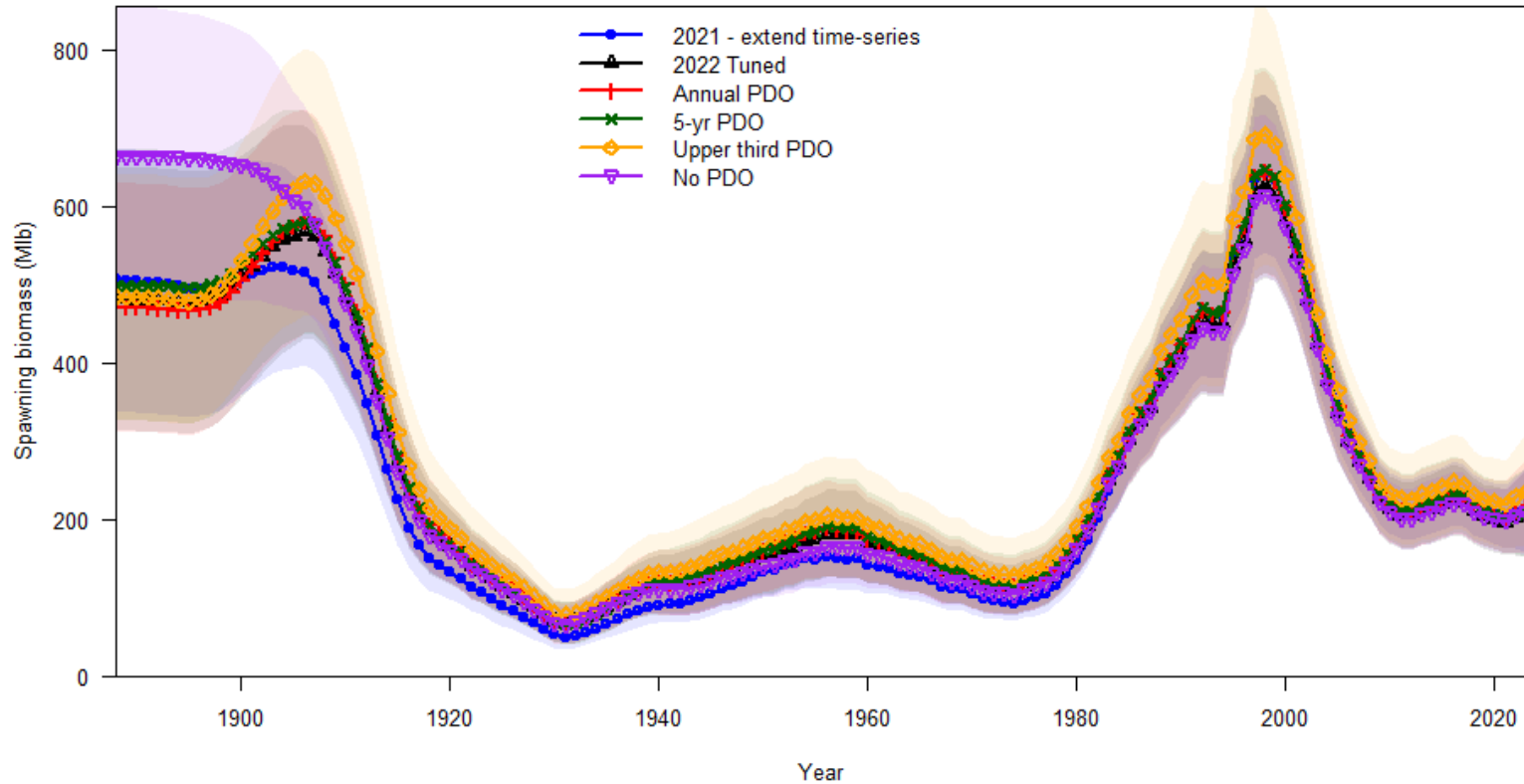
# Treatment of the PDO (Table 18)

| Treatment of the PDO                          | Model   |             |                  |             |
|---|---------|-------------|------------------|-------------|
|   | CW long |             | AAF long         |             |
|   | RMSE    | Coefficient | RMSE             | Coefficient |
| Status quo (binary regimes)                   | 0.42    | 0.37        | 0.38             | 0.35        |
| Annual deviations                             | 0.44    | 0.45        | 0.38             | 0.38        |
| 5-year moving average                         | 0.45    | 0.34        | 0.39             | 0.32        |
| Binary on largest 1/3 <sup>rd</sup> of values | 0.45    | 0.50        | Did not converge |             |
| Exclude PDO                                   | 0.48    | NA          | 0.42             | NA          |



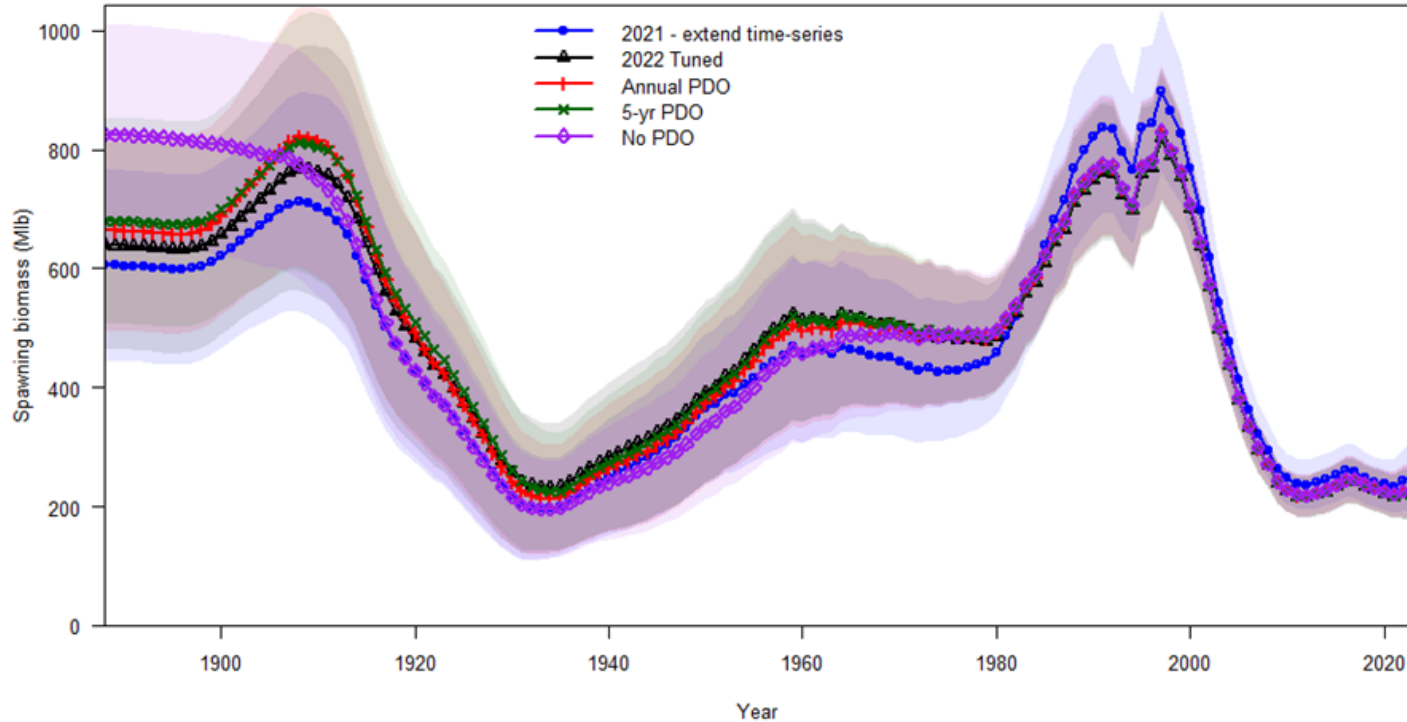
# Sensitivity analyses: PDO

Coastwide long



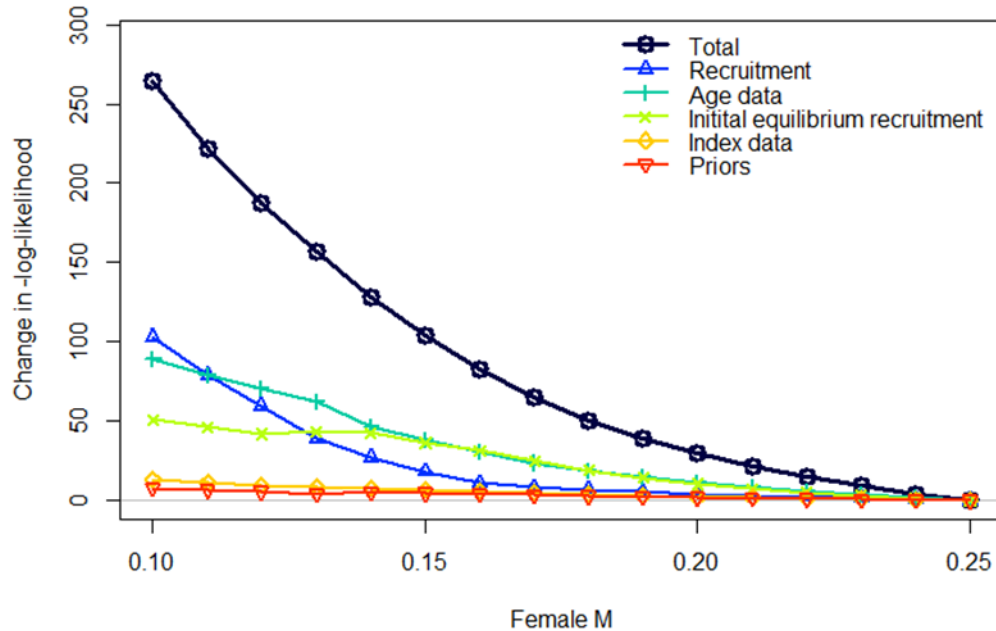
# Sensitivity analyses: PDO

AAF long



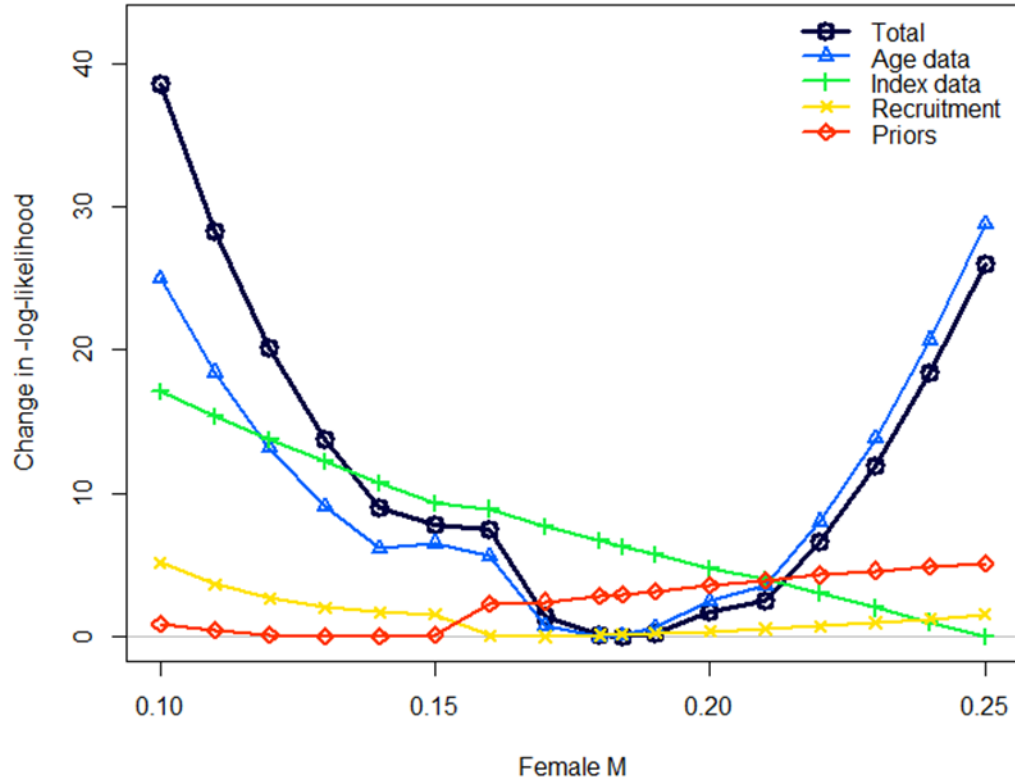
# Likelihood profiles: Female M

Coastwide short



# Likelihood profiles: Female M

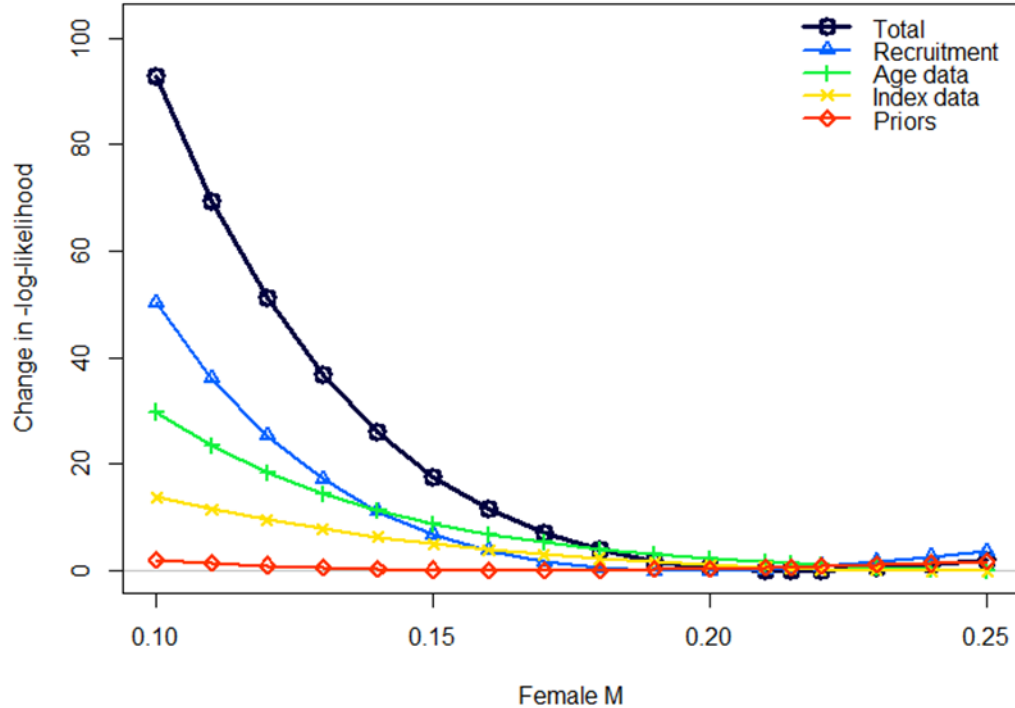
AAF long





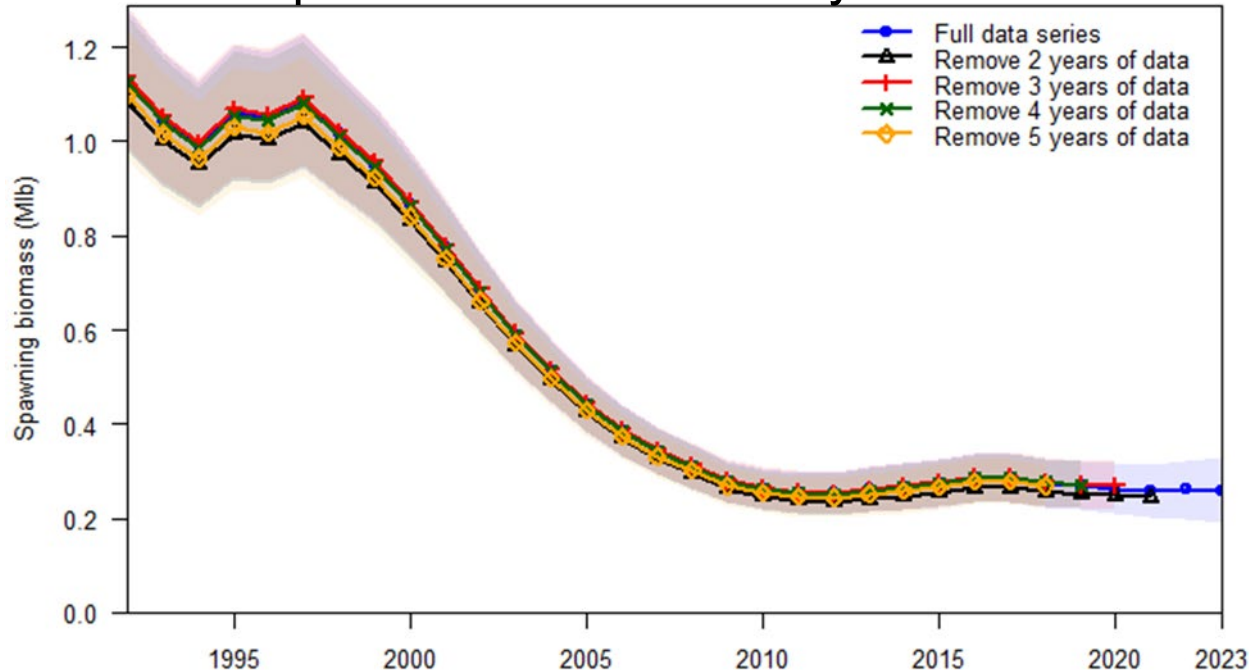
# Likelihood profiles: Female M

Coastwide long



# Retrospective analyses (Figures 64-67)

- Far better performance than previous models
  - Allowing the scale of male selectivity to vary largely separates the historical period from the recent year's data



# Other considerations

- No additional work done since the 2019 assessment on Bayesian models for 2022
- Spatially-explicit dynamics (e.g., alternative hypotheses about 4B, Russian waters) should continue to be considered as MSE and research program efforts are continued
- Uncertainty in sources of mortality generally addressed through sensitivity analyses; however, these could be integrated



# Outline (2)

- Evaluation of uncertainty
  - Sensitivity analyses
  - Likelihood profiles
  - Retrospective analyses
  - Other considerations
- Ensemble
  - Methods
  - Weighting based on predictive skill
  - Preliminary results for 2022
- Research priorities and future development



# Ensemble methods

- Random draws from each model for quantities of interest, including variance and covariance estimates
- Number ( $n$ ) of replicates from each model ( $m$ ) determines the weight ( $w$ ) within the set of models:

$$w_m = \frac{n_m}{\sum_m n_m}$$



# Weighting methods considered

- AIC
- Strength of retrospective patterns
- Fit to FISS index
- Expert opinion (*status quo*: equal weights)
- Prediction performance



# Properties of weighting methods

- Does not require the same data in all models
- Inherently penalizes over- or under-parameterized models
- Based on an objective method that can be applied annually even as models, data, and performance change
- Related closely to management quantities/projections

→ Prediction performance.

FISS index is a logical proxy for the spawning biomass and the biomass available to the fisheries and therefore fishing intensity and other management quantities.



# Mean Absolute Standardized Error (MASE)

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

- MASE >1: model skill is worse than the naïve prediction (last year's observation)
- 1: Equal to the naïve prediction
- <1: Better than naïve prediction
- 0: Perfect prediction





# Mean Absolute Standardized Error (MASE)

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

- ‘Standardized’ MASE accounts for differing variance estimates for annual observations
- Interpretation of values remains unchanged



# MASE weights

- For models with a MASE of  $\leq 1$ :

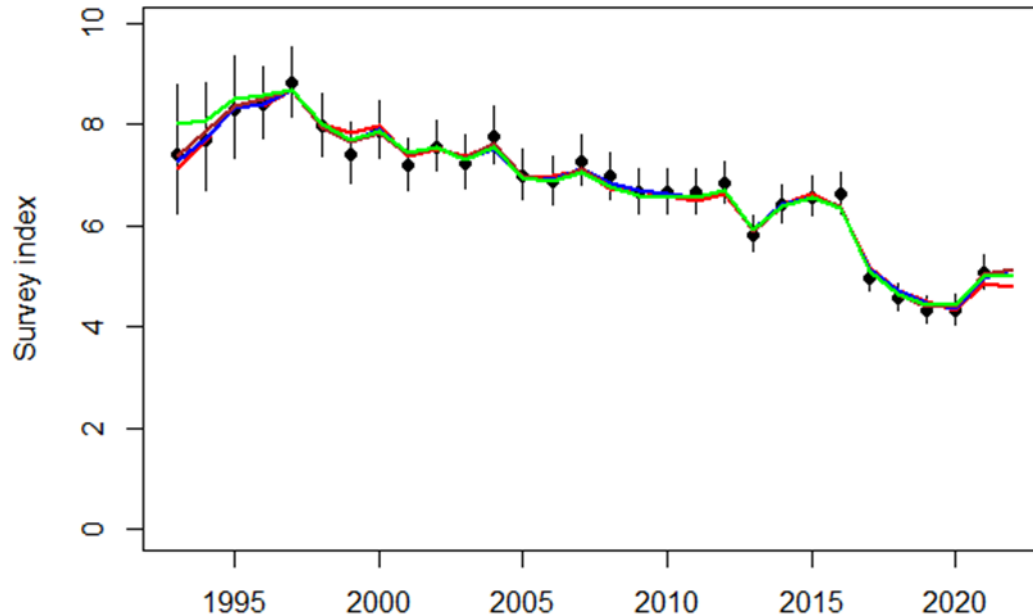
$$MASE\ weight_m = \frac{1 - MASE_m}{\sum_{m=1}^M 1 - MASE_m}$$

- A model with MASE = 1 gets zero weight
- A model with MASE = 0 gets maximum weight



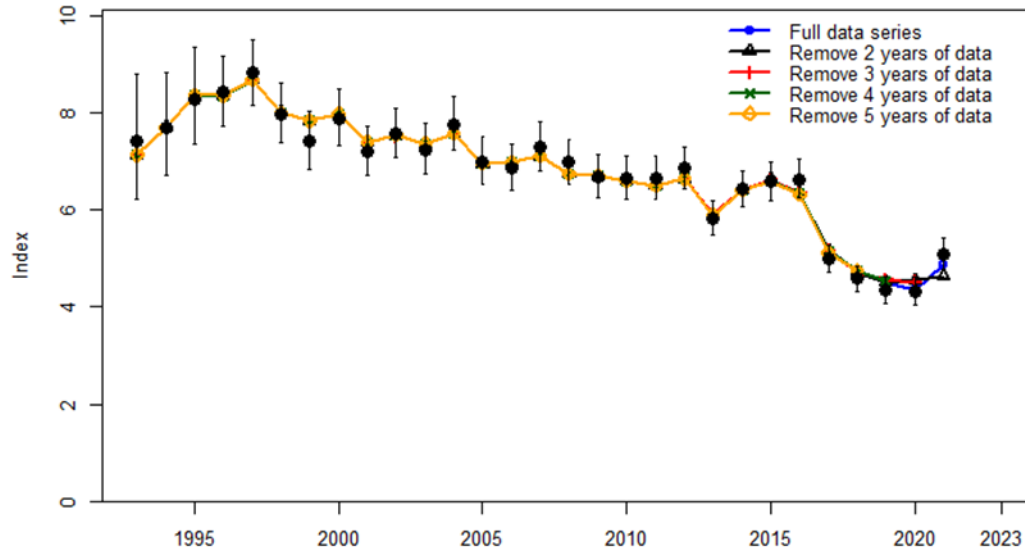
# Prediction period

- Want to know the performance in predicting the unobserved index (next year's observation)



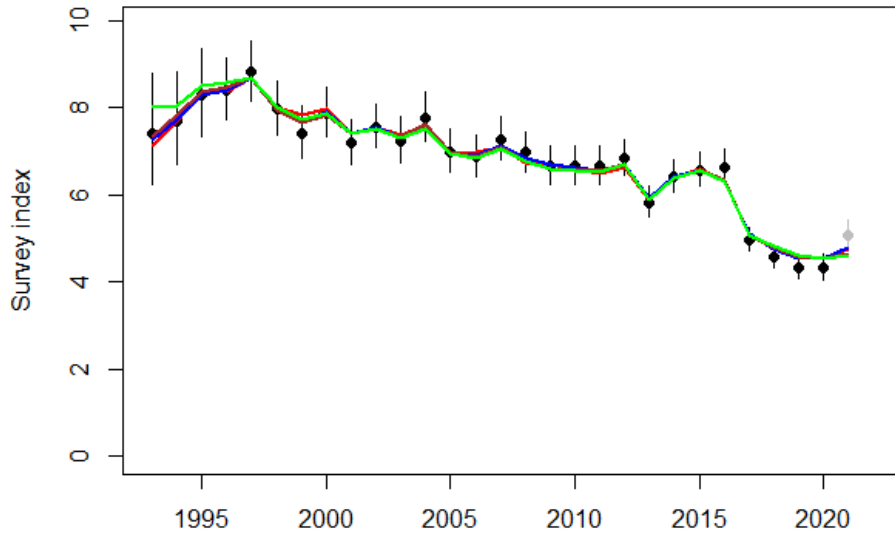
# Prediction period

- By removing previous year's data and predicting forward 'recent' skill can be quantified
  - Considered a 1, 2, 3 , or 4-year average

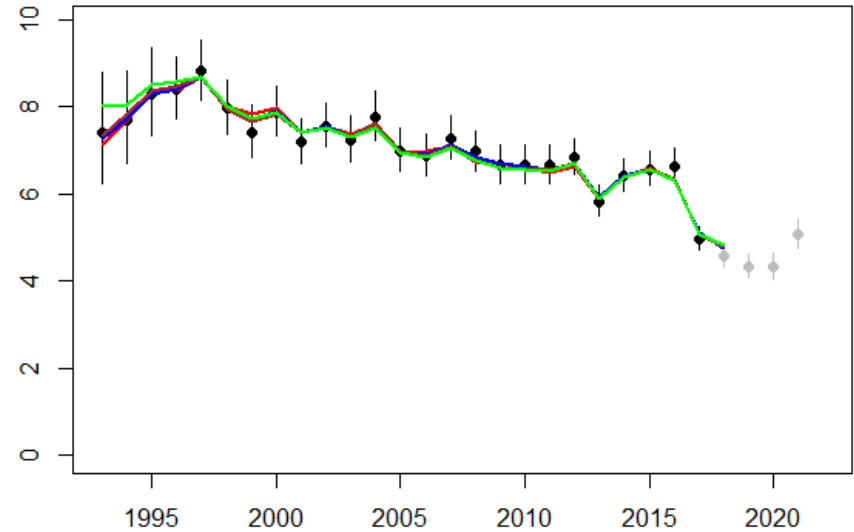


# Prediction period

Data through 2020

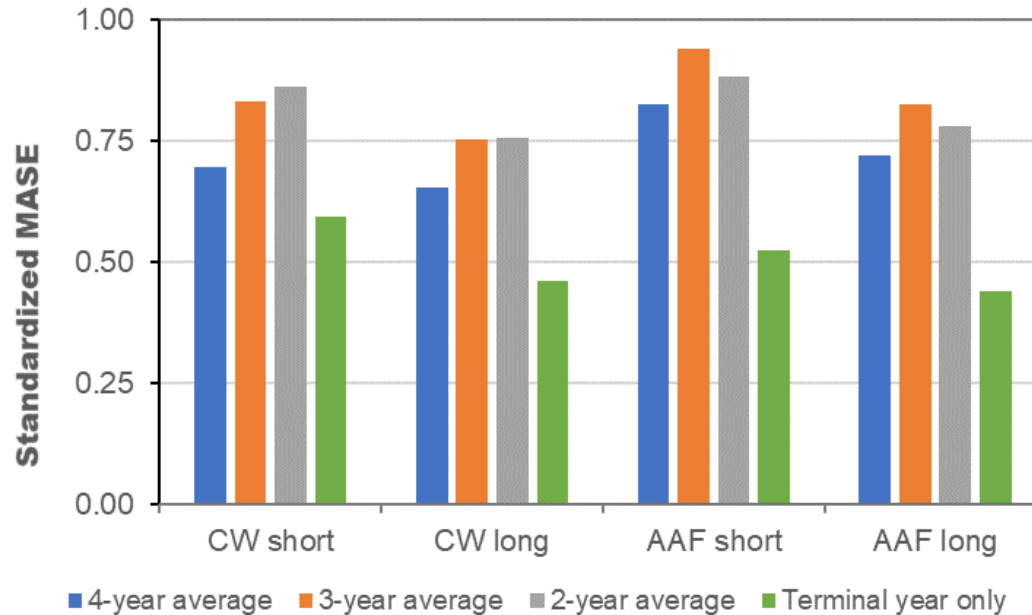


Data through 2017

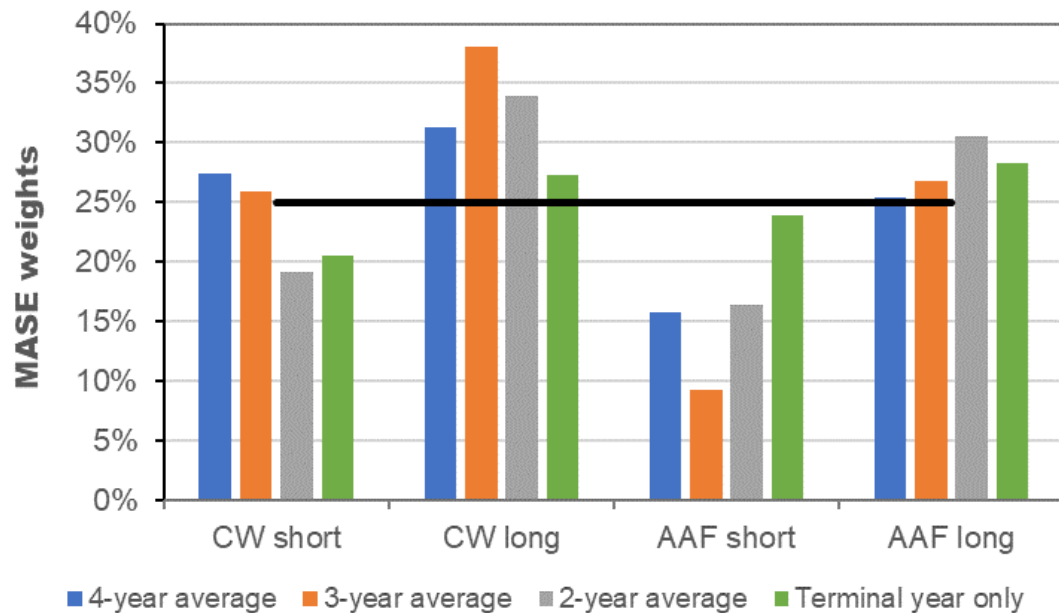


# MASE estimates (Table 19)

- All models better than the naïve prediction



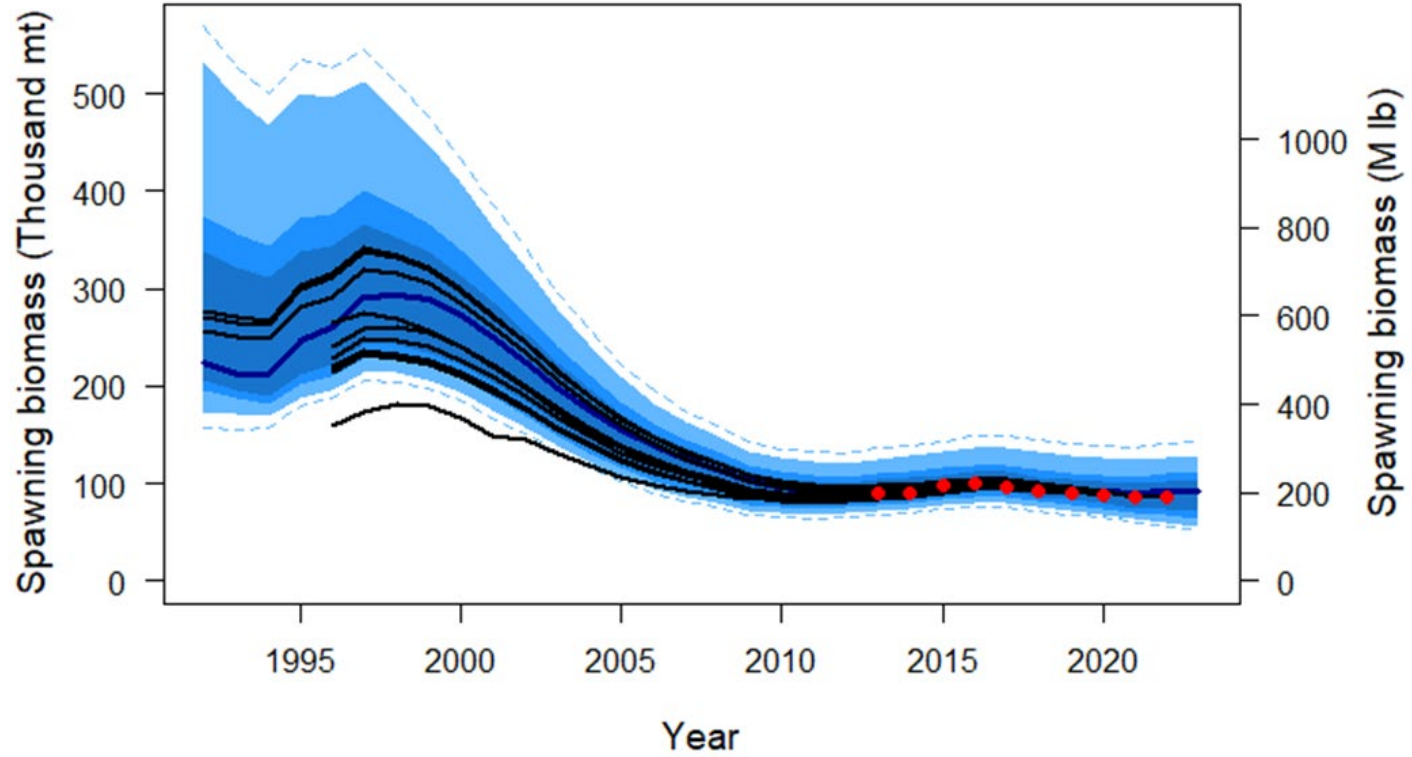
# MASE weights (Table =20)



*Status quo equal weights*

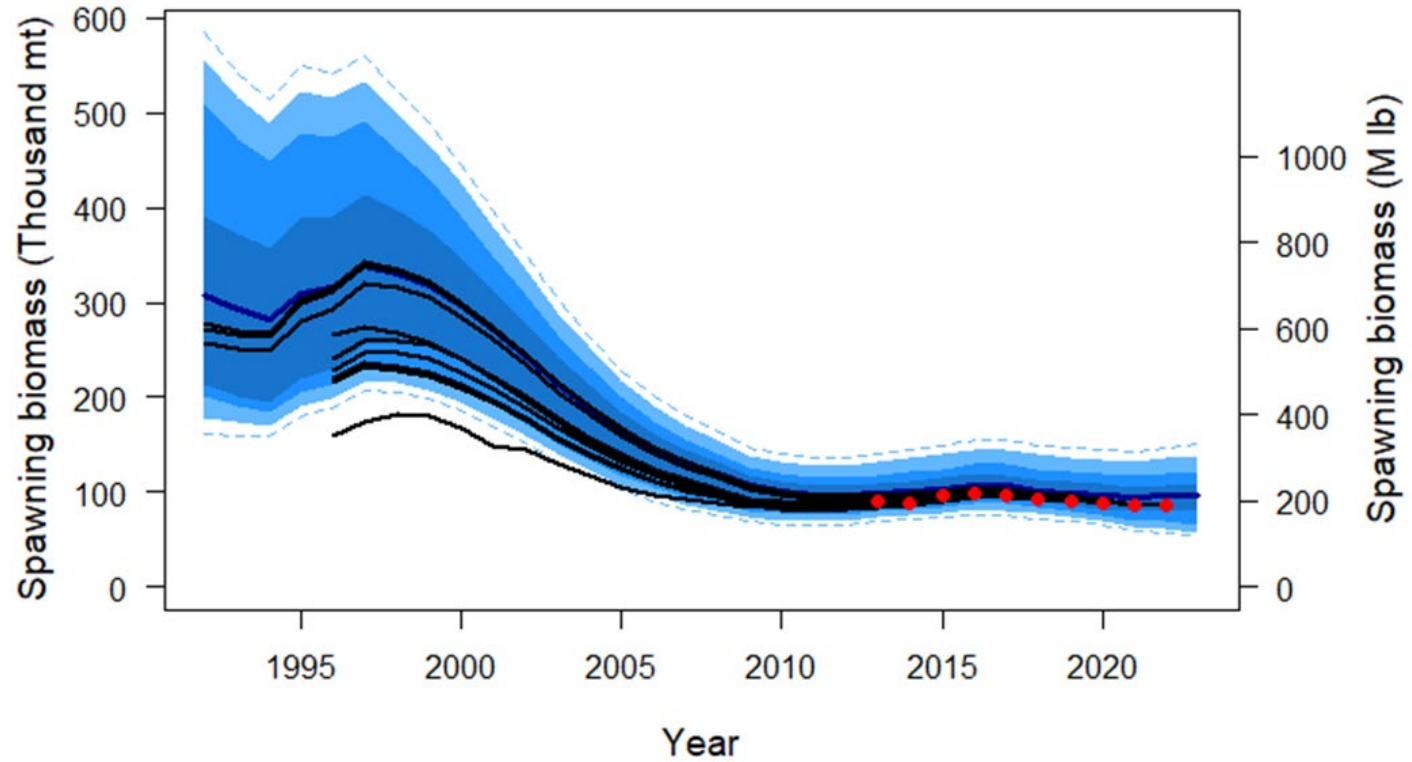


# Effect on trends: 3-year MASE

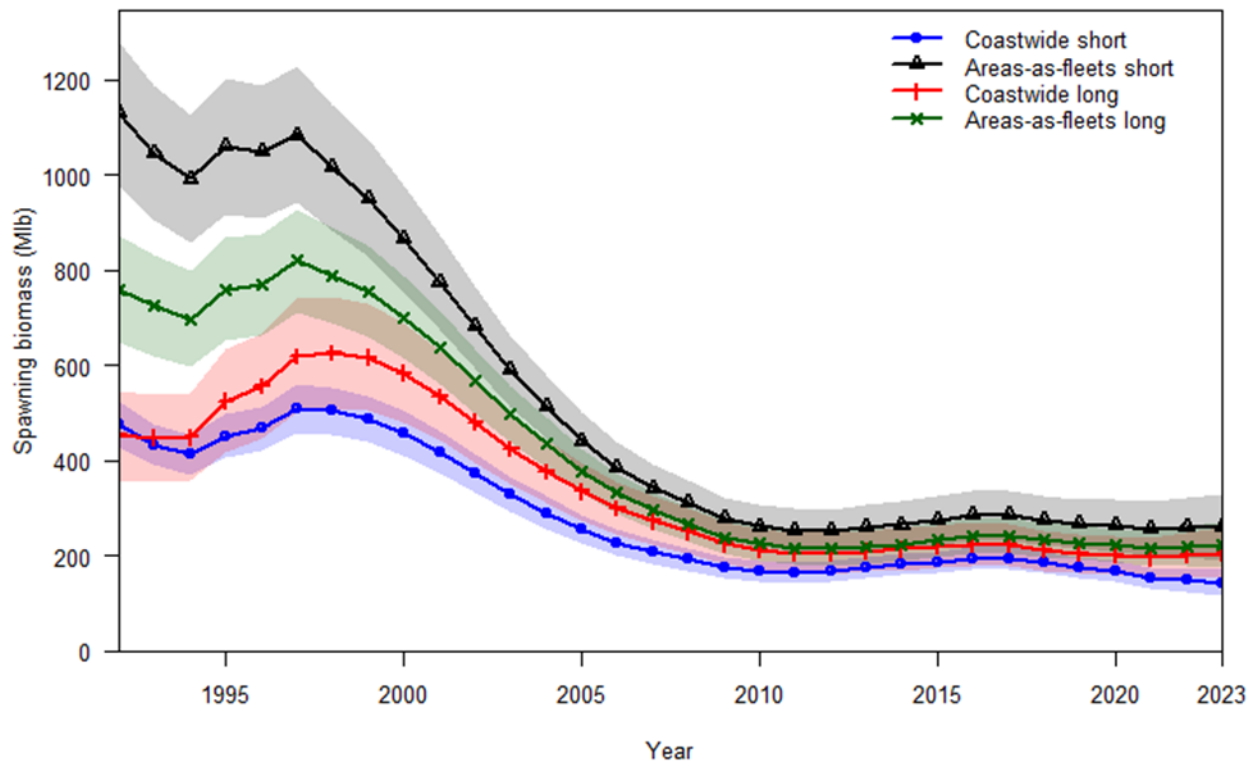




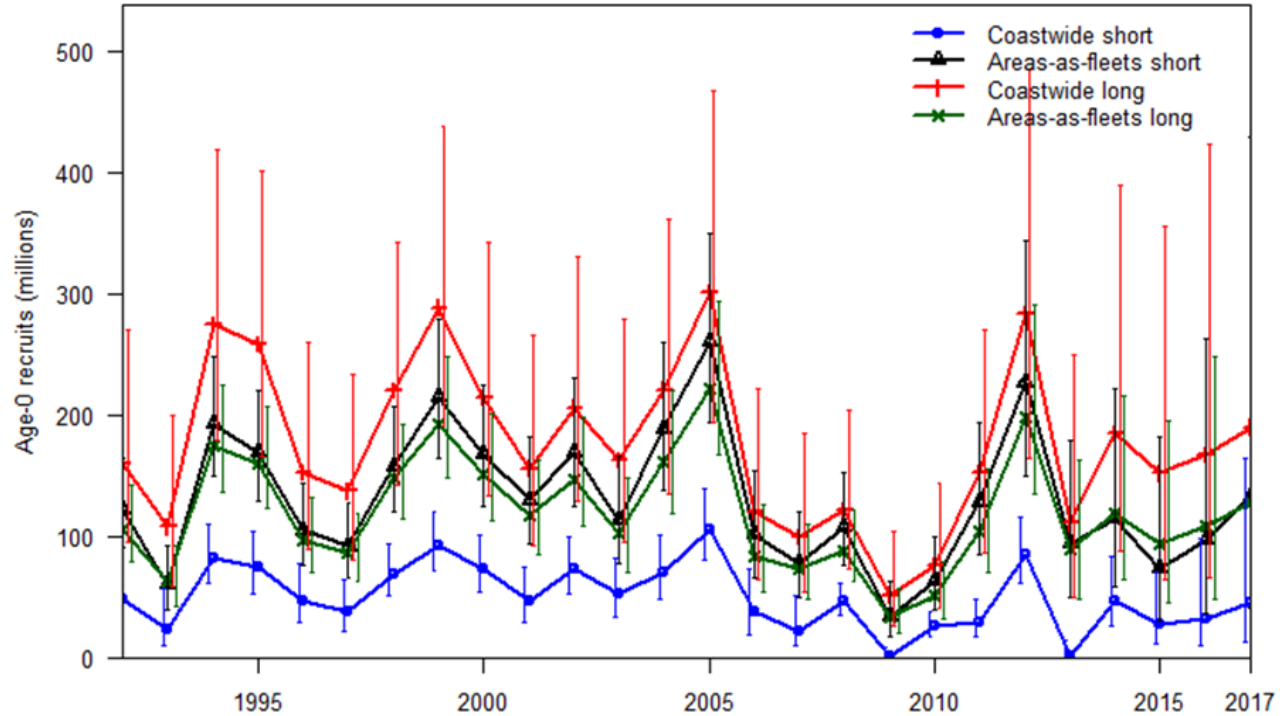
# Effect on trends: 1-year MASE



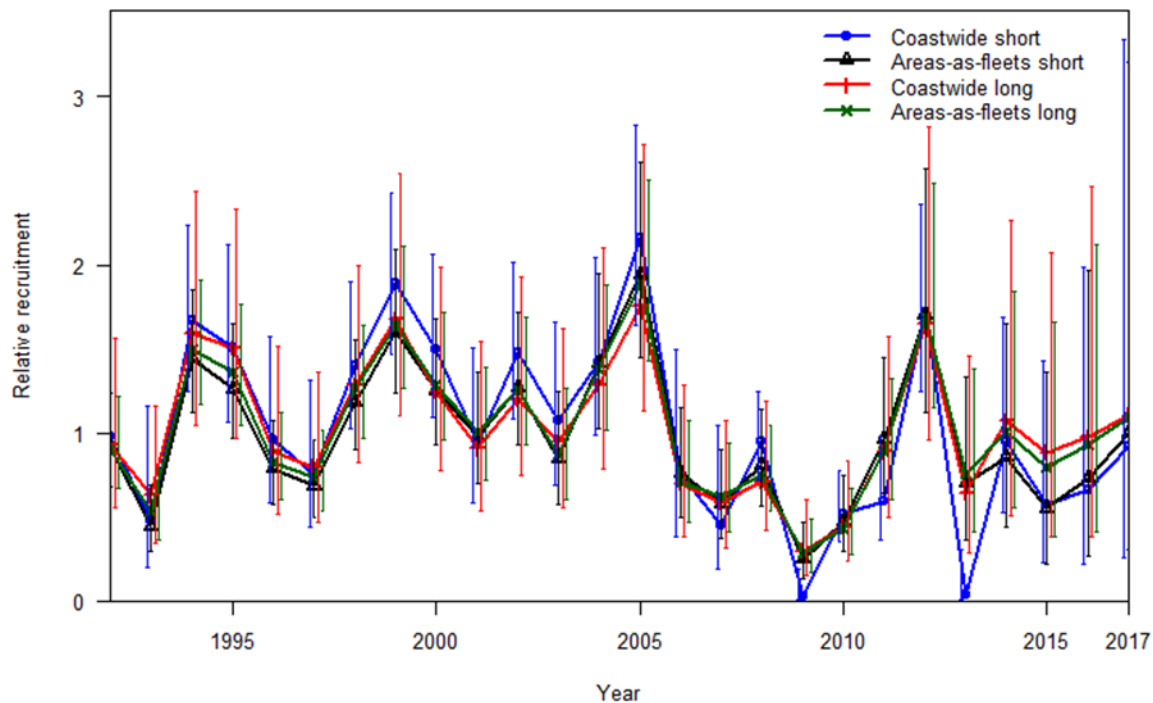
# Preliminary results



# Preliminary results



# Preliminary results



# Remaining changes for 2022

- Response to suggestions and comments from SRB020 (this meeting) and SRB021 (September)
- All 2022 data
- Sex-ratio of the 2021 commercial proportions at age



# Outline (2)

- Evaluation of uncertainty
  - Sensitivity analyses
  - Likelihood profiles
  - Retrospective analyses
  - Other considerations
- Ensemble
  - Methods
  - Weighting based on predictive skill
  - Preliminary results for 2022
- Research priorities and future development



# Research priorities

- Highest priorities are elevated to the 5-year research plan
- Longer list documents and provides a record of items needing further work



# Recommendation/s

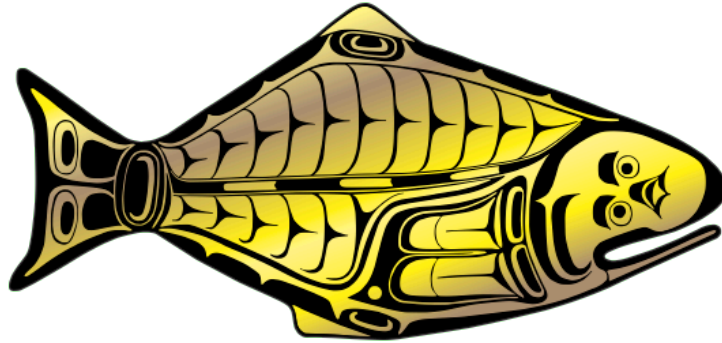
That the SRB:

- a) **NOTE** paper IPHC-2022-SRB020-07 which provides an overview of model development for 2022.
- b) **REQUEST** any changes to be made and reviewed at SRB021.
  - Key topics that would be helpful to consider:
    - Treatment of M
    - Ensemble weighting





**INTERNATIONAL PACIFIC**



**HALIBUT COMMISSION**

