Development of the 2025 Pacific halibut (*Hippoglossus stenolepis*) stock assessment

PREPARED BY: IPHC SECRETARIAT (I. STEWART AND A. HICKS; 9 MAY 2025)

Table of Contents

Summary

This document reports preliminary analyses in development of the 2025 Pacific halibut (*Hippoglossus stenolepis*) stock assessment. It follows the previous full stock assessments conducted in 2022 (Stewart and Hicks 2022) and 2019, including the independent peer review in 2019 (Stewart and Hicks 2019b; Stewart and Hicks 2020; Stokes 2019). Since the 2022 full stock assessment, two updates have been completed in 2023 (Stewart and Hicks 2024) and 2024 (Stewart and Hicks 2025) which included little change to the data or methods. Following the review of this document in June 2025 (SRB026), requested revisions will be considered and presented for additional review in September 2025 (SRB027), and the final 2025 assessment will be produced for the IPHC's Interim (IM101) and Annual (AM102) meetings. Updated data sources, including the results of the 2025 Fishery-Independent Setline Survey (FISS), logbook and biological data from the 2025 commercial fishery, and sex-ratio information from the 2024 commercial landings-at-age will be included for the final 2025 analysis.

Creating robust, stable, and well-performing stock assessment models for the Pacific halibut stock has historically proven to be challenging due to the highly dynamic nature of the biology, distribution, and fisheries (Stewart and Martell 2014). The stock assessment for Pacific halibut has evolved through many different modeling approaches over the last 30 years (Clark 2003). These changes have reflected improvements in fisheries analysis methods, changes in model assumptions, and responses to recurrent retrospective biases and other lack-of-fit metrics (Stewart and Martell 2014). The use of multiple models provides a solution to the endless search for a better stock assessment model and allows for structural as well as estimation uncertainty to be better captured. The IPHC adopted the ensemble approach for its 2012 stock assessment (Stewart et al. 2013a) and has continued to develop and refine the set of models used to provide tactical management information each year. The ensemble approach integrates the results of multiple hypotheses with the uncertainty associated with parameter estimation (Stewart and Martell 2015). This reduces potential for abrupt changes in management quantities as improvements and additional data are added to individual models (Stewart and Hicks 2018), and provides a more realistic perception of uncertainty than any single model, and therefore a stronger basis for probabilistic risk assessment.

Development of the current ensemble of stock assessment models began in 2012 with a single model using three alternative fixed values of natural mortality (Stewart et al. 2013a). In subsequent years, ensemble development included exploration of highly varied model approaches, including a Virtual Population Analysis (VPA) and a simple biomass production model (Cox et al. 2014) and a spatially explicit model including migration rates and recruitment distribution (Cox et al. 2017). The treatment of the historical data through long and short modelled time-series', and the treatment of spatial patterns via coastwide aggregation of data and an Areas-As-Fleets (AAF) approach have emerged as two critically important axes over which to describe the uncertainty in both the scale and trends of the Pacific halibut stock and population dynamics. Therefore, recent ensembles have included four equally weighted models representing a two-way cross of time-series length (short and long) and data aggregation (coastwide and by Biological Region).

Starting with the final 2024 stock assessment data, models and results (Stewart and Webster 2025; Stewart and Hicks 2025), this analysis provides a sequentially updated 'bridge' of the changes made thus far toward a preliminary assessment for 2025. This bridging analysis included a series of steps for which intermediate results and comparisons are provided. These steps included:

1) Extending the time series to include projected mortality based on limits adopted for 2025 (IPHC 2025),
2) updating to the newest stock synthesis software version (3.30.23.1; Methot Jr 2024),
3) updating the time-series information for the Pacific Decadal Oscillation, used as a covariate to the stock-recruitment relationship,
4) retuning the constraint on the scale of male time-varying fishery selectivity (the sex-ratio of the commercial fishery) and extending this variability into the forecast,
5) improving the bootstrapping approach to pre-model calculation of maximum effective sample sizes to include ageing imprecision (Hulson and Williams 2024),
6) re-tuning the process and observation error components of these models to achieve internal consistency within each,
7) and updating the maturity ogive to reflect the recent histology-based estimates produced by the IPHC's Biological and Ecosystem Sciences Branch.

Briefly, extending the time-series, updating the software version, and updating the treatment of the fishery sex-ratio all had no of very little effect on the model results. Moving to the new PDO covariate increased the estimated spawning biomass in the coastwide long model and little effect on the AAF long model. Adding the new bootstrapping results and retuning the sample sizes and process error variance terms for internal model consistency generally increased the estimated spawning biomass, except for the coastwide long model near the end of the time-series where it was nearly unchanged. Finally, updating the maturity ogive resulted in a larger spawning biomass across all four models and especially in the historical period. Convergence, sensitivity and retrospective analyses were performed on all models contributing to the ensemble. The coastwide long model was most sensitive to the fixed value of steepness, with lower values corresponding to higher spawning biomass, while the other three models showed little difference at higher or lower values. The spawning biomass estimated by the coastwide short model scaled nearly linearly with the fixed value of M; higher M corresponding to larger estimates of spawning biomass. Excluding the PDO relationship resulted in a larger estimated spawning biomass in the coastwide long model across the entire time-series but had little effect on the recent years in the AAF long model. Evaluation of potential increased whale depredation in recent years was unable to explain the reduced recruitment observed since 2005. Retrospective analyses showed generally downward trends as data were added for all but the coastwide short model. Jitter analyses indicate that the AAF models were the least robust to a wide range of initial parameter estimates; however, there was no evidence that convergence was not achieved for the results provided here.

In aggregate, the results of the preliminary ensemble indicate that the uncertainty in stock dynamics remains similar to previous assessments and high relative to that frequently reported

for many single-model or simple stock assessment analyses. This uncertainty will continue to be captured via the annual decision table (Stewart and Hicks 2025), reporting the trade-offs between yield and various stock and fishery risks. Given the challenges and uncertainties of the Pacific halibut population dynamics and stock assessment it is unlikely that future assessment models will provide substantially more precise and stable results, even as data time-series grow longer. In light of the uncertainty and variability within which the Pacific halibut management occurs, a robust management procedure, tested via the IPHC's Management Strategy Evaluation (MSE) process (Hicks and Stewart 2025) may provide a stronger basis for future management success and stability than annual decisions based on stock assessment results.

Data sources

The Pacific halibut data sources are collected with sampling designs created to produce results first for each IPHC Regulatory Area, and then to be aggregated to Biological Regions and to the entire range of the species in U.S. and Canadian waters (Figure 1). This section provides a brief overview of the key types of data available for analysis. A more in-depth summary can be found in the annual overview of data sources created each year and most recently for the 2024 stock assessment (Stewart and Webster 2025).
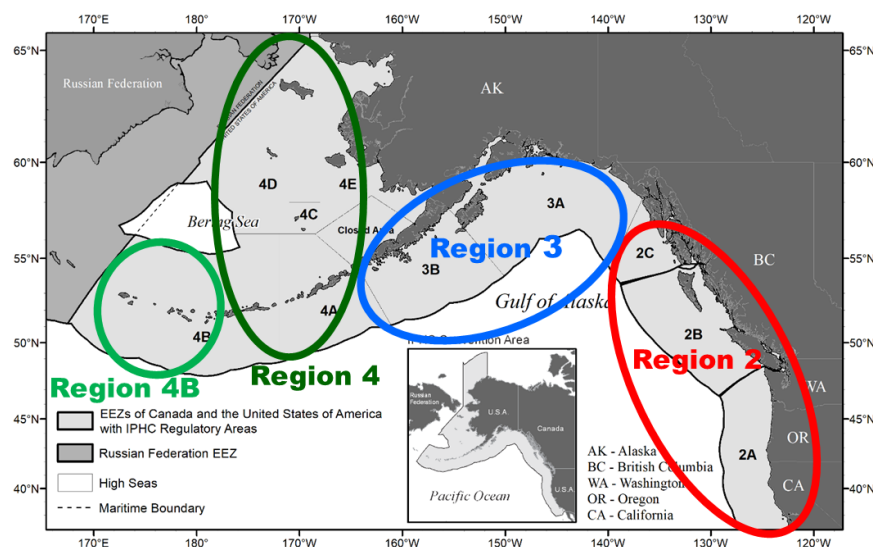

Figure 1. IPHC Regulatory Areas, Biological Regions, and the Pacific halibut geographical range within the territorial waters of Canada and the United States of America.

Overview of existing data

The time-series' of Pacific halibut data (described and plotted in much more detail in Stewart and Webster 2025) provide a rich historical record including mortality estimates, abundance indices (Catch-Per-Unit-Effort; CPUE) and age-composition data that extend back to the late 1800s and early 1900s (Figure 2). The IPHC's Fishery Independent Setline Survey (Ualesi et al. 2025; Webster 2025) provides the primary index of abundance and the most rich source of demographic information via individual weight, length and age data. The FISS includes Pacific halibut as young as 4-5 years old, which are below the IPHC's 32 inch (82 cm) minimum size limit (Stewart et al. 2021). Thus, these fish are observed several years prior to entry into the retained fishery landings which are sampled at the point of landing (Kong et al. 2022) and do not contain biological or catch-rate information on younger fish. Annual mortality estimates are provided to the IPHC from a variety of sources (Hutniczak et al. 2025) including the directed halibut fisheries (commercial, recreational and subsistence) as well as incidental mortality associated with discards in directed fisheries and discard mortality in non-directed fisheries ('bycatch') that are not allowed to legally retain Pacific halibut. Each of these sources have differing levels of precision and likely accuracy associated with the estimates used for stock assessment.
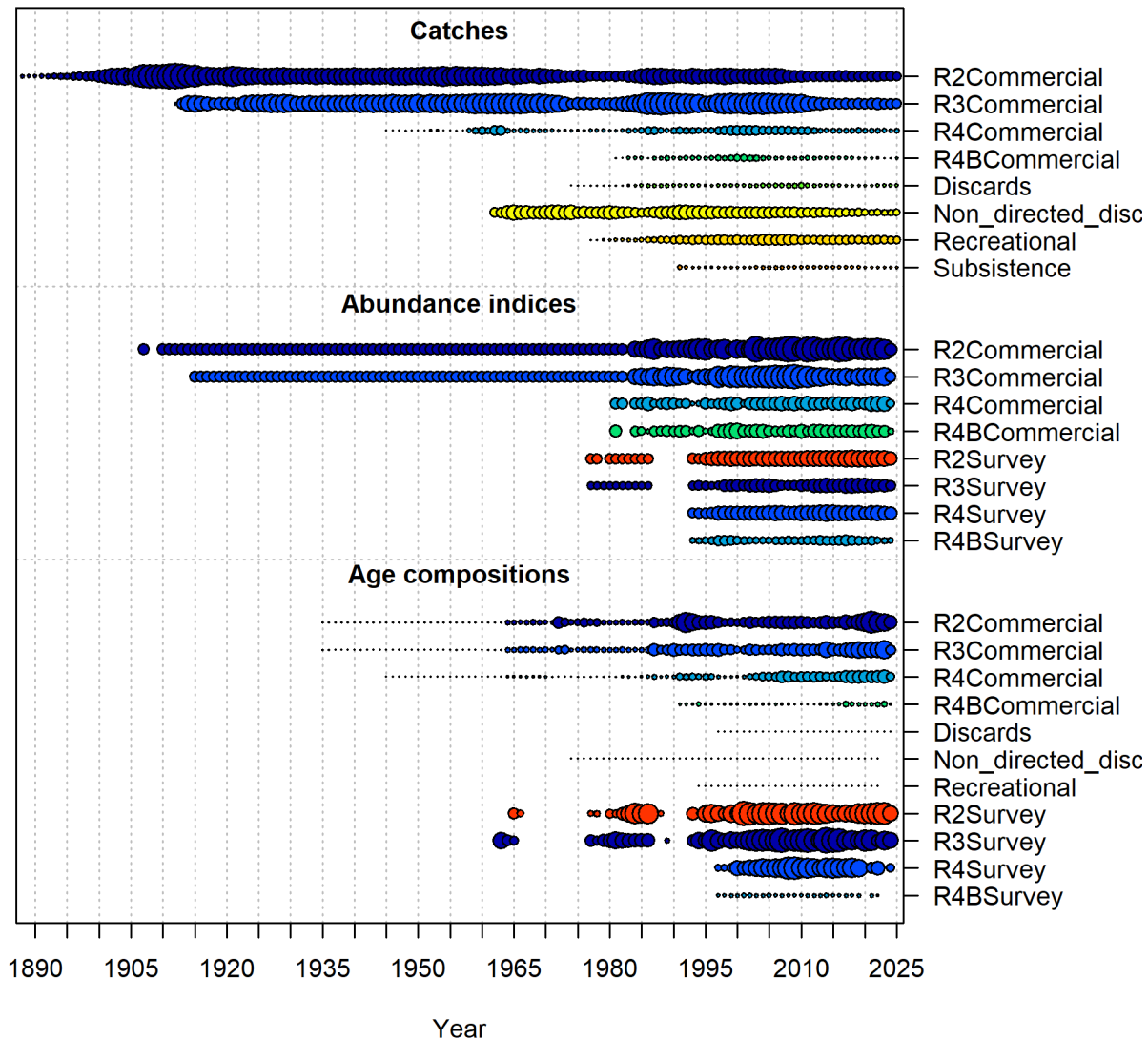
Figure 2. Data used in the stock assessment. Circle size is proportional to the magnitude of mortality (catches), inversely proportional to the variance (abundance indices) or proportional to the input sample size based on bootstrapping and prior to tuning for internal consistency (age-composition data).

Mortality

The industrial Pacific halibut fishery developed first off the west coast of the United States and Canada and sequentially moved to the north (Stewart and Webster 2025), only reaching full exploitation across all spatial areas in the last several decades. Mortality from non-directed discards increased rapidly with the arrival of foreign fleets into U.S. and Canadian waters in the 1960s. Recreational mortality has also increased over the time-series, although somewhat more gradually, since its initiation in the 1970s (Figure 3).

Figure 3. Time-series of mortality estimates by source.

Index data

The IPHC's FISS comprises the primary index of recent abundance and source of biological data for use in the stock assessment. Index values (Table 1) are used in this assessment in numbers of halibut captured per unit effort (NPUE). The recent time-series (1993-2024) is based on the output of the IPHC's space-time model (Webster 2025; Webster et al. 2020) which estimates the degree of spatial and temporal correlation among survey stations in order to predict trends in biomass and abundance across the entire r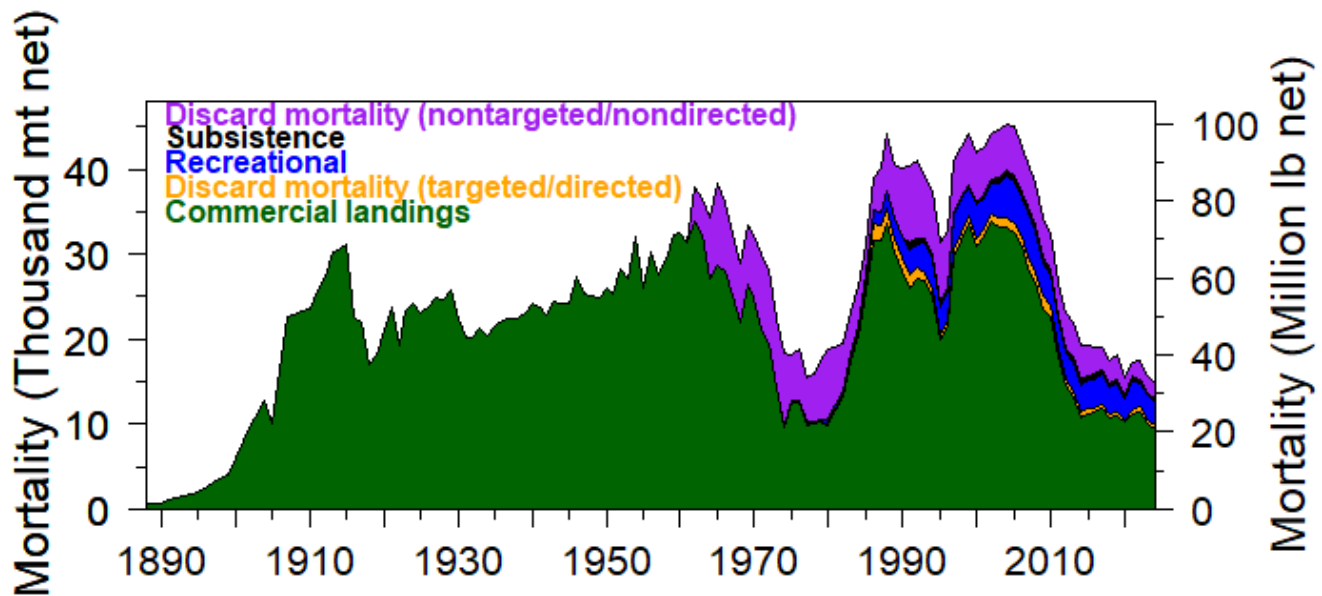ange of Pacific halibut within the IPHC Convention Area. This index provides precise trend information by IPHC Regulatory Area, even when annual sampling is reduced, which are weighted by the relative spatial bottom area and combined to Biological Regions and a coastwide index. The variances are summed, accounting for the square of the weights, and converted to log(SE) for use in the assessment model assuming log-normal error. There were geographically limited surveys conducting during 1963-1989, with summarized catch rates, but no variance estimates available from 1977 (Table 1). For the period prior to 1993 where there are no variance estimates, twice the recent average value is used, and for the coastwide series where spatial coverage is incomplete values are doubled again.

Commercial fishery CPUE (generally referred to as Weight-Per-Unit-Effort or WPUE as landings are recorded in weight) is reported through mandatory logbooks (voluntary only for vessels under 26 feet, 7.9 m, in length), collected by IPHC port samplers, or returned directly to the IPHC by mail. Commercial CPUE is available as far back as the early 1900s (Stewart and Webster 2025) providing a valuable historical record, but spanning a period of continuous fishery development and change, including an important transition to circle hooks in 1984 that substantially increased average catchability (Table 2-4).

Table 1. Modelled survey Numbers-Per-Unit-Effort (NPUE) and log(SE) 1993-2024, raw average observed NPUE 1977-1986; assumed values in italics.

Year	Region 2 Index	Region 2 log(SE)	Region 3 Index	Region 3 log(SE)	Region 4 Index	Region 4 log(SE)	Region 4B Index	Region 4B log(SE)	Coastwide Index	Coastwide log(SE)
1977	0.60	*0.107*	2.00	0.176	--	--	--	--	1.47	*0.227*
1978	0.80	*0.107*	1.30	0.176	--	--	--	--	1.11	*0.227*
1979	--	--	1.90	0.176	--	--	--	--	--	--
1980	1.20	*0.107*	2.50	0.176	--	--	--	--	2.01	*0.227*
1981	0.80	*0.107*	3.80	0.176	--	--	--	--	2.67	*0.227*
1982	1.84	*0.107*	3.80	0.176	--	--	--	--	2.87	*0.227*
1983	2.30	*0.107*	3.40	0.176	--	--	--	--	2.88	*0.227*
1984	6.74	*0.107*	11.60	0.176	--	--	--	--	9.30	*0.227*
1985	5.65	*0.107*	11.90	0.176	--	--	--	--	8.94	*0.227*
1986	4.54	*0.107*	7.80	0.176	--	--	--	--	6.26	*0.227*
1993	6.20	0.101	23.79	0.141	1.82	0.129	9.87	0.291	7.26	0.097
1994	7.41	0.104	23.42	0.120	2.10	0.111	10.12	0.263	7.56	0.080
1995	8.76	0.072	25.21	0.135	2.10	0.108	10.39	0.231	8.14	0.087
1996	7.76	0.059	26.40	0.173	2.32	0.094	10.59	0.182	8.36	0.111
1997	7.18	0.055	28.51	0.167	2.55	0.062	10.76	0.110	8.83	0.108
1998	6.20	0.054	24.60	0.084	2.65	0.063	11.10	0.111	7.98	0.054
1999	5.03	0.052	23.50	0.087	2.32	0.066	9.51	0.127	7.30	0.058
2000	5.58	0.055	25.26	0.080	2.47	0.062	8.43	0.151	7.77	0.053
2001	6.47	0.050	22.32	0.096	2.34	0.061	6.38	0.175	7.15	0.061
2002	6.40	0.051	24.42	0.069	2.22	0.059	4.66	0.205	7.40	0.047
2003	5.51	0.054	24.13	0.068	2.13	0.062	4.00	0.230	7.12	0.048
2004	5.02	0.052	27.55	0.067	2.13	0.059	3.73	0.214	7.70	0.049
2005	5.52	0.052	22.87	0.059	2.19	0.056	3.68	0.202	6.89	0.041
2006	5.42	0.052	21.79	0.087	2.25	0.049	4.21	0.197	6.72	0.057
2007	6.09	0.051	23.67	0.116	2.21	0.055	5.28	0.188	7.23	0.076
2008	6.14	0.050	21.34	0.127	2.49	0.061	5.30	0.167	6.94	0.079
2009	6.31	0.052	20.05	0.126	2.46	0.058	4.54	0.180	6.66	0.077
2010	6.11	0.050	20.17	0.087	2.32	0.052	4.28	0.179	6.55	0.055
2011	6.09	0.048	20.42	0.095	2.19	0.053	4.24	0.162	6.52	0.060
2012	7.26	0.047	21.10	0.060	2.14	0.050	3.82	0.154	6.79	0.039
2013	7.13	0.047	16.03	0.057	1.93	0.047	5.13	0.128	5.70	0.035
2014	7.30	0.046	19.06	0.056	1.97	0.044	4.47	0.135	6.32	0.036
2015	8.04	0.048	19.11	0.057	1.99	0.045	4.50	0.137	6.46	0.036
2016	8.09	0.046	19.44	0.064	1.88	0.049	5.14	0.122	6.51	0.040
2017	5.90	0.044	13.91	0.050	1.74	0.052	4.04	0.092	4.92	0.032
2018	5.23	0.041	12.69	0.050	1.66	0.052	4.05	0.128	4.53	0.031
2019	5.36	0.042	11.36	0.055	1.62	0.055	4.02	0.146	4.26	0.033
2020	5.05	0.042	11.76	0.054	1.57	0.070	3.95	0.176	4.26	0.036
2021	5.73	0.044	15.40	0.062	1.50	0.053	3.81	0.139	5.04	0.040
2022	5.81	0.045	12.74	0.080	1.45	0.049	3.58	0.204	4.48	0.048
2023	6.00	0.045	12.04	0.074	1.42	0.059	3.74	0.253	4.37	0.044
2024	6.64	0.066	12.20	0.111	1.38	0.074	3.90	0.290	4.49	0.066

Table 2. Commercial fishery Weight-Per-Unit-Effort (WPUE) 1907-1949 and log(SE); assumed values in italics.

Year	Region 2 Index	Region 2 log(SE)	Region 3 Index	Region 3 log(SE)	Region 4 Index	Region 4 log(SE)	Region 4B Index	Region 4B log(SE)	Coastwide Index	Coastwide log(SE)
1907	280.00	*0.100*	--	--	--	--	--	--	280.00	*0.100*
1910	271.00	*0.100*	--	--	--	--	--	--	271.00	*0.100*
1911	237.00	*0.100*	--	--	--	--	--	--	237.00	*0.100*
1912	176.00	*0.100*	--	--	--	--	--	--	176.00	*0.100*
1913	128.94	*0.100*	--	--	--	--	--	--	129.00	*0.100*
1914	124.13	*0.100*	--	--	--	--	--	--	124.00	*0.100*
1915	118.02	*0.100*	266.10	*0.100*	--	--	--	--	118.00	*0.100*
1916	114.60	*0.100*	202.80	*0.100*	--	--	--	--	137.00	*0.100*
1917	81.80	*0.100*	157.90	*0.100*	--	--	--	--	98.00	*0.100*
1918	87.50	*0.100*	125.40	*0.100*	--	--	--	--	96.00	*0.100*
1919	82.30	*0.100*	129.90	*0.100*	--	--	--	--	93.00	*0.100*
1920	84.10	*0.100*	147.90	*0.100*	--	--	--	--	96.00	*0.100*
1921	76.46	*0.100*	141.17	*0.100*	--	--	--	--	88.00	*0.100*
1922	62.44	*0.100*	133.79	*0.100*	--	--	--	--	73.00	*0.100*
1923	56.68	*0.100*	149.97	*0.100*	--	--	--	--	78.00	*0.100*
1924	55.39	*0.100*	109.13	*0.100*	--	--	--	--	74.00	*0.100*
1925	51.21	*0.100*	94.63	*0.100*	--	--	--	--	68.00	*0.100*
1926	51.67	*0.100*	93.73	*0.100*	--	--	--	--	67.00	*0.100*
1927	48.83	*0.100*	86.32	*0.100*	--	--	--	--	65.00	*0.100*
1928	47.27	*0.100*	72.34	*0.100*	--	--	--	--	58.00	*0.100*
1929	38.55	*0.100*	70.79	*0.100*	--	--	--	--	51.00	*0.100*
1930	34.44	*0.100*	65.91	*0.100*	--	--	--	--	46.00	*0.100*
1931	38.48	*0.100*	76.17	*0.100*	--	--	--	--	50.00	*0.100*
1932	47.50	*0.100*	83.49	*0.100*	--	--	--	--	60.00	*0.100*
1933	50.16	*0.100*	83.99	*0.100*	--	--	--	--	63.00	*0.100*
1934	54.07	*0.100*	74.97	*0.100*	--	--	--	--	62.00	*0.100*
1935	61.77	*0.100*	97.57	*0.100*	--	--	--	--	76.00	*0.100*
1936	54.66	*0.100*	96.70	*0.100*	--	--	--	--	71.00	*0.100*
1937	61.48	*0.100*	109.99	*0.100*	--	--	--	--	80.00	*0.100*
1938	70.33	*0.100*	114.29	*0.100*	--	--	--	--	88.00	*0.100*
1939	61.90	*0.100*	112.21	*0.100*	--	--	--	--	80.00	*0.100*
1940	61.71	*0.100*	116.38	*0.100*	--	--	--	--	81.00	*0.100*
1941	62.54	*0.100*	122.26	*0.100*	--	--	--	--	85.00	*0.100*
1942	65.43	*0.100*	132.54	*0.100*	--	--	--	--	90.00	*0.100*
1943	72.24	*0.100*	131.27	*0.100*	--	--	--	--	95.00	*0.100*
1944	86.84	*0.100*	149.23	*0.100*	--	--	--	--	110.00	*0.100*
1945	79.69	*0.100*	130.86	*0.100*	--	--	--	--	102.00	*0.100*
1946	83.78	*0.100*	123.82	*0.100*	--	--	--	--	101.00	*0.100*
1947	86.30	*0.100*	114.56	*0.100*	--	--	--	--	99.00	*0.100*
1948	88.61	*0.100*	112.20	*0.100*	--	--	--	--	99.00	*0.100*
1949	85.01	*0.100*	105.89	*0.100*	--	--	--	--	95.00	*0.100*

Table 3. Commercial fishery Weight-Per-Unit-Effort (WPUE) 1950-1991 and log(SE); assumed values in italics.

Year	Region 2 Index	log(SE)	Region 3 Index	log(SE)	Region 4 Index	log(SE)	Region 4B Index	log(SE)	Coastwide Index	log(SE)
1950	87.66	*0.100*	103.60	*0.100*	--	--	--	--	95.00	*0.100*
1951	87.63	*0.100*	108.93	*0.100*	--	--	--	--	96.00	*0.100*
1952	95.58	*0.100*	128.86	*0.100*	--	--	--	--	110.00	*0.100*
1953	128.65	*0.100*	134.32	*0.100*	--	--	--	--	131.00	*0.100*
1954	137.97	*0.100*	127.43	*0.100*	--	--	--	--	133.00	*0.100*
1955	122.20	*0.100*	116.32	*0.100*	--	--	--	--	119.00	*0.100*
1956	132.02	*0.100*	126.05	*0.100*	--	--	--	--	129.00	*0.100*
1957	100.95	*0.100*	119.84	*0.100*	--	--	--	--	110.00	*0.100*
1958	101.96	*0.100*	139.96	*0.100*	--	--	--	--	121.00	*0.100*
1959	98.67	*0.100*	160.62	*0.100*	--	--	--	--	129.00	*0.100*
1960	105.02	*0.100*	156.08	*0.100*	--	--	--	--	132.00	*0.100*
1961	96.00	*0.100*	159.79	*0.100*	--	--	--	--	127.00	*0.100*
1962	84.76	*0.100*	136.89	*0.100*	--	--	--	--	115.00	*0.100*
1963	77.73	*0.100*	123.89	*0.100*	--	--	--	--	105.00	*0.100*
1964	75.27	*0.100*	120.10	*0.100*	--	--	--	--	100.00	*0.100*
1965	86.47	*0.100*	107.07	*0.100*	--	--	--	--	99.00	*0.100*
1966	82.59	*0.100*	112.72	*0.100*	--	--	--	--	100.00	*0.100*
1967	81.44	*0.100*	113.00	*0.100*	--	--	--	--	101.00	*0.100*
1968	86.58	*0.100*	111.62	*0.100*	--	--	--	--	103.00	*0.100*
1969	81.53	*0.100*	105.07	*0.100*	--	--	--	--	95.00	*0.100*
1970	73.62	*0.100*	103.67	*0.100*	--	--	--	--	91.00	*0.100*
1971	76.05	*0.100*	96.31	*0.100*	--	--	--	--	89.00	*0.100*
1972	69.47	*0.100*	82.87	*0.100*	--	--	--	--	78.00	*0.100*
1973	64.41	*0.100*	62.13	*0.100*	--	--	--	--	63.00	*0.100*
1974	60.89	*0.100*	61.95	*0.100*	--	--	--	--	61.00	*0.100*
1975	61.87	*0.100*	66.76	*0.100*	--	--	--	--	61.00	*0.100*
1976	44.39	*0.100*	61.91	*0.100*	--	--	--	--	55.00	*0.100*
1977	64.17	*0.100*	65.57	*0.100*	--	--	--	--	63.00	*0.100*
1978	54.06	*0.100*	68.47	*0.100*	--	--	--	--	71.00	*0.100*
1979	55.80	*0.100*	67.33	*0.100*	--	--	--	--	75.00	*0.100*
1980	59.54	*0.100*	116.09	*0.100*	--	--	--	--	94.00	*0.100*
1981	73.84	*0.100*	148.86	*0.100*	136.84	*0.100*	99.00	0.078	111.00	*0.100*
1982	71.85	*0.100*	181.34	*0.100*	98.68	*0.100*	--	--	127.00	*0.100*
1984	151.95	0.045	491.33	0.046	386.90	*0.100*	161.00	0.103	316.00	0.035
1985	161.59	0.051	535.06	0.039	456.18	0.099	234.00	0.160	352.00	0.034
1986	137.26	0.035	506.00	0.042	308.70	0.062	238.00	0.372	315.00	0.041
1987	135.53	0.027	490.38	0.036	360.93	0.159	220.00	0.111	316.00	0.038
1988	168.40	0.054	560.55	0.042	405.68	0.105	224.00	0.122	363.00	0.036
1989	154.92	0.042	507.69	0.031	387.41	0.078	268.00	0.094	353.00	0.025
1990	194.64	0.043	403.54	0.036	370.26	0.095	209.00	0.103	315.00	0.029
1991	170.62	0.039	375.02	0.041	367.06	0.157	329.00	0.085	314.00	0.038

Table 4. Commercial fishery Weight-Per-Unit-Effort (WPUE) 1992-2024 and log(SE); assumed values in italics.

Year	Region 2 Index	Region 2 log(SE)	Region 3 Index	Region 3 log(SE)	Region 4 Index	Region 4 log(SE)	Region 4B Index	Region 4B log(SE)	Coastwide Index	Coastwide log(SE)
1992	167.66	0.040	413.39	0.048	324.01	0.117	280.00	0.095	315.00	0.035
1993	200.04	0.031	439.11	0.096	399.87	0.448	218.00	0.220	369.00	0.100
1994	175.74	0.027	362.77	0.049	343.14	0.333	197.00	0.101	302.00	0.069
1995	190.73	0.025	439.48	0.043	330.22	0.100	189.00	0.336	326.00	0.037
1996	208.81	0.042	505.01	0.046	427.58	0.138	269.00	0.185	387.00	0.039
1997	237.52	0.035	498.02	0.026	417.44	0.107	275.00	0.064	400.00	0.025
1998	221.23	0.029	512.59	0.036	411.86	0.089	287.00	0.058	402.00	0.025
1999	249.48	0.079	475.49	0.024	385.64	0.061	310.00	0.045	390.00	0.023
2000	227.94	0.036	492.21	0.025	403.74	0.082	318.00	0.046	396.00	0.020
2001	202.84	0.039	454.52	0.029	363.00	0.213	270.00	0.076	358.00	0.042
2002	214.81	0.032	466.46	0.025	296.56	0.082	245.00	0.081	356.00	0.020
2003	208.95	0.018	439.27	0.024	251.12	0.072	196.00	0.068	325.00	0.018
2004	192.88	0.028	425.79	0.026	235.23	0.072	202.00	0.061	315.00	0.019
2005	178.98	0.024	387.69	0.023	219.59	0.063	238.00	0.093	293.00	0.017
2006	180.22	0.024	360.70	0.022	178.26	0.064	218.00	0.111	268.00	0.019
2007	155.80	0.022	338.41	0.023	154.12	0.054	231.00	0.109	246.00	0.016
2008	135.02	0.018	314.08	0.022	162.55	0.071	193.00	0.069	227.00	0.018
2009	152.95	0.020	277.22	0.020	174.43	0.054	189.00	0.100	220.00	0.018
2010	185.68	0.034	242.32	0.024	143.97	0.079	143.00	0.062	203.00	0.020
2011	180.42	0.019	226.65	0.025	143.25	0.056	165.00	0.103	196.00	0.015
2012	193.96	0.020	214.08	0.032	137.37	0.074	149.00	0.067	193.00	0.021
2013	192.78	0.026	189.98	0.033	122.70	0.072	127.00	0.064	178.00	0.017
2014	210.44	0.026	182.93	0.039	116.04	0.092	146.00	0.070	183.00	0.022
2015	217.37	0.024	224.46	0.045	136.04	0.065	149.00	0.076	202.00	0.025
2016	212.66	0.019	216.22	0.044	128.30	0.066	123.00	0.083	196.00	0.020
2017	212.49	0.020	218.98	0.037	129.11	0.077	119.00	0.076	202.00	0.020
2018	195.67	0.027	189.88	0.055	115.12	0.058	134.00	0.071	177.00	0.028
2019	184.34	0.027	213.12	0.037	101.45	0.100	115.00	0.085	179.00	0.022
2020	175.50	0.024	215.84	0.040	100.92	0.081	105.00	0.059	178.00	0.022
2021	178.43	0.025	194.57	0.041	127.08	0.049	88.00	0.057	168.00	0.024
2022	154.56	0.024	133.57	0.038	122.22	0.049	80.00	0.085	134.00	0.022
2023	138.94	0.029	114.76	0.036	85.35	0.052	84.00	0.090	114.00	0.018
2024	135.08	*0.069*	105.23	*0.100*	80.98	*0.157*	115.00	*0.278*	112.00	*0.062*

Age data

At each FISS station, otoliths are sampled randomly at rates selected to generate approximately 1500 per IPHC Regulatory Area per year. The number of stations contributing to the annual age information varies considerably over the time-series, with Biological Region 3 the most heavily sampled, followed by Region 2, Region 4 and far fewer samples collected in Region 4B (Table 5). There are also a small number of geographically limited surveys from the period 1963-1966 for which there are age samples, but no corresponding index. Otoliths from the commercial fishery landings are also sampled in proportion to the weight of the catch with different rates by IPHC Regulatory Area (Hutniczak et al. 2025). This has led to a relatively larger number of commercial trips sampled in Biological Region 2 over most of the historical period, with Region 3, Region 4, and Region 4B each contributing fewer samples (Table 6-7).

Table 5. Number of stations contributing to FISS age data (1963-2024).

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1963	--	236	--	--	236
1964	--	305	--	--	305
1965	121	146	--	--	267
1966	66	--	--	--	66
1977	58	100	--	--	158
1978	62	98	--	--	160
1979	--	104	--	--	104
1980	80	101	--	--	181
1981	72	102	--	--	174
1982	154	148	--	--	302
1983	192	101	--	--	293
1984	241	198	--	--	439
1985	166	103	--	--	269
1986	178	97	--	--	275
1988	72	--	--	--	72
1989	--	33	--	--	33
1993	66	70	--	--	136
1994	14	147	--	--	161
1995	103	120	--	--	223
1996	198	424	--	--	622
1997	211	424	220	74	929
1998	228	507	100	42	877
1999	332	554	61	82	1,029
2000	239	548	149	83	1,019
2001	330	520	146	83	1,079
2002	313	555	154	82	1,104
2003	323	516	153	82	1,074
2004	327	523	145	70	1,065
2005	340	507	144	81	1,072
2006	317	526	240	84	1,167
2007	330	538	176	73	1,117
2008	338	549	166	76	1,129
2009	333	537	171	84	1,125
2010	333	521	172	76	1,102
2011	358	549	166	79	1,152
2012	354	522	168	71	1,115
2013	364	528	167	78	1,137
2014	381	556	227	76	1,240
2015	352	529	239	81	1,201
2016	350	538	220	72	1,180
2017	371	521	166	118	1,176
2018	466	537	167	77	1,247
2019	482	560	167	81	1,290
2020	370	494	--	--	864
2021	393	550	77	37	1,057
2022	321	266	117	27	731
2023	378	411	--	--	789
2024	175	157	55	--	387

Table 6. Number of commercial fishing trips contributing to fishery age data (1935-1982); historical values in italics are assumed.

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1935	*50*	*50*	--	--	*100*
1936	*50*	*50*	--	--	*100*
1937	*50*	*50*	--	--	*100*
1938	*50*	*50*	--	--	*100*
1939	*50*	*50*	--	--	*100*
1940	*50*	*50*	--	--	*100*
1941	*50*	*50*	--	--	*100*
1942	*50*	*50*	--	--	*100*
1943	*50*	*50*	--	--	*100*
1944	*50*	*50*	--	--	*100*
1945	*50*	*50*	*5*	--	*100*
1946	*50*	*50*	*5*	--	*100*
1947	*50*	*50*	*5*	--	*100*
1948	*50*	*50*	*5*	--	*100*
1949	*50*	*50*	*5*	--	*100*
1950	*50*	*50*	*5*	--	*100*
1951	*50*	*50*	*5*	--	*100*
1952	*50*	*50*	*5*	--	*100*
1953	*50*	*50*	*5*	--	*100*
1954	*50*	*50*	*5*	--	*100*
1955	*50*	*50*	*5*	--	*100*
1956	*50*	*50*	*5*	--	*100*
1957	*50*	*50*	*5*	--	*100*
1958	*50*	*50*	*5*	--	*100*
1959	*50*	*50*	*5*	--	*100*
1960	*50*	*50*	*5*	--	*100*
1961	*50*	*50*	*5*	--	*100*
1962	*50*	*50*	*5*	--	*100*
1963	*50*	*50*	*5*	--	*100*
1964	116	100	14	--	230
1965	118	106	12	--	238
1966	102	113	12	--	228
1967	125	133	20	--	278
1968	135	132	14	--	282
1969	113	102	12	--	227
1970	97	125	18	--	241
1971	82	77	9	--	168
1972	552	196	3	--	752
1973	311	262	5	--	578
1974	153	68	3	--	226
1975	234	76	7	--	320
1976	332	135	7	--	476
1977	247	138	7	--	401
1978	241	120	4	--	377
1979	125	101	6	--	244
1980	140	113	1	--	262
1981	146	90	7	--	248
1982	168	137	11	--	316

Table 7. Number of commercial fishing trips contributing to fishery age data (1983-2024).

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1983	133	106	23	6	268
1984	170	90	9	13	282
1985	171	99	14	2	286
1986	158	152	34	1	345
1987	531	498	76	12	1,117
1988	278	258	19	16	571
1989	318	371	39	24	752
1990	491	560	50	3	1,104
1991	718	496	62	12	1,288
1992	1,027	478	61	20	1,586
1993	959	471	65	11	1,506
1994	896	474	89	31	1,490
1995	887	468	72	37	1,464
1996	859	437	76	27	1,399
1997	676	429	183	58	1,346
1998	515	277	127	47	966
1999	454	303	118	24	899
2000	512	358	119	27	1,016
2001	505	233	117	13	868
2002	561	284	163	53	1,061
2003	545	266	118	49	978
2004	491	200	75	9	775
2005	461	193	125	13	792
2006	483	256	81	22	842
2007	429	218	95	12	754
2008	385	221	98	11	715
2009	432	240	68	14	754
2010	354	260	97	25	736
2011	383	224	83	14	704
2012	421	217	81	13	732
2013	455	196	73	14	738
2014	426	221	64	8	719
2015	476	192	119	15	802
2016	466	164	112	15	757
2017	410	175	106	17	708
2018	337	178	105	17	637
2019	409	199	116	10	734
2020	406	176	47	12	641
2021	379	160	43	11	593
2022	467	190	60	11	728
2023	495	220	64	9	788
2024	511	216	161	14	902

As has been the case since the 2015 stock assessment (Stewart and Martell 2016), all age data used in the stock assessment is aggregated into bins of ages from age-2 to age-25, with age 2 representing a 'minus' group including all fish of age 2 and younger, and age 25 representing a 'plus' group including all fish age 25 and older. For years prior to 2002 (except the survey ages from 1998 which were re-aged in 2013), surface ages were the standard ageing method,

replaced by break-and-bake in recent years. Because surface ages are known to be biased at older ages (Forsberg and Stewart 2015), the age data are aggregated at a lower 'plus' group, age 20+, for all years where this was the primary method.

Beginning with the 2019 stock assessment, sex-specific fishery age data has been available via the collection of fin clips and subsequent genetic assay based on sampling begun in 2017. The processing of these samples lags one-year, thus for the preliminary 2025 stock assessment there were seven years of sex-specific fishery age compositions used (2017-2023). They are compiled in an identical manner to the standard fishery age data, but delineating males and females through the weighting and aggregation up to Biological Regions and coastwide. The sex-specific fishery age compositions for 2024 will be available for the final 2025 stock assessment later this year, along with re-aged FISS data from 2023 and 2024. The re-ageing of the 2023 and 2024 samples was prompted by increasingly difficult ring identification in recent years leading to some patterns that appeared inconsistent with previous year's data (tracking of year-classes).

Other biological and fishery information
There are several other sources of information contributing to the stock assessment models. These include:

1) the time-series of the Pacific Decadal Oscillation (PDO) index
2) the maturity ogive
3) priors on natural mortality (M)
4) fecundity information
5) estimated weight-at-age
6) the length-weight relationship
7) ageing error (bias and imprecision)
8) data based 'priors' on bycatch, discard, and recreational selectivity

The PDO index and the maturity schedule were the subjects of specific development and sensitivity analyses reported in this stock assessment. There have been no significant changes to the treatment of other sources of information since the 2015 stock assessment (Stewart and Martell 2016), except for the length-weight relationship which was updated as part of the 2022 stock assessment based on an analysis conducted in 2021 (Webster and Stewart 2022). Directly measured weights have been collected during the FISS (since 2019) and the commercial sampling (since 2015) and used directly in the stock assessment data preparation. Therefore, the length-weight relationship is primarily used for estimation by domestic agencies of mortality in weight from piece counts (this is relevant to non-directed discard mortality, recreational mortality and subsistence mortality).

All other sources of information are updated (where appropriate) and described each year in the annual overview of data sources (Stewart and Webster 2022). For convenience, the treatment of each is briefly summarized in Table 8.

Table 8. Summary of other information sources contributing directly to stock assessment input files (Stewart and Webster 2025).

Input	Summary	Key assumptions
Pacific Decadal Oscillation index[1]	Monthly values averaged and compiled into a binary index for each year based on assignment to 'positive' and 'negative' phases. **Updated as part of the bridging analysis for 2025.**	Used as a binary indicator rather than annually varying values.
Maturity	Trimmed Generalized Additive Model based on calibrated visual estimates from 2002-2024; 50% female maturity at 11.0 years old. **Updated as part of the bridging analysis for 2025.**	Based on histological assessments, treated as age-based and time-invariant.
Priors on natural mortality (*M*)	Prior on age 3+ *M* based on longevity (Hamel and Cope 2022) and elevated values for ages 0-2 based on life history theory and analyses of other flatfish.	Age 55 is a reasonable proxy for longevity based on multiple observations of male and female halibut age 50 and greater.
Fecundity	Assumed to be proportional to body weight.	Temporal variability is included via changes in weight-at-age.
Weight-at-age	Reconstructed from survey and fishery information by Biological Region.	Historical variability has been similar for female and male Pacific halibut.
Length-weight relationship	Not used directly in the assessment, most of the historical data relies on a constant average length-weight relationship.	Measured weights are used preferentially where available.
Ageing error	Pacific halibut are relatively easy to age accurately and with a high degree of precision using the break-and-bake method (Clark 2004a, 2004b; Clark and Hare 2006; Piner and Wischnioski 2004). Surface ages are biased and less precise (Stewart 2014).	Multi-decadal comparison suggest that accuracy and precision have not changed appreciably over the entire historical record (Forsberg and Stewart 2015).
Bycatch selectivity prior	Age-distributions are created from weighted and aggregated length frequencies from a variety of sources and age-length keys from trawl surveys.	Due to incomplete sampling, poor data quality in many years, and other uncertainties, data are considered unreliable for estimation of recruitment.
Discard selectivity prior	Age-distributions of sub-legal (<32 inch) Pacific Halibut captured by the FISS are used as a proxy for poorly sampled directed commercial fishery discards.	Survey data may not be representative of commercial fishing behavior but are currently the only source of information on the age range of discarded fish.
Recreational selectivity prior	Weighted age-frequency data from the IPHC Regulatory Area 3A recreational fishery are the only comprehensive source available.	These data may not be representative of all recreational mortality but provide the best information currently available.

---

[1] Data can be accessed at: https://www.ncei.noaa.gov/access/monitoring/pdo/

External information on M

The approach taken to natural mortality remains unchanged from that in the 2022 stock assessment (Stewart and Hicks 2022). That analysis was thoroughly documented and is briefly summarized here. It is based on two primary concepts stemming from the 2021 CAPAM workshop (Hamel et al. 2023):

1) Although results are varied, simulations experiments have generally indicated that estimation of *M* is preferable to the use of fixed values, where this is possible. The use of informative priors is frequently necessary, with the most common prior based on longevity.
2) Elevated *M* at the youngest ages/smallest sizes should be expected due to increased size-dependent predation mortality.

For the 2022 Pacific halibut assessment both of these concepts were included in the four stock assessment models. An age independent prior on M was developed based on published meta-analyses (Hamel 2014; Hamel and Cope 2022), which uses the prediction interval based on a meta-analysis of the maximum observed age for a wide range of species. Both male and female Pacific halibut have been observed to age-55 (with multiple fish of both sexes exceeding age-50 indicating that this is likely to be an accurate estimate of longevity, and not an artifact of a single case of ageing imprecision). The prior median is given by:

$$M = \frac{5.4}{Age_{max}}$$

which results in a value of 0.0982, and a log(SD) of 0.438. With such a large variance, this prior is only weakly informative (Figure 4), but still may provide some stability for estimation of *M*.
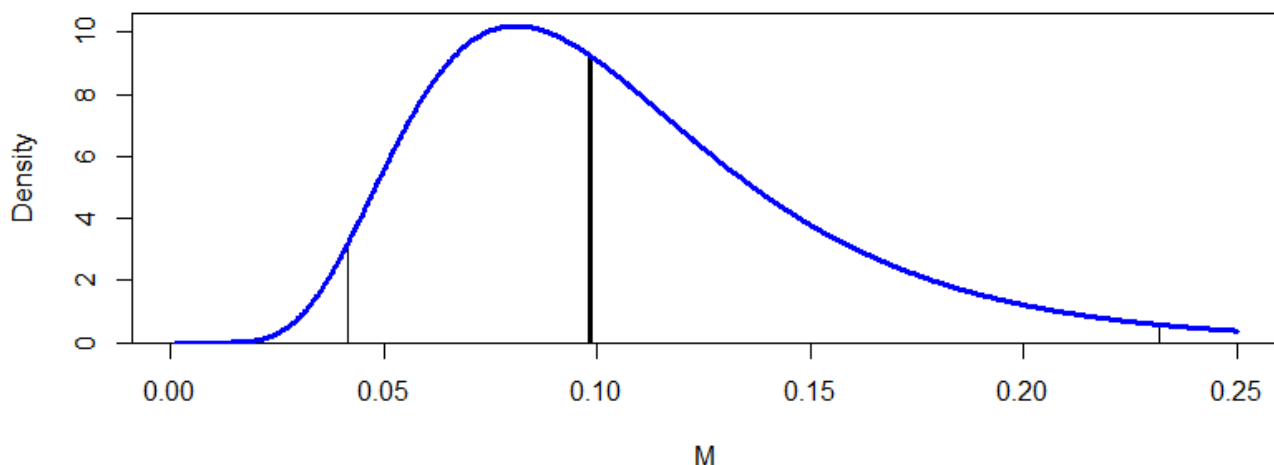
Figure 4. Informative prior for *M*. Thick vertical line denotes the median, thin lines the 2.5 and 97.5 percentiles of the distribution.

Pacific halibut were compared to other flatfish species via a summary of sex specific asymptotic size (either L_{inf} or L_{old}, depending on the parameterization) and *M* from all available Northeast

Pacific flatfish stock assessments (Stewart and Hicks 2022). Complete data were available for 26 stocks spread over four geographical regions comprising the U.S. West Coast, British Columbia, Gulf of Alaska, and the Bering Sea. A key result of this comparison was that flatfish with maximum sizes of >35 cm had natural mortality values both above and below those used for Pacific halibut. As Pacific halibut reach this size by age-3, this suggests that strongly elevated natural mortality due to predation common across flatfish species is likely to occur primarily below this age. Ecosystem models that include predator-prey dynamics generally suggest much higher M for the youngest age classes of NE Pacific groundfish (Adams et al. 2022). Where this information has been applied in other assessments used for management advice somewhat arbitrary scalars such as 1.5 x M for age 2, 2 x M for age-1 and 3 x M for age-0 are generally consistent with ecosystem models (e.g., Ianelli et al. 2021). Applying this general approach to Pacific halibut allows for size-dependent M that is consistent with theoretical concepts but does not appreciably change the natural mortality used for ages represented in observed fishery and survey data (exclusively age 2+). With little to no data at these youngest ages, any effect is likely to 'scale out' in the absolute estimates of recruitment deviations; however, when an index of recruitment is evaluated (i.e., the PDO in this assessment; see sensitivity analyses below) it may be important to include elevated M at these ages.

Bootstrapping input sample sizes for age compositions

Data weighting in the Pacific halibut stock assessment historically relied on the number of sampled FISS stations and number of sampled commercial fishery trips as a starting point for all models. Investigation of alternative tuning procedures and likelihoods was necessarily conditioned on those starting values, yet those starting values had not been evaluated specifically until the 2022 stock assessment. That analysis followed the method developed in Stewart and Hamel (2014) for estimating the maximum effective sample size based on the actual distribution and weighting of both the samples and the fish within samples. The effective sample size calculated in this manner is analogous to a minimum variance estimate – the actual effective sample size may be lower than calculated if not all strata are fully sampled (measurement error), of the source of the data differs from that assumed in the assessment (structural or process error). However, the effective sample size cannot be larger than the bootstrapped value simply due to the among and within sample variability and the sample sizes achieved. Thus, although time-consuming to produce, the approach provides an objective starting point for data weighting, and a logical upper bound on sample sizes used in the stock assessment models. Since the development of the 2022 stock assessment an improved method was developed that included ageing imprecision as part of the inherent variability in the observations (Hulson and Williams 2024). This new method not only resamples hauls (or trips) and fish within those hauls, but also individual ages from an empirical matrix of multiple age estimates from the same otoliths. Specifically, for a given observed age, a random age is drawn from all fish that had the original age assigned by one reader and a different age by a second read. In this way, the precision of each realization of the observed age composition information is reduced based on how likely that specific age would be given multiple reads. For Pacific halibut there are two ageing methods, break-and-bake (Table 9) and surface (Table 10), each with a differing degree of imprecision.

Table 9. Distribution of multiple reads of the same otoliths using the break-and-bake method.

First read \ Second read	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25+
2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	79	9	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	26	411	44	5	1	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
5	-	1	216	931	63	11	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	8	450	1,122	166	18	6	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	1	17	1,004	1,849	332	48	10	2	1	-	2	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	4	38	735	2,572	451	86	20	10	1	1	-	1	-	-	-	-	-	-	-	-	1
9	-	-	-	-	7	74	680	2,753	496	115	25	8	3	1	3	-	-	1	-	-	-	-	-	-
10	-	-	-	-	2	7	112	749	2,961	513	123	38	9	5	2	1	-	1	1	-	-	-	-	-
11	-	-	-	-	-	2	11	119	683	2,678	595	104	33	11	3	2	2	1	1	-	-	-	-	1
12	-	-	-	-	2	-	6	26	134	709	2,527	540	132	63	13	4	2	3	2	-	-	-	-	-
13	-	-	-	1	-	1	1	7	34	154	676	1,872	417	132	22	9	4	2	1	-	-	-	-	-
14	-	-	-	-	-	-	-	4	13	42	146	483	1,377	436	112	32	8	5	1	-	1	1	-	1
15	-	-	-	-	-	-	1	13	16	67	153	262	623	1,414	356	141	51	10	5	1	3	1	-	-
16	-	-	-	-	-	-	2	1	4	4	13	40	110	349	822	230	68	18	13	-	2	-	1	-
17	-	-	-	-	-	-	1	1	2	7	20	26	43	119	295	646	173	62	16	4	-	2	1	1
18	-	-	-	-	-	-	2	3	2	4	20	26	34	42	118	226	510	159	55	27	8	2	1	2
19	-	-	-	-	-	-	1	-	-	1	-	3	7	13	19	72	176	336	107	33	9	3	1	-
20	-	-	-	-	-	-	-	-	-	2	10	13	23	25	30	67	135	189	420	121	24	19	7	5
21	-	-	-	-	-	-	-	-	-	1	-	1	1	2	2	7	12	34	86	183	66	13	7	8
22	-	-	-	-	-	-	1	-	-	-	-	1	-	1	-	-	8	14	38	91	128	38	22	17
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	9	14	22	59	113	39	35
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	1	2	4	8	26	42	88	50
25+	-	-	-	-	-	-	-	-	-	2	1	1	1	-	3	5	17	19	17	28	30	48	82	801

Table 10. Distribution of multiple reads of the same otoliths using the surface method.

First read	Second read																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
2	439	72	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	22	1,111	108	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	54	1,123	74	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	1	45	420	79	9	1	-	-	-	-	-	-	-	-	-	-	-	-
6	-	1	6	49	803	95	16	2	2	-	-	-	-	-	-	-	-	-	-
7	-	-	-	15	140	1,197	196	24	7	3	1	-	-	-	-	-	-	-	-
8	-	-	-	2	22	245	1,802	343	65	12	2	-	2	1	-	-	-	-	-
9	-	-	-	2	2	45	378	2,460	434	84	22	3	3	-	1	-	-	-	-
10	-	-	-	-	3	13	77	470	2,348	411	97	31	16	3	3	-	-	-	-
11	-	-	-	-	1	3	13	102	479	2,068	489	120	34	13	2	1	-	1	-
12	-	-	-	-	2	2	8	26	88	438	1,963	492	113	49	16	6	1	-	-
13	-	-	-	-	2	-	2	7	18	100	442	1,468	373	95	15	6	3	1	2
14	-	-	-	-	-	-	1	1	2	25	86	302	1,031	250	59	24	12	7	4
15	-	-	-	-	-	1	-	1	3	5	29	76	200	577	156	42	20	7	1
16	-	-	-	-	-	-	-	-	2	-	8	20	46	142	335	91	25	15	11
17	-	-	-	-	-	-	-	-	-	-	8	4	19	33	91	209	69	20	21
18	-	-	-	-	-	-	-	-	-	1	1	-	5	18	24	58	131	42	23
19	-	-	-	-	-	-	-	-	-	-	-	1	2	2	6	15	44	79	43
20+	-	-	-	-	-	-	-	-	-	1	-	-	-	1	2	8	21	35	222

The results of this updated bootstrapping analysis indicated that the effective sample size across all composition data was approximately 1.5 times the raw number of samples collected, more variable and much lower than the 2022 analysis (Figure 7). There were differences between the fishery data and the FISS data and among geographical aggregations (Table 11) present with and without accounting for ageing imprecision, but the largest effect of the updated bootstrapping approach was on the historical data due to the lower precision of the surface method. On average, the effective sample size for the FISS data decreased by about 40%, the sexed fishery data by about 60% and the unsexed fishery data (including the historical surface ages) by about 80%.

Because the early fishery data are unavailable in current IPHC data bases, age compositions prior to 1991 were unable to be bootstrapped. Instead, the average relationship between the number of samples and the bootstrapped effective sample size was used to approximate effective sample sizes for use as starting values in the assessment models. Bootstrapped FISS (Table 12) and fishery (Table 13-14) maximum effective sample sizes are provided below.
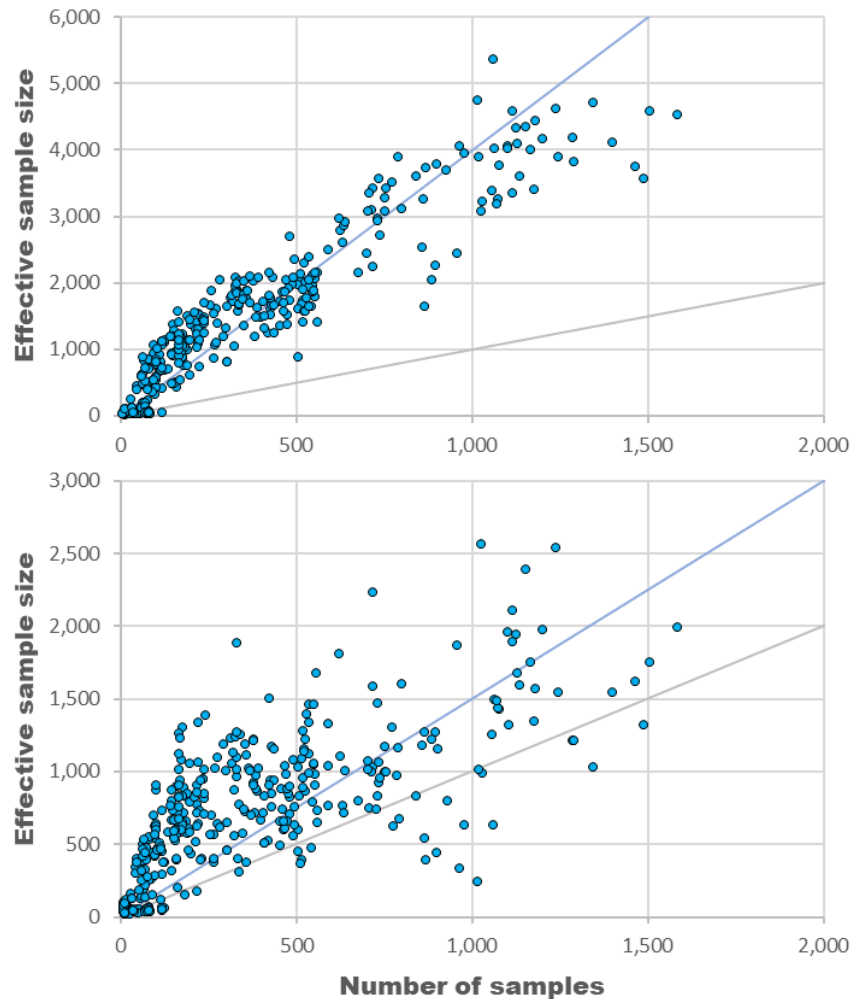
Figure 5. Number of samples vs. bootstrapped effective sample size for all FISS and fishery age compositions data. Upper panel indicates the relationship estimated for the age data used 2022 stock assessment, lower panel indicates the relationship estimated for the age data used in this stock assessment including ageing imprecision. Grey line indicates a 1:1 relationship, blue line indicates a 4:1 relationship (upper panel), or 1.5:1 relationship (lower panel). Note that the y-axes differ between the two panels.

Table 11. Summary of bootstrapping results by data type and spatial aggregation.

Data type	Aggregation	Mean effective N per sample
FISS	Coastwide	2.3
FISS	Region 2	3.6
FISS	Region 3	3.1
FISS	Region 4	4.5
FISS	Region 4B	0.6
All fishery	Coastwide	1.2
All fishery	Region 2	1.6
All fishery	Region 3	2.9
All fishery	Region 4	4.6
All fishery	Region 4B	3.2
Sexed fishery	Coastwide	1.6
Sexed fishery	Region 2	1.9
Sexed fishery	Region 3	3.8
Sexed fishery	Region 4	6.2
Sexed fishery	Region 4B	5.9

Table 12. Bootstrapped effective sample size for FISS age data (1963-2024).

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1963	--	943	--	--	943
1964	--	456	--	--	456
1965	371	316	--	--	371
1966	187	--	--	--	187
1977	126	419	--	--	521
1978	139	285	--	--	395
1979	--	453	--	--	453
1980	249	642	--	--	884
1981	212	901	--	--	1,076
1982	536	752	--	--	1,004
1983	849	622	--	--	1,184
1984	1,383	578	--	--	1,151
1985	1,121	597	--	--	1,022
1986	1,303	561	--	--	1,096
1988	129	--	--	--	129
1989	--	124	--	--	124
1993	527	436	--	--	584
1994	88	871	--	--	897
1995	871	729	--	--	1,063
1996	1,053	1,503	--	--	1,807
1997	818	907	175	48	796
1998	393	446	152	34	535
1999	962	905	252	58	989
2000	701	788	729	56	1,008
2001	1,881	956	663	70	1,429
2002	1,227	982	586	59	1,320
2003	868	1,088	661	43	1,435
2004	1,233	1,155	792	50	1,495
2005	1,249	1,024	765	59	1,483
2006	1,055	1,217	847	38	1,747
2007	1,095	1,463	832	43	2,108
2008	772	1,047	1,258	32	1,671
2009	1,266	1,123	1,230	33	1,943
2010	1,012	1,278	937	39	1,956
2011	1,111	1,456	1,011	40	2,387
2012	1,183	1,139	828	40	1,888
2013	1,021	866	673	42	1,592
2014	904	1,678	940	52	2,536
2015	734	1,396	1,014	38	1,970
2016	571	1,335	855	35	1,570
2017	693	888	823	48	1,344
2018	655	987	906	36	1,538
2019	897	731	751	42	1,212
2020	1,000	1,078	--	--	1,272
2021	1,020	1,049	275	44	1,250
2022	1,126	399	461	16	1,467
2023	1,211	831	--	--	1,164
2024	601	594	184	--	966

Table 13. Bootstrapped effective sample size for commercial fishery age data (1964-2011).

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1964	166	195	40	--	182
1965	169	207	34	--	189
1966	146	221	34	--	181
1967	179	260	57	--	220
1968	193	258	40	--	224
1969	162	199	34	--	180
1970	139	244	52	--	191
1971	117	150	26	--	133
1972	790	383	9	--	596
1973	445	512	14	--	458
1974	219	133	9	--	179
1975	335	148	20	--	254
1976	475	264	20	--	377
1977	354	270	20	--	318
1978	345	234	11	--	299
1979	179	197	17	--	193
1980	200	221	3	--	208
1981	209	176	20	--	197
1982	240	268	32	--	251
1983	190	207	66	--	213
1984	243	176	26	--	224
1985	245	193	40	--	227
1986	226	297	98	--	274
1987	760	973	218	--	886
1988	398	504	55	--	453
1989	455	725	112	--	596
1990	703	1,094	143	--	875
1991	1,586	754	355	73	1,210
1992	2,565	849	353	76	1,991
1993	1,865	977	382	44	1,746
1994	1,270	882	261	156	1,322
1995	1,215	1,040	324	52	1,617
1996	1,180	936	268	51	1,542
1997	799	1,170	152	31	1,031
1998	393	638	60	52	330
1999	491	644	68	45	439
2000	368	372	42	38	238
2001	600	395	115	22	386
2002	649	613	199	47	633
2003	476	560	291	28	629
2004	559	432	387	36	621
2005	736	460	373	32	671
2006	707	562	449	42	827
2007	753	767	698	29	1,167
2008	882	642	624	37	997
2009	846	816	464	20	999
2010	720	700	541	17	917
2011	717	731	528	21	1,010

Table 14. Bootstrapped effective sample size for commercial fishery age data (2012-2024). 2017-2023 represent bootstrapping of the sex-specific age data.

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
2012	522	623	544	20	832
2013	655	607	474	42	952
2014	937	1,338	415	32	2,231
2015	654	822	449	45	1,596
2016	601	724	508	83	995
2017	553	549	417	104	766
2018	399	611	466	61	760
2019	664	711	474	43	1,072
2020	860	628	394	54	1,102
2021	1,216	654	342	53	1,324
2022	878	655	378	67	1,024
2023	823	905	495	97	1,301
2024	833	510	387	25	1,149

Pacific Decadal Oscillation (PDO) Indices

The PDO (Mantua et al. 1997) was been identified and used as a covariate to the scale of Pacific halibut recruitment for decades (Clark and Hare 2002; Clark et al. 1999). Monthly values were averaged to generate an annual deviation which was then assigned a binary 'regime' (Figure 6). The previous approach[2] included 1900 through the present, but there were differences in how the most recent years had been calculated when compared with earlier years. In 2023, the methods for generating the time-series changed and were applied to a longer period. There is now a longer time-series available[3], including the period from 1854 to the present, using consistent methods across the entire period. However, in recalculating the index the period of years over which it was standardized was also changed, so not only do the individual years differ but the transition between regimes also differs, most importantly at the end of the time-series where all values after 1997 would be considered part of a negative regime (i.e., no more than three consecutive positive values; Figure 7). The transition between these two data sets is described below as part of the bridging analysis for 2025.

---

[2] Data from 1950 available here: https://oceanview.pfeg.noaa.gov/erddap/tabledap/cciea_OC_PDO.htmlTable?time,PDO
[3] Data available here: https://www.ncei.noaa.gov/access/monitoring/pdo/
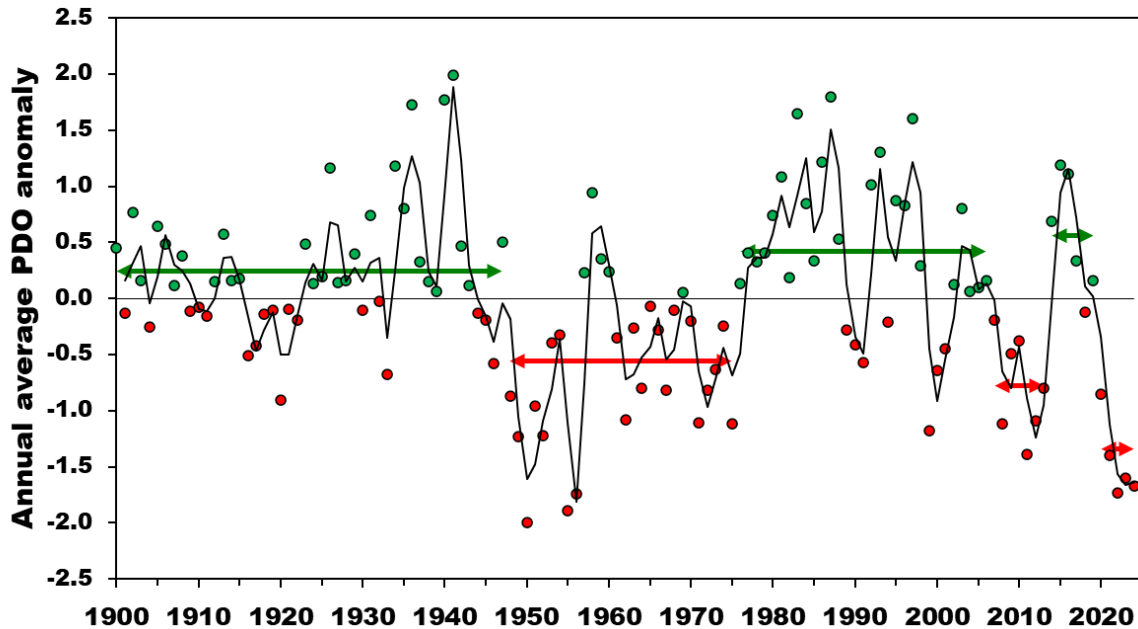
Figure 6. Time-series of PDO anomalies used in previous stock assessments. Lines with arrows indicate positive (green) and negative (red) regimes.
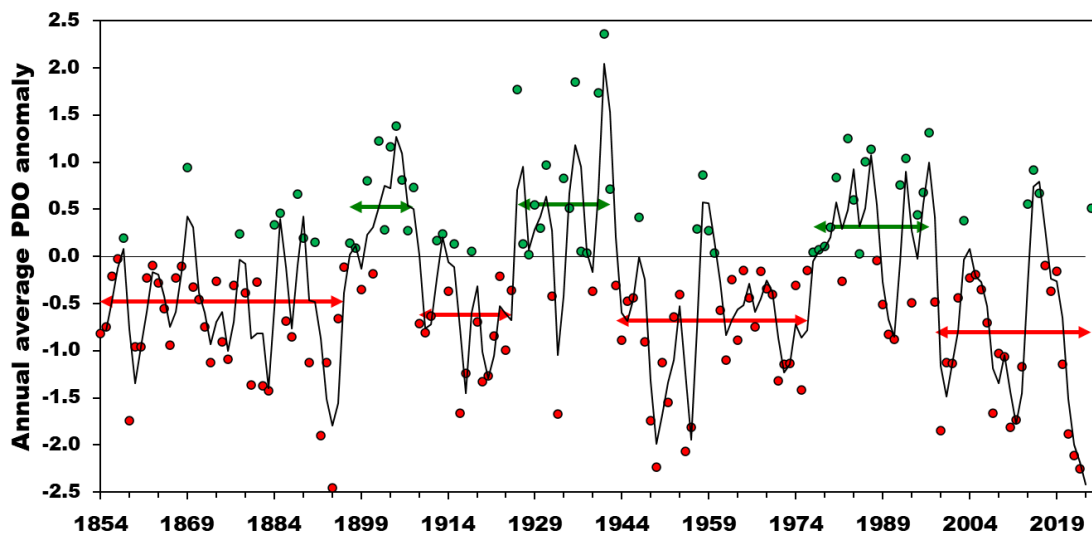
Figure 7. Updated time-series of PDO anomalies used in this stock assessment. Lines with arrows indicate positive (green) and negative (red) regimes.

Maturity

The maturity ogive used in Pacific halibut stock assessments, based on the work of Clark and Hare (2006), has remained unchanged since 2006. That analysis used visually estimated maturity from a subset of the IPHC Regulatory Areas collected from 2002-2004 to generate a logistic relationship, which was then truncated to be zero below the first age for which a fish had been observed to be mature (age 8). In 2025, an extensive analysis of histological

maturity was completed by the IPHC (*IPHC-2025-SRB026-06*). This is described in detail in the referenced document. Briefly, Histological maturity ogives were estimated with Generalized Additive Models (GAMs) fit by Biological Region to data collected between 2022 and 2024. This information was then used to calibrate the corresponding visual estimates of maturity, and those extending back to 2002. The reason for truncating this analysis in 2002 is the change in ageing method from break-and-bake (2002+) to biased surface methods (<=2001), which would create a bias in the maturity ogive for older fish. The curves for each Biological Region over the time-series were combined into coastwide ogives based on the relative abundance of Pacific halibut in each Region. Finally, a time-series average was calculated with and without truncating the youngest ages for which no Pacific halibut have been observed to be mature (Ages 5-6; Figure 8). The transition between the older curve and the newer curve is described below as part of the bridging analysis for 2025 and the effect on model results of truncating the youngest ages was explored as part of the sensitivity analyses.
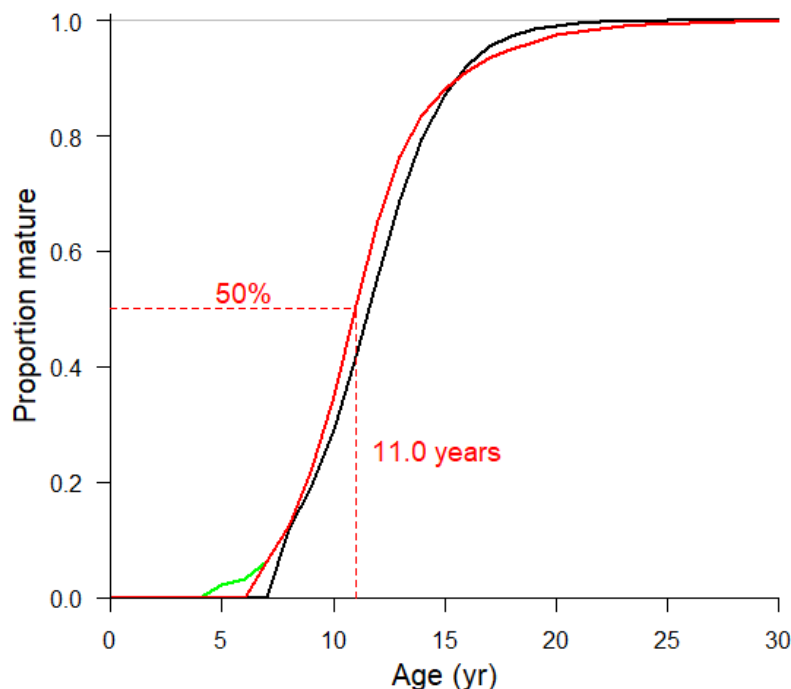
Figure 8. Historical maturity ogive from Clark and Hare (2006; black line), updated curve (green line) and updated curve truncated to be zero below age 7 (red line).

Model development

Multimodel approach

Creating robust, stable, and well-performing stock assessment models for the Pacific halibut stock has proven extremely challenging due to the highly dynamic nature of the biology, distribution, and fisheries (Stewart and Martell 2014). The stock assessment for Pacific halibut has evolved through many different modeling approaches over the last 30 years (Clark 2003; Clark and Hare 2006). These changes have reflected improvements in fisheries analysis methods, changes in model assumptions, and responses to recurrent retrospective biases and

other evidence of model mis-specification and concurrent degradation of model performance (Stewart and Martell 2014). Perhaps the most influential of these changes was the transition from separate IPHC Regulatory Area-specific assessment models to a coastwide model in 2006, as the understanding of adult movement among areas was substantially updated by the results of the IPHC's extensive PIT-tagging experiment in 2003-2009 (Clark and Hare 2006; Webster et al. 2013). Some simulation studies have found that dividing a migratory population into several discrete assessment units tends to overestimate the total biomass (e.g., Li et al. 2014; McGilliard et al. 2014).

Although recent modelling efforts have created some new alternatives, no single model satisfactorily approximates all aspects of the available data and scientific understanding. Building on simpler approaches in 2012 and 2013, in 2014, the current ensemble of four stock assessment models, representing a two-way cross of short vs. long time series', and aggregated coastwide vs. AAF models was developed for the full assessment analysis and review in 2015 (Stewart and Martell 2016). The models were further improved in 2019 to accommodate sex-specific age composition data from the commercial fishery (Stewart and Hicks 2019b) and again in 2022 to improve the input sample sizes for age composition data and to better inform estimates of natural mortality (Stewart and Hicks 2022).

AAF models are commonly applied when biological or sampling differences among geographical areas make coastwide summary of data sources problematic (Waterhouse et al. 2014). AAF models continue to treat the population dynamics as a single aggregate stock, but fit to each of the spatial datasets individually, allowing for differences in selectivity and catchability of the fishery and survey among regions. In addition, AAF models more easily accommodate temporal and spatial trends in where and how data have been collected, and fishery catches have occurred. This is achieved through explicitly accounting for missing information in some years, rather than making assumptions to expand incomplete observations to the aggregate coastwide level. Both aggregating the data into a single series and approximating spatial dynamics via AAF approaches may be useful under some circumstances; however, there is no clear best-performing configuration under all conditions. Not surprisingly, models that most closely match the biology, which is only known under simulated conditions, tend to perform the best (Punt et al. 2015).

To capture the structural uncertainty inherent among the Pacific halibut stock assessment models, it is necessary to use multi-model inference, here referred to as an 'ensemble' of models (e.g., Ianelli et al. 2016; Karp et al. 2018; Stewart and Martell 2015). The ensemble approach, applied in many fields in addition to fisheries (Du 2014; Hamill et al. 2012), recognizes that there is no "perfect" or "true" assessment model, and that a robust risk assessment can be best achieved via the inclusion of multiple models in the estimation of management quantities and the uncertainty about these quantities (Stewart and Martell 2015). This stock assessment is based on the approximate probability distributions derived from an ensemble of models, thereby incorporating the uncertainty within each model as well as the uncertainty among models. This approach reduces potential for abrupt changes in management quantities as improvements and additional data are added to individual models (Stewart and Hicks 2018), and provides a more

realistic perception of uncertainty than any single model, and therefore a stronger basis for risk assessment.

The current ensemble explicitly captures two critically important dimensions of uncertainty: how the time-series data are used via short and long models, and how the spatial information is treated in the models via data aggregation to the level of Biological Regions treated as separate fleets (AAF) or to the coastwide level. Inclusion of these sources of structural uncertainty results in wider confidence intervals than are commonly seen in single-model stock assessments (Stewart and Hicks 2019a). More detail on how the models are weighted and integrated can be found in the Ensemble section below.

Structural rationale

Consistent with analyses since 2015, this stock assessment is implemented using the generalized software stock synthesis (Methot and Wetzel 2013b), a widely used modeling platform developed at the National Marine Fisheries Service. This platform allows for a wide range of structural choices with regard to biology and growth, catchability, selectivity, spatial processes, stock-recruitment dynamics as well as error distributions and integrated projections. A benefit of using this code is that it is well documented, and the inputs and output formats are standardized (Methot Jr et al. 2021b), regardless of model configuration, allowing easy interpretation of model files and rapid evaluation of the results without re-running the fitting algorithm using the r4ss package (Taylor et al. 2021) implemented in the R programming language (https://cran.r-project.org/).

A primary structural stock assessment model choice is whether or not to model growth explicitly (and often parametrically) or empirically. Many stock assessments assert/estimate a growth function of some type and rely on this growth function to translate between numbers and biomass for model calculations. This approach has the benefits of allowing direct fitting to observed length observations, interpolating and/or extrapolating predictions for years where direct observations may be missing, as well as direct inclusion of the potential effects of selectivity at length on the observed data. The cost of such an approach is that growth can be an extremely complex process, varying over time, space and by cohort (via density dependence). When there is appreciable growth variability, a great deal of complexity may be required to adequately model this population process, even before sampling and selectivity issues have been addressed. Failure to account for this type of variability can lead to poor fits to composition data, potentially biasing the assessment results (Maunder et al. 2015, and subsequent special issue papers).

Pacific halibut show a very high degree of growth variability, with a 20 year-old female potentially as small as 60 cm or as large as 240 cm (Figure 9), males do not reach the same maximum sizes but also show a high degree of variability. Both sexes show nearly linear growth over their entire lifetime making the use of common parametric growth relationships (e.g., the von Bertalanffy curve) unreasonable. Previous efforts to fit to length data in the assessment model have shown little information content and a very high computational overhead.
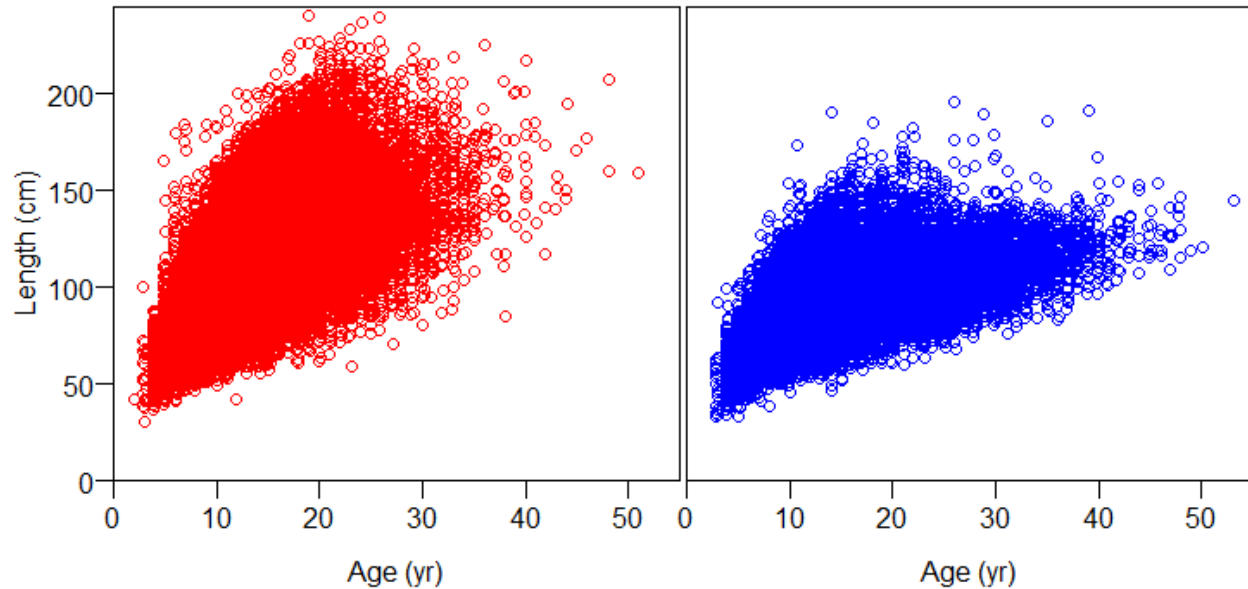
Figure 9. Observed length at age for female (left panel) and male (right panel) Pacific halibut collected by the FISS through 2023.

The Pacific halibut stock assessment models, like many other stock assessments with relatively complete age and size information, take a simpler approach to growth by using empirically derived weights-at-age. The empirical weight-at-age approach has the benefit of reducing complexity with regard to growth modelling but has several costs in other modelling areas. These include the need for more complexity in modelling selectivity, particularly where some of the selectivity process may be a function of size rather than age alone. This is the case for Pacific halibut, where the interaction of changes in size-at-age, gear selectivity that is likely at least partially a function of fish size, and a minimum size limit thus requires the treatment of selectivity-at-age as a time-varying process (Stewart and Martell 2014). However, the treatment of selectivity as time-varying appears to be a necessity for Pacific halibut even if treated as a function of size; static selectivity for a spatially aggregated model in the face of changes in availability was identified as a primary contributor to severe historical retrospective patterns (Stewart and Martell 2014).

There are relatively few examples of stock assessments used for management purposes that are explicitly spatial: modelling movement among areas, distributing recruitment events, and tracking spatial variability in biological characteristics (e.g., McGilliard and Palsson 2021; Stewart et al. 2009). Most such cases rely on low rates of movement to allow for estimation of recruitment distribution among areas. More frequently assessments either aggregate the available data across spatial heterogeneity (preferably weighting appropriately such that the aggregate information reflects the underlying distribution), or retain separate data series representing spatial areas, but fit to them in the context of a single instantaneously mixing population model (the AAF approach). These methods for dealing implicitly with spatial dynamics are by necessity gross approximations, with performance properties specific to a particular

application that are unknown, and almost certainly depend on the true underlying processes. Some simulation studies have shown that fisheries operating in different areas with differing selectivity schedules can be reasonably approximated by an AAF approach (e.g., Waterhouse 2014). Other studies have found acceptable performance of AAFs when simulating actual spatial variability (e.g., Hurtado et al. 2014, McGilliard et al. 2014); however additional studies have found that combining spatial data into weighted aggregates also performs acceptably and may be more stable than more complex AAF approaches (Punt et al. 2015, Li et al. 2015). A primary conclusion from simulation-based studies is that if the true underlying process is well-represented, then models reflecting these dynamics tend to perform well (Bosley et al. 2021; Goethel and Berger 2017). Unfortunately, for Pacific halibut it is not clear whether aggregated or AAF models might be the best choice as neither approach accurately represents the complex spatial dynamics.

The choice of how long a time-series to model generally represents a compromise among: data availability, data quality, model complexity, and technical convenience (e.g., data preparation and model convergence times). As assessment model time series' are extended to include more historical data, the quality of those data generally becomes increasingly lower as standardization of sampling programs has a greater likelihood of having changed appreciably. In the case of Pacific halibut, fishery-independent survey information has been reasonably comprehensive since approximately 1997, and sufficient to support the IPHC's geostatistical model since 1993 (Webster 2018; Webster et al. 2020). Current fishery sampling approaches have also not changed dramatically over the same period. The completeness of this time period with regard to data availability was one of the primary incentives for stock assessment models used by the IPHC since 2006 to begin the modelled period in 1996. Notable differences prior to that period included the transition in the survey and fishery from "J" to circle hooks (1984), variable and much less comprehensive survey coverage (<1997), lack of access to raw historical fishery data (ages, catch rates, etc.; <1991), and many others. The costs of using only a relatively short time-series include: a lack of integration between harvest strategy calculations derived from the full historical period, a lack of perspective on recent trends, the need for careful treatment of initial model conditions, inability to estimate some model parameters with only a shorter time-series, and increased sensitivity to additional data, as each year represents a greater fraction of the total information available in the model. These trade-offs prompted the development of the first long time-series model in 2013, with the recognition that neither the short or long time-series approach was clearly superior, and that differences in the results reflected a meaningful source of uncertainty in the assessment results.

All of the halibut models considered here treat male and female halibut separately. Like many broadcast spawning fishes, there is a basic assumption that spawning is likely to be limited primarily by female spawning output and not by male abundance (at least over a reasonable range of sex-ratios; this is generally not a concern except for cases such as some crab stocks where fishery mortality may operate primarily on males). If the sex-ratio could be expected to be stable over time, it might be reasonable to structure assessment models without regard to sex and/or just assume half of the mature biomass represented females. This is a common approach

for species where there is little dimorphic growth. However, for Pacific halibut, highly dimorphic growth interacting with gear selectivity for larger fish, and a fishery in which there are strong incentives to target the larger females (due to the minimum size limit and graduated price structure) results in sex-ratios of the catch and of the landings skewed largely toward females. Historical modelling suggested that the potential for a static assumption regarding sex-ratio could lead to a highly biased interpretation of stock status and that females and males are best modelled separately.

In aggregate, these considerations led to the choice of four stock assessment models during the 2014 assessment process: a two-way cross of: coastwide vs. AAF data structuring, and long vs. short time-series. Each of these models explicitly treated male and female halibut separately and employed empirical weight-at-age rather than an explicit growth function. All models fit to both fishery and survey index trends and age compositions and allowed for temporal variability in selectivity and catchability. Additional alternative modelling approaches were considered, including a simple surplus production model and a Virtual Population Analysis model. Both of these approaches suggested that recent removals and stock trends were on a similar scale to the four models included in that assessment (Stewart and Martell 2015) but presented sufficiently substantial issues in interpretation or application to the management process that they were not formally included in that stock assessment.

General model configuration

There are a number of basic technical settings and features that are common to all four stock assessment models described here. This section provides an overview, which is supplemented by a description of specific individual model details below.

The stock synthesis software separates inputs into several files read in prior to model estimation including the primary data file, the primary control file (including parameter setup and estimation switches), the weight-at-age file, the forecast file (including settings for reference point calculations), and the starter file (including some general estimation and reporting switches and settings). Each of these input files for each of the four stock assessment models described here are included in the background documents, along with the primary report file of estimated and derived quantities and the directory of summary and diagnostic figures created using the R package r4ss (Taylor et al. 2021). A full summary of supplemental material is provided in [Appendix A](). Note that not all automatically created diagnostic material, nor all of the model output is relevant to the model configurations employed here.

These models were configured to make use of relatively standard stock assessment practice in the population structuring. There were no seasonal dynamics, and catches were assumed to be removed halfway through the year via Pope's approximation. This approach does not require iterative estimation of fleet- and year-specific fishing mortality rate parameters (often reducing model run times) and should reasonably approximate the dynamics unless fishing mortality rates are extremely high or within year growth increments very large. Catches were input in thousands of pounds (net weight; head-off and gutted, approximately 75% of round weight), so that the mean weight-at-age inputs were in net pounds and the numbers-at-age are tracked in thousands

of individuals. Population dynamics contain ages 0-30, and female and male halibut are modelled separately in the underlying dynamics.

The input data were partitioned via a fleet structure of: the directed fishery (by area in the AAF models), discard mortality from the directed fishery, non-directed discard mortality ('bycatch'), recreational, subsistence, and survey (FISS; by area in the AAF models). Table 15 summarizes the data and key features of each model (note that changes from the 2024 model are described in greater detail below). Age data were partitioned by sex (the vectors for each year contain females, then males, such that the sex-ratio is inherently included in the age compositions), where this information was available and assigned the appropriate ageing method in the data file (see section above). Where few fish contribute to the 'tails' of the age distributions for each fleet and year combination, the model was set to automatically aggregate observations and predictions at each of the low and high ages with proportions less than 0.1%. This choice avoids large vectors of zeroes in the multinomial calculations. The model was also set up to add a very small constant (0.0001) to all age proportions in order to stabilize the computation.

All model growth specifications were bypassed in order to use the empirical weight-at-age approach; therefore, the settings in the control file and the results included in model outputs related to these settings are not meaningful (this includes length-at-age, weight-at-length, and maturity-at-length; these are all integrated directly in the weight-at-age inputs). The weight-at-age file also included a matrix of spawning output-at-age representing the product of annual weight-at-age (a matrix) and the vector of maturity-at-age (Stewart and Webster 2025).

For most estimated parameters, uniform priors were implemented, with bounds sufficiently wide to avoid maximum likelihood estimates falling on or very near a bound, unless the bound was structurally logical. Exceptions included process-error deviations, which are constrained by variance parameters, and the longevity-based prior on natural mortality, as described above. Table 16 summarizes the counts of estimated parameters in each model. Natural mortality was allowed to differ by sex, with the value for male halibut estimated in all four models, and the value for females in all but the short coastwide model. Treatment of both the stock-recruitment relationship and the initial conditions at the start of the modelled time-series differed among the four models and are described below.

The double-normal selectivity parameterization is used in all four models, as it represents a flexible, but still parametric approach that can easily be made time-varying via just one or two parameters with annual deviations. There are more flexible nonparametric selectivity options, but these generally require all the parameters to vary over time, creating a substantial increase in complexity. The double-normal selectivity can be easily configured to be either asymptotic or dome-shaped, by adjusting the width of the peak and/or descending slope and final selectivity parameters. It also includes an option for male selectivity to be offset from female selectivity, based directly on the parameters of the selectivity curve, such that time-varying selectivity for one sex can be mapped into temporal variability for both sexes without estimating a second set of deviation parameters. The double-normal was implemented for all model fleets, with at least

the ascending limb of selectivity (ascending width and peak parameters) allowed to vary over time for all four models (described further below).

Table 15. Comparison of structural assumptions among models.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Modelled period[1]	1992+	1888+	1992+	1888+
Data partitions	N/A	N/A	Regions 2, 3, 4, 4B	Regions 2, 3, 4, 4B
Directed Fishery fleets	1	1	4	4
Other fishing fleets	4	4	4	4
Survey fleets	1	1	4	4
Fishery CPUE (weight)	1992+	1907+	1992+	1907+, 1915+, 1981+, 1981+
Fishery age data years	1992+	1935+	1992+	1935+, 1935+, 1945+, 1991+
Survey CPUE (numbers)	1993+	1977+	1993+, 1993+, 1997+, 1997+	1977+, 1977+, 1997+, 1997+
Survey age data years	1993+	1963+	1993+, 1993+, 1997+, 1997+	1965+, 1963+, 1997+, 1997+
Weight-at-age	Aggregate	Aggregate	Areas 2, 3, 4	Areas 2, 3, 4
Female M	Fixed at 0.15	Estimated	Estimated	Estimated
Male M	Estimated	Estimated	Estimated	Estimated
Stock-recruit relationship	B-H	B-H	B-H	B-H
Initial conditions estimated	R_{init} N-at-age: 1-19	R_0, N-at-age: 1-29	R_{init}, N-at-age: 1-19	R_0, N-at-age: 1-29
Environmental regime effects on recruitment	No	Estimated	No	Estimated
Steepness (h)	0.75	0.75	0.75	0.75
$\sigma_{\text{recruitment deviations}}$	1.0	0.54	0.72	0.50
Survey selectivity	Asymptotic, by sex	Asymptotic, by sex	Domed, by sex (R2, R3) Asymptotic, by sex (R4, R4B)	Domed (R2, R3), Asymptotic (R4, R4B)
Fishery selectivity	Asymptotic, by sex	Asymptotic, by sex	Domed, by sex (R2, R3) Asymptotic, by sex (R4, R4B)	Domed, by sex (R2, R3) Asymptotic, by sex (R4, R4B)
Scale of male fishery selectivity	Estimated, time-varying	Estimated, time-varying	Estimated, time-varying	Estimated, time-varying
Non-directed discard selectivity	Domed	Asymptotic	Domed	Domed
Recreational selectivity	Asymptotic	Domed	Domed	Domed
Discard selectivity	Domed, by sex	Domed, by sex	Domed, by sex	Domed, by sex
Subsistence selectivity	Mirrored to recreational	Mirrored to recreational	Mirrored to recreational	Mirrored to recreational

[1]Mortality estimates for 2025 were projected based on adopted IPHC limits.

As has been the case in all recent halibut models, the catch-per-unit-effort index derived from the directed halibut fishery is included in each of the models, but the catchability is allowed to vary over time. In principle, there are many factors which can create changes in the proportionality of the catch-rate in a fishery with the underlying population. The most obvious of these are abrupt changes in fishing methods, such as the change from "J" to circle-hooks in 1984. This type of change was accommodated (in the long time-series models) via an unconstrained deviation on catchability in that year (effectively a separate q for the two parts of the time series). Beyond abrupt changes, there are many factors that can 'drift' over time but may not be so obvious, including technological improvements, changes in spatial areas or times of year being fished, targeting of areas with large vs. small fish, etc. This type of change suggests a random walk in catchability, which was the approach taken in all four models here. To implement this, a catchability parameter was estimated for the first year for which index data were available, and then a deviation (from the previous year's value, not the mean) was estimated for each subsequent year of the time-series. The annual catchability deviations were constrained by a single σ for each fleet. The iterative tuning algorithm for identifying the internally consistent values for each σ is described below.

In all models, fit to the age data used a multinomial likelihood with initial input sample sizes based on the revised bootstrap results described above, subsequently adjusted downward via a multiplicative scalar for each fleet in the control file (more discussion below). Indices of abundance from both the FISS and commercial fishery (by area in the AAF models) were fit using a log-normal likelihood and input log(SE)s based on the space-time modelling (FISS) or the between trip variability (fishery). Survey indices were fit in numbers of fish to avoid converting catch in numbers to weights in the data and then weights back to numbers in the model predictions (as informally recommended by the Scientific Review Board in 2014). Weight-per-unit-effort is the native scale for the fishery indices based on logbook records.

Using the method first developed for the 2015 assessment, discard mortality, bycatch and recreational selectivity are estimated, but the age composition data are down-weighted to avoid imparting any significant information on recruitment strengths from these uncertain and potentially non-representative data sets. In this way, the data that are available serve as an informative 'prior' on the selectivity for each of these fleets, and therefore propagate some uncertainty associated with selectivity estimation (vs. simply specifying selectivity as fixed parameters), but do not strongly inform other model parameters and population dynamics estimates.

Discards in the directed commercial fishery are treated as a separate fleet in each model. This approach was taken for several reasons: discard rates may be a function of spatial fishing effort and not simply contact selectivity as is often assumed in stock assessments - there has been little relationship between the magnitude of discards and the magnitude of commercial landings when this has been evaluated for previous reviews. Further, modelling discards with a retention curve in the empirical weight-at-age approach within SS does not allow for separate mean weight-at-age vectors to be applied to landings and discards (which may differ significantly for younger ages due to the size limit). Sex-specific selectivity curves were estimated in each model

informed by the observations from the sublegal fish captured by the setline survey. The selectivity was configured to be a double normal, with female halibut offset from male halibut to account for the dimorphic growth (the opposite of all other fleets), and the relative scale of females to males estimated directly. Both sexes were allowed to be dome-shaped, with differing descending limbs. Because the sublegal survey age data were already included in the likelihood as part of the survey age compositions, it would be a misrepresentation of the uncertainty to naively fit them again equally as part of the discard data set. Instead, previous analyses showed that down-weighting these data such that they had a very small input sample size had little appreciable effect on the model results but still allowed for the direct estimation of selectivity. This approach lends itself to direct inclusion of observer data on discarded halibut when/if sampling expansion methods that are representative of the entire fleet become available.

Table 16. Comparison of estimated parameter counts among models.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Static				
Female M	--	1	1	1
Male M	1	1	1	1
$Log(R_0)$	1	1	1	1
Initial R_0 offset	1	--	1	--
Environmental link coefficient	--	1	--	1
Fishery catchability	1	2	4	7
Survey catchability	1	4	--[1]	4
Fishery selectivity	5	5	21	20
Discard selectivity	8	7	5	5
Non-directed discard selectivity	4	2	3	3
Recreational selectivity	5	6	5	6
Survey selectivity	5	5	21	18
Total static	32	35	63	67
Time-varying[2]				
Recruitment deviations[3]	57	171	57	171
Fishery catchability deviations	37	118	148	322
Fishery selectivity deviations	85	200	345	668
Survey selectivity deviations	108	159	270	324
Total deviations	287	648	820	1,485
Total	319	683	883	1,552

[1]The analytic solution is used for this catchability parameter.
[2]Includes five uninformed forecast years, in order to propagate uncertainty.
[3]Includes deviations representing the initial age structure at the beginning of the modelled time-period.

Bycatch and recreational selectivity curves were also allowed to be dome-shaped given the relative frequency of younger halibut in the observed distributions. Where descending limb

parameters were estimated to be at the upper bounds, these parameters were fixed (making the curves asymptotic) to avoid any negative behavior during minimization and approximation of the variance in model quantities via the Hessian matrix. Since the 2019 assessment, sex-specific age composition data for the recreational fishery has become available (Stewart and Webster 2025), and so additional offset parameters were added to allow for sex-specific selectivity as in the treatment of the discards. Because of the down-weighting of the data for these series, and the unknown or potentially poorly spatially representative nature of the data themselves, no attempt was made to allow these selectivity curves to vary over time.

The presence of both observation error (in the indices and age composition data) and process error (in fishery catchability and selectivity for the survey and fishery) creates a challenge for standard weighting and tuning practices employed in many assessment models. Specifically, if process error is not modelled (and/or a fixed value is asserted), the input sample sizes (and sometimes index variances) can be relatively easily iteratively tuned or estimated (Maunder 2011). This approach is useful for reducing the potential effects of outliers, lack-of-fit, or model misspecification with regard to composition data (Francis 2011). At the other extreme, if the observation error is assumed to be known (and assigned a fixed value), then the degree process error can be estimated via random effects, or iteratively tuned using a maximum likelihood-based approximation (the 'Thompson and Lauth method'; Annex 2.1.1 in Thompson and Lauth 2012). When data are sufficient, both components can be iteratively, or by more statistically rigorous means, estimated simultaneously (Thorson 2019; Thorson et al. 2016).

The general goal for the treatment of process error in selectivity and catchability and observation error in the data is to first reduce clear signs of bias to the degree possible and then to achieve internal consistency among error distributions and sample sizes/variances. In all four models developed here, the initial input sample sizes, derived from the revised bootstrapping analysis described above were considerably larger than commonly applied weighting for stock assessment models would suggest (Table 12-14). These values were iteratively reduced based on evaluation of three considerations: the relative magnitude of the standardized Pearson residuals, comparison of the input value for each fleet with the harmonic mean effective sample size which is an unbiased estimator for a set of independent multinomial samples (Stewart and Hamel 2014), and the scaling suggested by the Francis (2011) method (as implemented in the r4ss package). For almost all fleets and all models, this approach led to a substantial reduction from initial sample sizes. In no cases were the input values increased from the maximum values derived via bootstrapping.

Starting from a small value for the input σ for each fleet and parameter combination where temporal variability was allowed, process error was increased until the tuned value was consistent with the degree of variability observed among the deviations (SE_{devs}^2) and the average uncertainty of the deviations themselves $\bar{\sigma}_{dev}^2$. This approach is very close to that outlined by Thompson and Lauth (2012) and is consistent with the preferred method for tuning this and other types of process error (such as recruitment deviations) in stock synthesis (Methot and Taylor 2011; Methot et al. 2019):

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

In addition to providing internal consistency, this approach makes intuitive sense: under perfect information the average variance of the deviations will be zero and the variability among the deviations will exactly match the process error, conversely, under no information the variance of the deviations will be the input constraint. After initial process error tuning, the input sample sizes were adjusted downward until the weights suggested by the fit to the mean age over the time series were approximately equivalent to the input values (the "Francis method'; Francis 2011). There were only minor changes to the tuned σ values required after iteration of the input sample sizes, suggesting the two processes were relatively separable and stable; further there were only minor changes in the process error variances in this assessment relative to the 2019 and 2022 assessment despite the revised input sample sizes.

As a final model-building step, models were regularized via adjusting parameterizations through removing and/or fixing selectivity parameters that consistently remain stuck to bounds or are not contributing to the likelihood in a meaningful way (<1% correlation with other model parameters). This regularization does not include forecast recruitment deviations, which are expected to be uncorrelated with other model parameters (and the objective function) but are 'estimated' in order to appropriately propagate the uncertainty in recent recruitments into forecasts.

The tuning approach for the stock-recruitment relationship was very similar, ensuring that the input σ governing recruitment variability was consistent with the observed variability and variance estimates; the calculation for this tuning is automated in the r4ss package, and the output was used as a guide for the scale of the bias correction, including ramps to and from the peak value consistent with the information content of the data and variability in the deviations observed in the output. This step is important for recruitment variability as it also provides for a better approximation for the bias correction in recruitment deviations (Methot and Taylor 2011) in the 'main' or best informed period of the time-series of recruitments. Again here, after initial tuning, little change was observed across alternative models or from previous results.

In the end, this tuning process provides a model that is internally consistent: the error distributions are commensurate with the fit to the data and the degree of process error is consistent with the signal (information content) in the data. Importantly, accounting for process error in selectivity was the primary solution for historically observed retrospective patterns in the Pacific halibut stock assessment models (Stewart and Martell 2014). Tuning diagnostics and results specific to each model are provided below.

Coastwide short
The initial conditions for a model starting after an extensive historical fishery and appreciable recruitment variability must be structured to avoid simple assumptions that may have strong effects on the subsequent time-series. For the coastwide short model, the initial conditions included estimating the population numbers at age 1-19 in the first year of the model (1992 after extension of the time-series; see below). Since the age data available for the initial year were aggregated at age-20 (due to the historical use of the surface ageing method), there was no

specific information on additional individual year-classes. To accommodate a non-equilibrium value in the plus group, an offset to initial equilibrium recruitment (via a single time 'block') was also estimated. The effect of these two approaches was to essentially decouple the numbers-at-age at the beginning of the time-series from any equilibrium assumptions.

As in previous assessments, the coastwide short model employed a Beverton-Holt stock recruitment relationship with estimated equilibrium recruitment level (R_0) setting the scale of the stock-recruit relationship. Steepness (h) was fixed at a value of 0.75 for this and all other models, an assumption that has been explored extensively in previous assessments. Fixing steepness, but iteratively solving for the internally consistent level of recruitment variability generally does not have a large effect on year-class strengths where data are informative, but does have very strong effects on direct estimates of Maximum Sustainable Yield (Mangel et al. 2013); however, this quantity is not of specific interest for the Pacific halibut assessment. A summary of the number of estimated parameters contributing to each aspect of the model is provided in Table 16.

Age-based selectivity for female halibut for both the FISS and commercial fishery was estimated using the double normal, forced to be asymptotic once it reached peak selectivity. This required two parameters: the ascending width of the curve and the age at which the peak selectivity is reached. Both parameters are allowed to vary over time with a random walk of annual deviations. These deviations were initiated in the first year for which age composition data were available, and extend into the forecast period (three years) to propagate the variance associated with potential future changes in selectivity. Male selectivity for the survey was estimated via offsets to the female ascending width and peak parameters, and a third parameter defining the scale of male selectivity relative to that for females. Male selectivity offset parameters for the fishery allow for the time-varying process to apply to both males and females with only two additional parameters. The scale of male selectivity for both the survey and fishery were allowed to vary over time as a random walk. For the fishery, these deviations are estimated beginning in 2018, since the sex-specific age composition derived from genetic analyses begins in 2017. In aggregate, there were five estimated base parameters each for the survey and fishery and annual deviations on the ascending limb parameters (Table 16).

Coastwide long
Initial conditions for the coastwide long time-series model include the initial age structure and a long period of uninformed recruitments with the model period beginning in 1888 and the first age data available for 1935 (Table 15); therefore, there was a substantial 'burn in' for recruitment variability prior to any data. The treatment of the stock-recruitment function in the coastwide long model was substantially different from that of the coastwide short model. Consistent with historical IPHC analyses (Clark and Hare 2002, 2006) and previous stock assessments, the coastwide long model allowed for the possibility that the scale of the stock-recruitment function is correlated with the regimes of the Pacific Decadal Oscillation (PDO; Mantua et al. 1997). To implement this approach, a Beverton-Holt relationship was used, parameterized with an estimated value for the equilibrium recruitment level (R_0) parameter, and a fixed value of steepness (h) of 0.75. The annual average of the PDO index (see description above for updates

to this index) was converted to a binary indicator (PDO_{regime}) where productive regimes (e.g., 1977-1997) were assigned a value of 1.0, and poor regimes (e.g., 1943-1976) a value of 0.0. These regimes were linked to the scale of the stock-recruit function via an adjusted equilibrium level of recruits (R_0') based on an estimated coefficient (β) creating an offset to the unadjusted value:

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

The adjusted equilibrium recruitment value was then used in the stock-recruit function with bias-corrected annual deviations:

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

This parameterization has the desirable property that if there is no correlation between the putative environmental index and underlying mean recruitment, the β parameter will be estimated at a value of 0.0 and the recruitment estimates will be unaffected. In that case R_0' is simply equal to R_0. As was the case for the coastwide short time-series model, fixing steepness precludes the naïve use of MSY estimates.

The approach to selectivity in the coastwide long model was identical to that in the coastwide short model. Selectivity deviations on the ascending limb parameters of the fishery and survey series were initiated in the first year for which age composition data were available for both the fishery (1935) and the survey (1963).

Natural mortality (M) is estimated separately for males and females in the coastwide long model using the informative prior described above.

AAF short
The AAF short model was configured very similarly to the coastwide short model. The most notable difference was in the treatment of selectivity for the survey and fishery in Biological Regions 2 and 3: these were allowed to be dome-shaped relative to the coastwide population dynamics. Implementing dome-shaped selectivity for these four model fleets requires the addition of a third selectivity parameter defining the width of the descending limb. This additional parameter was not allowed to vary over time. Similar to the coastwide long model, the three parameters defining the annual male offset to female selectivity for the commercial fishery in each area were only estimable beginning with the 2017 sex-ratio data. Temporal variability in selectivity parameters occurred over a slightly longer range of years in the AAF short model, as there were Region-specific survey data available for the entire time-series from Biological Regions 2 and 3. Beginning with the 2022 assessment, the AAF short model estimates female and male M.

AAF long
The only structural differences between the AAF long and AAF short models were the years over which deviations in recruitment, selectivity and catchability are estimated. The AAF long model

treated the stock-recruitment function in the same manner as the coastwide long model, including the PDO as an estimated covariate to equilibrium recruitment.

Changes from 2024

In the intervening period between the last full stock assessment analysis in 2022 and this preliminary analysis for 2025, the length and information content of the data sets has grown, and new information, such as the revised bootstrapping results (described above) has become available. Changes to specific data sets have been documented in the recent assessments and their effects evaluated individually in each year (Stewart and Hicks 2024, 2025). Key changes for 2025 included:

1) Extending the time series to include projected mortality based on limits adopted for 2025 (IPHC 2025),
2) updating to the newest stock synthesis software version (3.30.23.1; Methot Jr 2024),
3) updating the time-series information for the PDO, used as a covariate to the stock-recruitment relationship,
4) retuning the constraint on the scale of male time-varying fishery selectivity (the sex-ratio of the commercial fishery) and extending this variability into the forecast,
5) improving the bootstrapping approach to pre-model calculation of maximum effective sample sizes to include ageing imprecision (Hulson and Williams 2024),
6) re-tuning the process and observation error components of these models to achieve internal consistency within each,
7) and updating the maturity ogive to reflect the recent histology-based estimates produced by the IPHC's Biological and Ecosystem Sciences Branch.

The sequential effects on the model results of each of these changes are described below as a 'bridging' analysis from the 2024 stock assessment.

Extending the time-series

In order to provide for transparent comparisons from this preliminary stock assessment through the final results for 2025, the initial step in this analysis was to extend the modelled time-series to 2025, using the projected mortality associated with the limits set by the IPHC (IPHC 2025). Weight-at-age was assumed to remain constant from 2024 to 2025; however, it will be updated prior to the final 2025 sock assessment when the new data become available. No other information was needed for this single year projection and all model results and parameter estimates remained unchanged relative to the final 2024 stock assessment.

Software version update

The Pacific halibut stock assessment has updated to newer versions of the stock synthesis software (Methot and Wetzel 2013a; Methot and Wetzel 2013b) as new features have been added, and in order to avoid major changes as input/output changes have evolved over time. The 2024 stock assessment was implemented in version 3.30.22.1 (Methot et al. 2024), which was updated to 3.30.23.1 (Methot Jr 2024) for the 2025 stock assessment. The results were unaffected as there were no changes made that were related to any of the features used for the

analyses of Pacific halibut; therefore, for simplicity, this step has been omitted from the bridging figures below.

PDO index

As described above, a revised PDO index using consistent methods for an extended time-series (1854-present) is now available. In order to compare how this new series explained the historical recruitment both the effect size (% difference in average recruitment over positive and negative regimes) and the Standard Deviation (SD) of the recruitment deviations were summarized for the previous index and the updated index. A lower SD implies less residual variability in recruitment and conversely more of the process explained by the underlying stock-recruitment curve and the environmental effect.

The effects of updating the PDO index differed between the CW long and AAF long models. For the coastwide long model, the updated index resulted in a larger regime effect (62% vs 59% higher average recruitment during a positive regime) and a slightly lower SD of the estimate recruitment deviations (0.364 vs 0.375; Table 17). For the AAF long model the updated index resulted in a slightly lower effect size (50% vs 53%) and no change in the SD of estimated recruitment deviations (Table 17). The estimated historical time-series of spawning biomass was adjusted to better align with the revised regime definitions in both models (Figure 10-11). Over the most recent portion of the time-series the CW long estimated spawning biomass was scaled downward, while the AAF long estimated spawning biomass was virtually unchanged (Figure 11-12). In aggregate, there was no strong support for remaining with the previous index and therefore the index was updated for the preliminary 2025 stock assessment. Further evaluation of the treatment PDO is provided as part of the sensitivity analyses described below.

Table 17. Comparison of effect size and SD of the recruitment deviations for the CW and AAF long models for the previous and updated PDO indices.

Model	Positive regime effect (increase)		SD of recruitment deviations	
	AAF long	CW long	AAF long	CW long
Previous index	53%	59%	0.322	0.375
Updated index	50%	62%	0.322	0.364
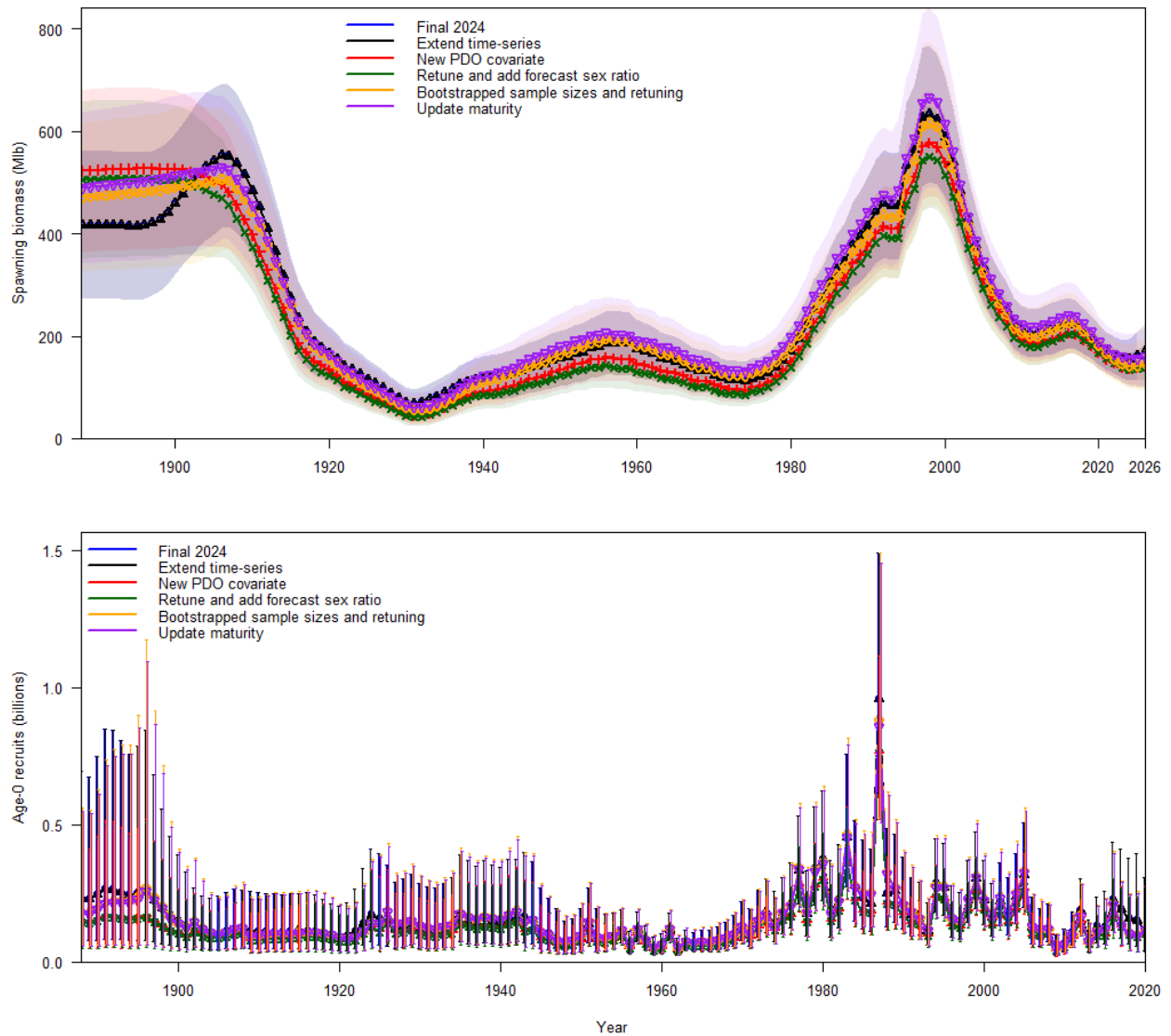
Figure 10. Comparison of estimated spawning biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2024 to preliminary 2025 coastwide long models.
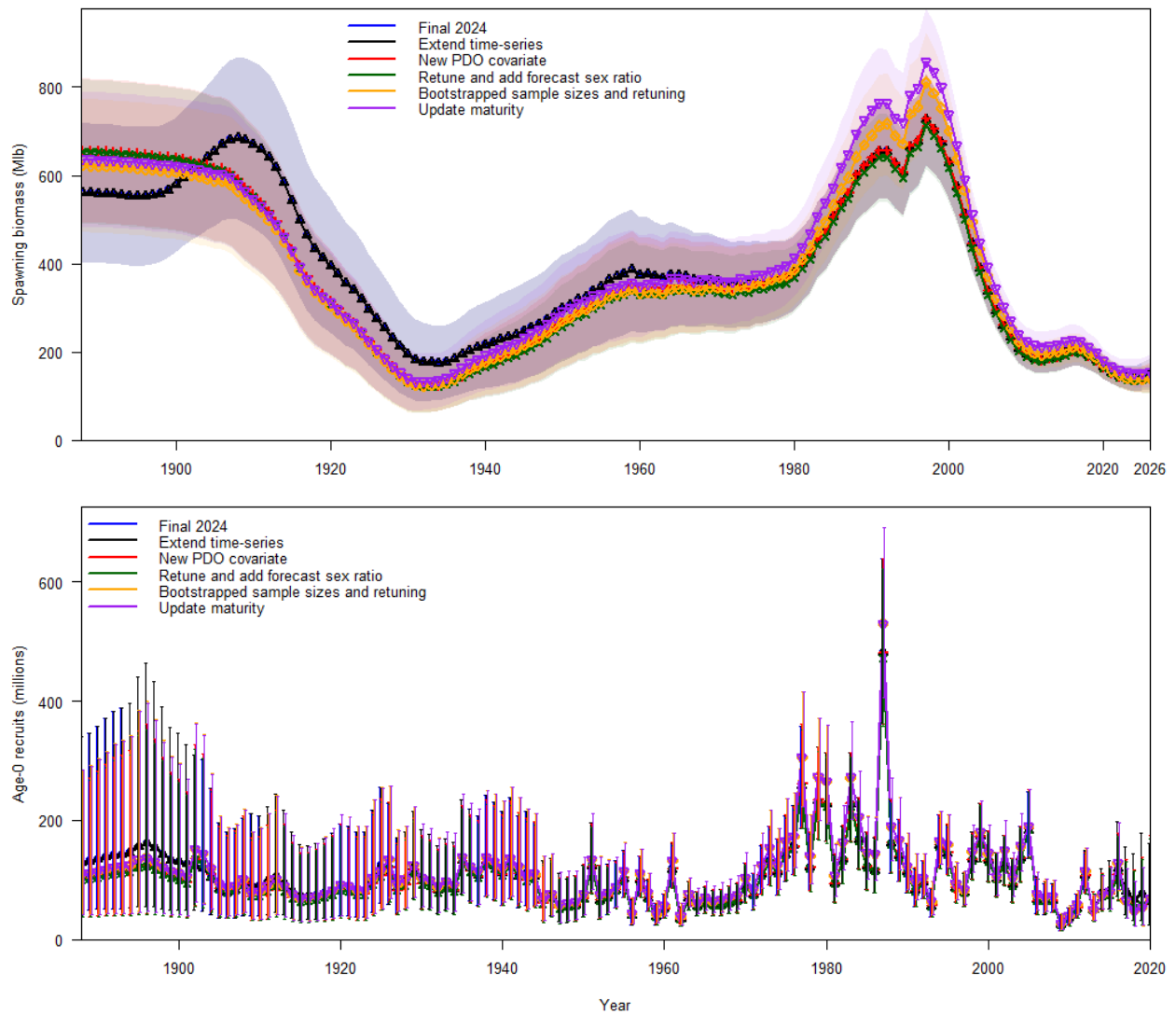
Figure 11. Comparison of estimated spawning biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2024 to preliminary 2025 AAF long models.
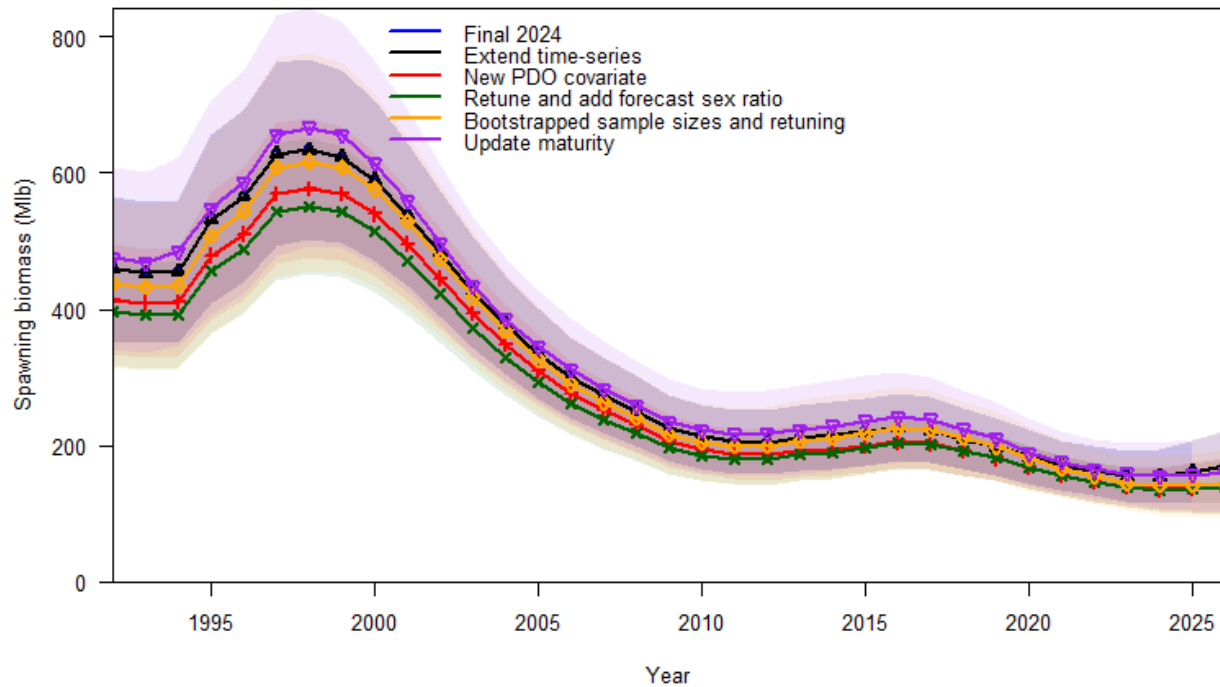
Figure 12. Comparison of recent estimated spawning biomass (1992-2026) over sequential changes from the 2024 to preliminary 2025 coastwide long models.
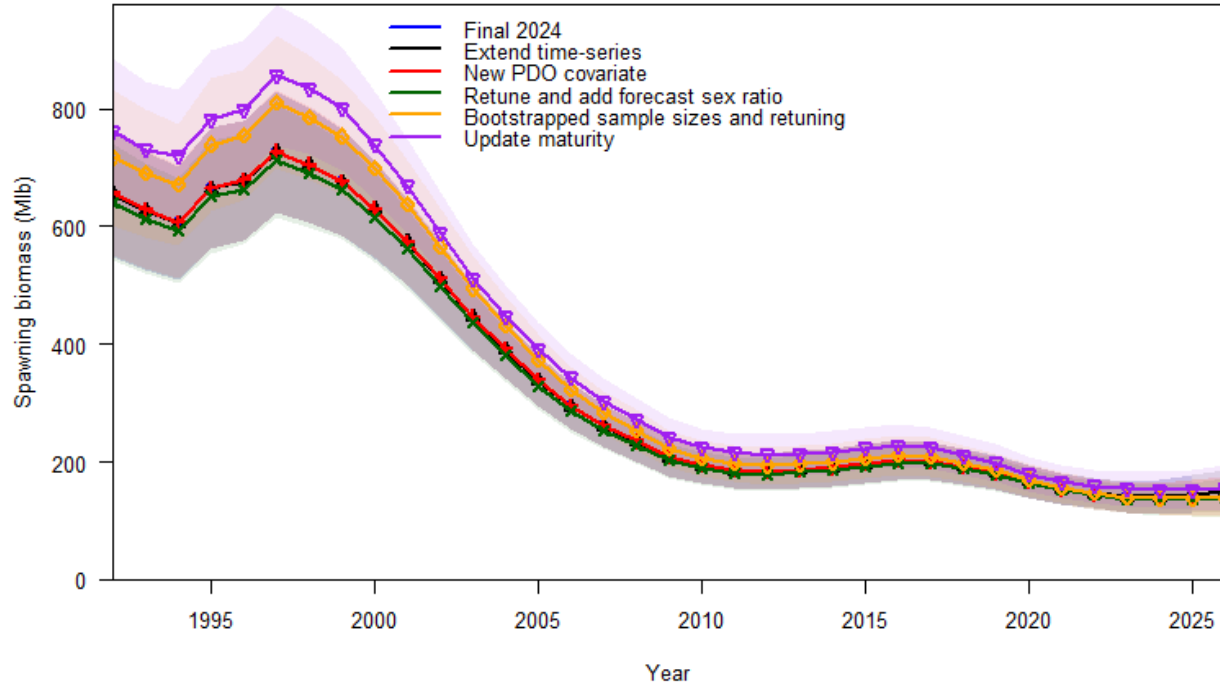
Figure 13. Comparison of recent estimated spawning biomass (1992-2026) over sequential changes from the 2024 to preliminary 2025 AAF long models.

Treatment of male selectivity
The next step in the bridging analysis was to revisit the treatment of time-varying male selectivity scale relative to female selectivity. In the offset approach used to define male selectivity, the scale (asymptote) of male selectivity is parameterized as a random walk beginning in 2018. This approach was first implemented in the 2019 stock assessment, and extended in the 2022 stock assessment as the sex-specific age composition data grew from two years to four years. However, the sigma constraining the random walk was still poorly informed in 2022 and the initial value used in 2019 (0.02, for a parameter that can logically vary from 0.0 to 1.0) had not been iteratively tuned. Further, the time-varying deviations in fishery selectivity were extended into the forecast period as part of the 2024 stock assessment (Stewart and Hicks 2025), in order to propagate the variance associated with unknown future selectivity but the male scale parameters were not included in this extension. For the preliminary 2025 stock assessment the deviations were extended to the full forecast period for consistency with all other modelled deviations (recruitment, fishery catchability, selectivity) and the sigmas iteratively retuned to be consistent with the estimated variability. The results of this change were negligible for all four models (Figure 10-15).

Updated sample sizes and data weighting
The next step in the bridging analysis was to replace the previously-used bootstrapped sample sizes with the updated bootstrapped maximum effective sample sizes (including ageing imprecision) described above. The effective sample sizes and process deviations were then iteratively retuned (as described above) to regain model internal consistency. Due to the substantial downweighting of the historical age composition information relative to the more recent data (as described above) this bridging step had a relatively large effect on model estimates of spawning biomass over the historical period for all four models (Figure 10-15). For the two long time-series models there was little change over the most recent spawning biomass estimates (Figure 12-13); however, the two short time-series models both estimated a larger recent spawning biomass for this bridging step, apparently largely as a result of increased estimates of the 2012 and 2016 year-classes (Figure 14-15).

Updated maturity ogive
The final step in the bridging analysis was to replace the historical maturity ogive with the newly estimated relationship described above and in *IPHC-2025-SRB026-06*. As is expected, shifting the maturity ogive toward younger fish results in a larger estimate of spawning biomass across all four models and across the entire time-series (Figure 10-15). Despite the upward scaling of this new information there was little effect on the trends. This is because the updated maturity is not treated as a time-varying process (a single estimate is applied to the entire time-series) and the only feedback to modelled dynamics is through the stock-recruitment function which estimated quite variable recruitment deviations.
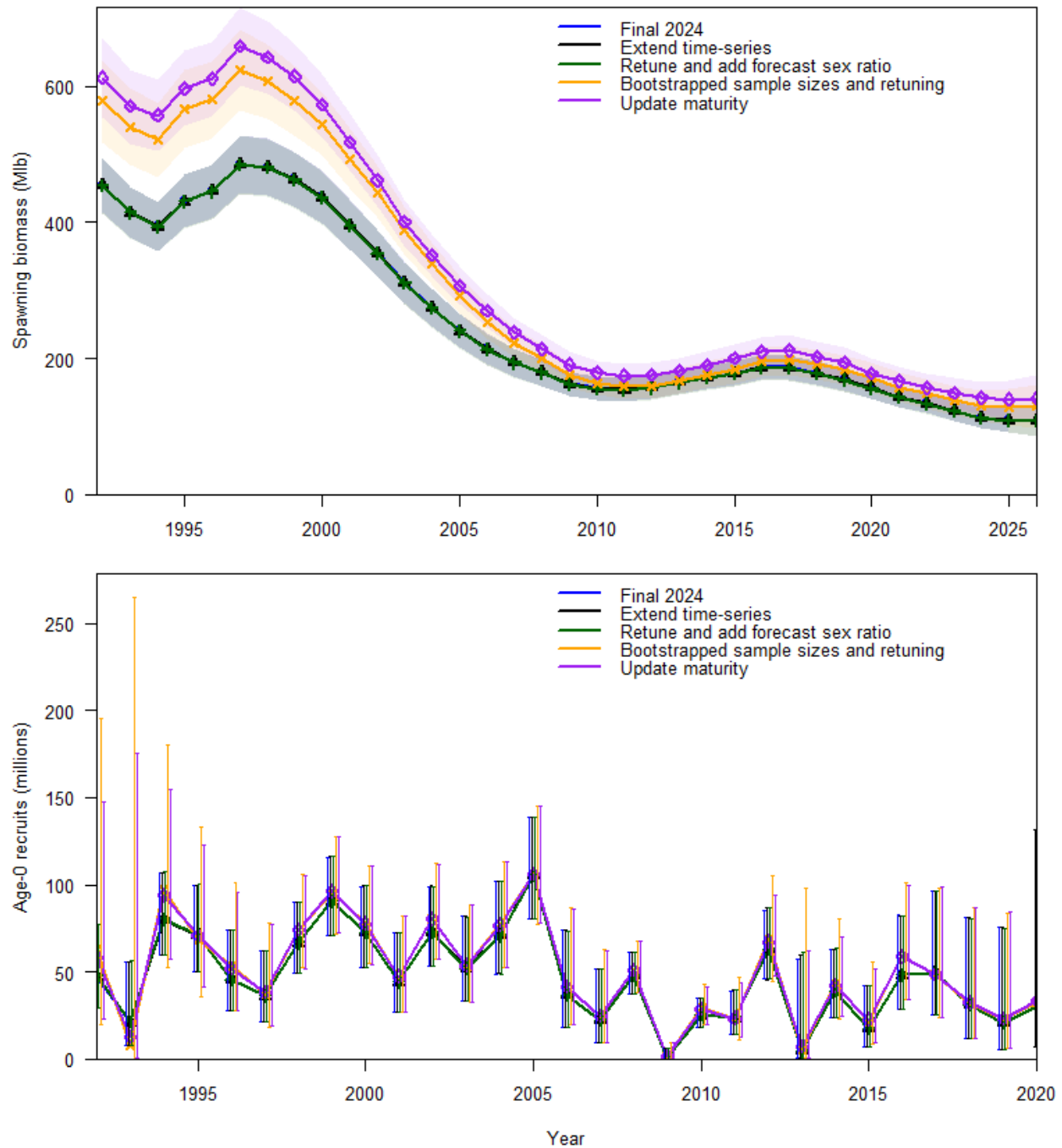
Figure 14. Comparison of estimated spawning biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2024 to preliminary 2025 coastwide short models.
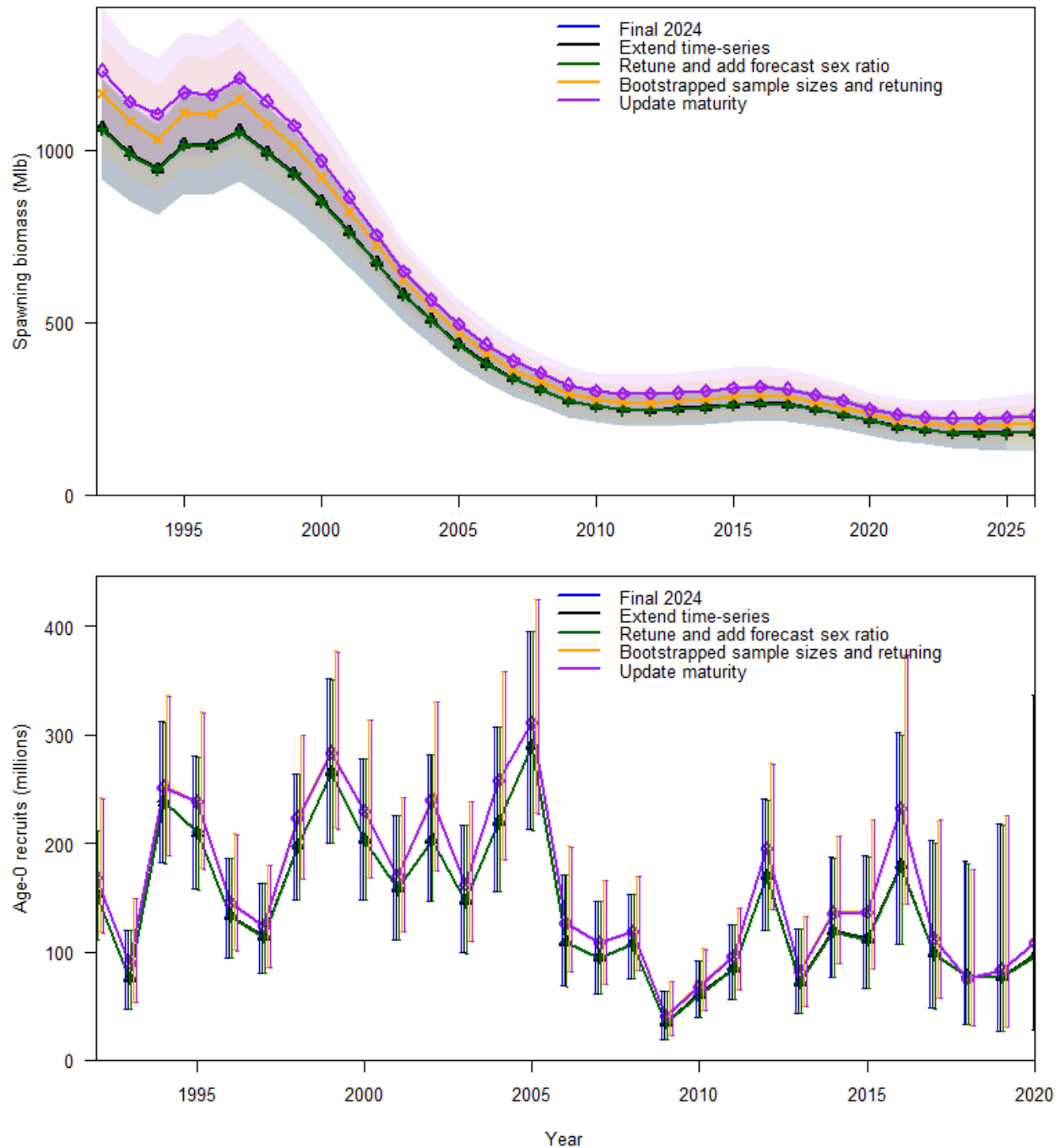
Figure 15. Comparison of estimated spawning biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2024 to preliminary 2025 AAF short models.

Convergence criteria
Standard tools for monitoring convergence criteria include assessing the maximum gradient component, sensitivity to alternative phasing and initial values, use of overdispersed starting points or 'jitter analyses', as well as likelihood profiles, and Bayesian integration.

Wherever parameters were hitting bounds either the bounds were adjusted (if biologically plausible) or the parameters were fixed. For example, the descending limb of the 4B commercial fishery in the AAF models was estimated to be at the bound of 1.0 (as has been the case for all recent assessments) and so was fixed at this value. This approach reduces the likelihood that variances calculations will be (undesirably) effected by parameters stuck to bounds but does require periodic revisitation to ensure that the signal for parameters hitting bounds remains, and that fixing those parameters does not have an appreciable effect on the maximum likelihood solution.

For this preliminary 2025 assessment, all individual models all had a maximum gradient component < 0.002. A series of preliminary and intermediate runs did not indicate any signs that the estimates reported here represented local minima, nor did the models have difficulty producing a positive definite Hessian matrix under the range of alternative and sensitivity analyses (some presented in this document, but many used only for development). Both the AAF models did have trouble resolving the historical deviation parameters under some starting and phasing (the order in which the parameters were added to the minimization) configurations. In the stock synthesis framework random walk deviations cannot be estimated prior to the base selectivity parameters (and must start from a value of 0.0), therefore it is difficult to establish a general pattern of estimating the scaling parameters first and then adding less influential parameters later in the estimation phases. For this reason, the setup for these two models utilizes a parameter file, starting estimation for subsequent models at or near the solution from a previous run. Whenever a parameter file is used, it is important to periodically (and especially for the final model) rerun the model from dispersed starting points.

Convergence was explored for all four models specifically through a 'jitter' analysis perturbing all parameter values simultaneously and repeating minimization. A strong test using this method provides over-dispersed starting values such that the model is traversing a broad range of parameter space to ensure that the Maximum Likelihood Estimate (MLE) does not represent a local minimum that might be a poorer solution than another point in the likelihood space. For each of the four models 100 sets of dispersed starting points were used to initiate minimization for this analysis. Convergence to the MLE occurred for 46/100 for the CW long model, 26/100 for the CW short model, 65/100 for the AAF short model, and 50/100 for the AAF long model. None of the solutions resulted in a minimum that was better than the MLE. Although true convergence to a global minimum can never be proven, all convergence criteria indicate that the results of the preliminary 2025 assessment provide a robust solution.

Individual model diagnostics and results
This section provides more detail on the specific diagnostics and results of each of the four assessment models. It is not intended to provide the fit and residuals to every data component,

but to summarize the basic performance of the model and specifically highlight areas of potential deficiency. Figures showing comprehensive diagnostics and results and the full report files, as output directly from stock synthesis, are provided electronically as described in Appendix A. Each model section finishes with a brief summary of the relative strengths and weaknesses of that model.

Coastwide short

Predictions of both the fishery and survey indices of abundance fit the observed data very well in the coastwide short model (Figure 16). Prior to the 2019 assessment, a small amount of process error was allowed on fishery catchability. In the 2019 analysis this process error was effectively zero and was turned off for that and the 2021-2024 stock assessments. Re-evaluation as part of this analysis via iterative tuning of the annual catchability deviations indicated a small value was again consistent with model fit and was therefore included. The predicted aggregate age distributions also matched the observed distributions well, for both the fishery and survey indicating that the selectivity parameterization was generally capturing differences in both the age-structure and the sex-ratio (Figure 17).

The coastwide short model tuning resulted in a higher weight on the coastwide FISS ages than for the commercial fishery age data (Table 18). The discard, non-directed discard and recreational age data were all intentionally heavily down-weighted (as described above) and so input sample sizes were not iterated to larger values, despite fits to the data that implied a higher weight. Fit to the annual FISS age compositions were generally good (Figure 18), although some patterning was visible in the standardized residuals (Figure 19). Specifically, there was a clear pattern of negative residuals in the plus group for male halibut; however, this was almost imperceptible in the fits themselves due to the very small observed and predicted values in this age bin. The fits to the annual fishery data were also acceptable, noting some patterning associated with the 1987 cohort and ages 15 and 20 in the most recent decade (Figure 20-21). The implied fit to the sex ratio information for the commercial fishery (Figure 22) was similar to that for the FISS (Figure 23); both show year-to-year variability in the scale and patterns. Additional diagnostics and diagnostic figures (such as fits to the down-weighted annual compositions for the discard, bycatch, and recreational fleets) are included in the background materials.

Neither the FISS nor the fishery female selectivity was estimated to have a highly variable ascending limb over the short time-series (Figure 24). The estimated fishery selectivity showed a small increase in the selection of males at the end of the time-series, somewhat the opposite of that estimated for the FISS (Figure 25), perhaps a function of the catch distribution shifting toward the Eastern side of the stock where fast-growing males are much more common. For the discard fleet, estimated selectivity included fewer and younger females than males (Figure 26). Estimated selectivity for the non-directed discards fleet showed a peak at ages 4-5 and a slightly domed relationship. Recreational/subsistence selectivity was shifted to the left of the commercial fishery discards (and therefore the FISS).
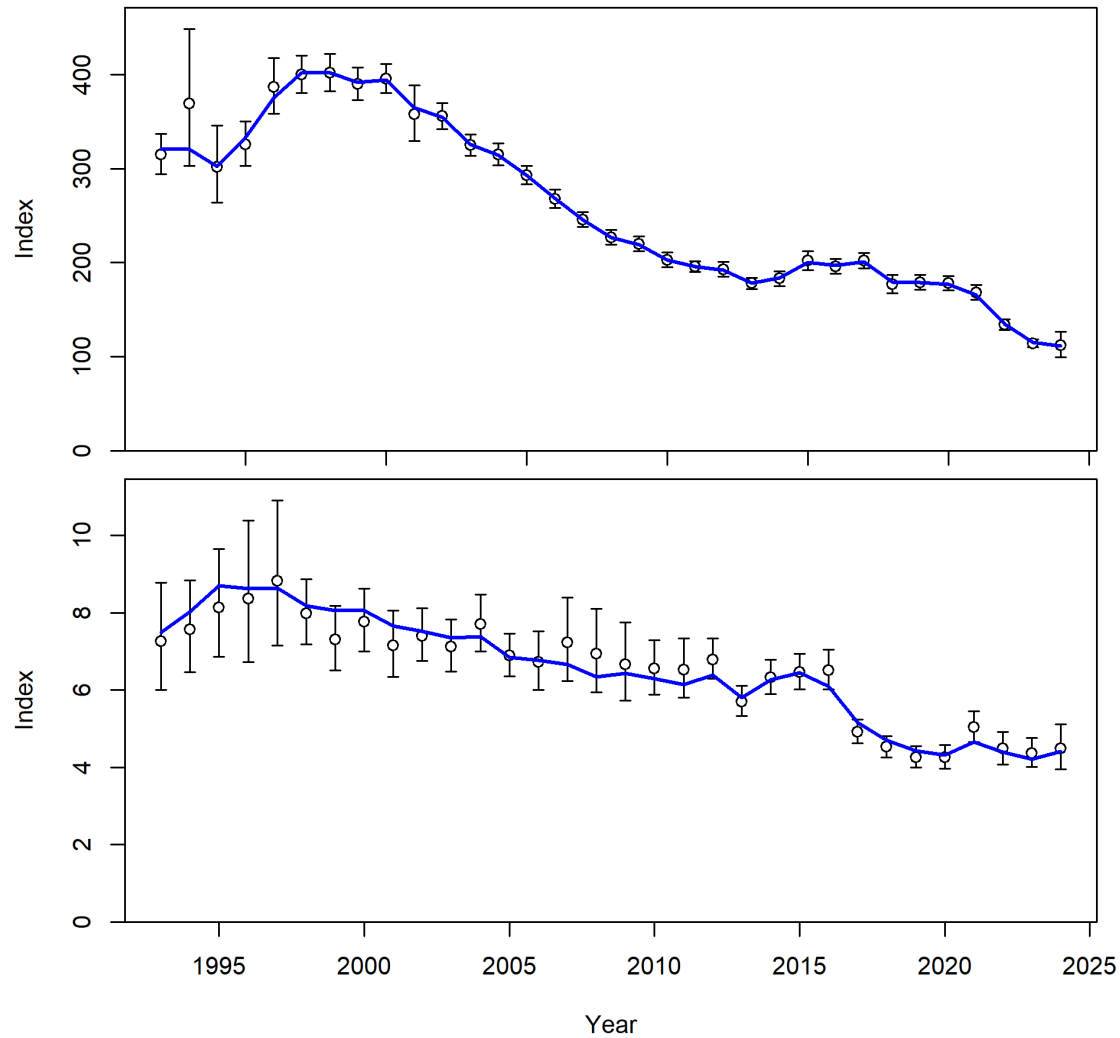
Figure 16. Fit to fishery (upper panel) and FISS (lower panel) indices of abundance in the coastwide short model; note that the scale of the y-axes differ as do the units (the fishery index is in weight and the FISS in numbers).
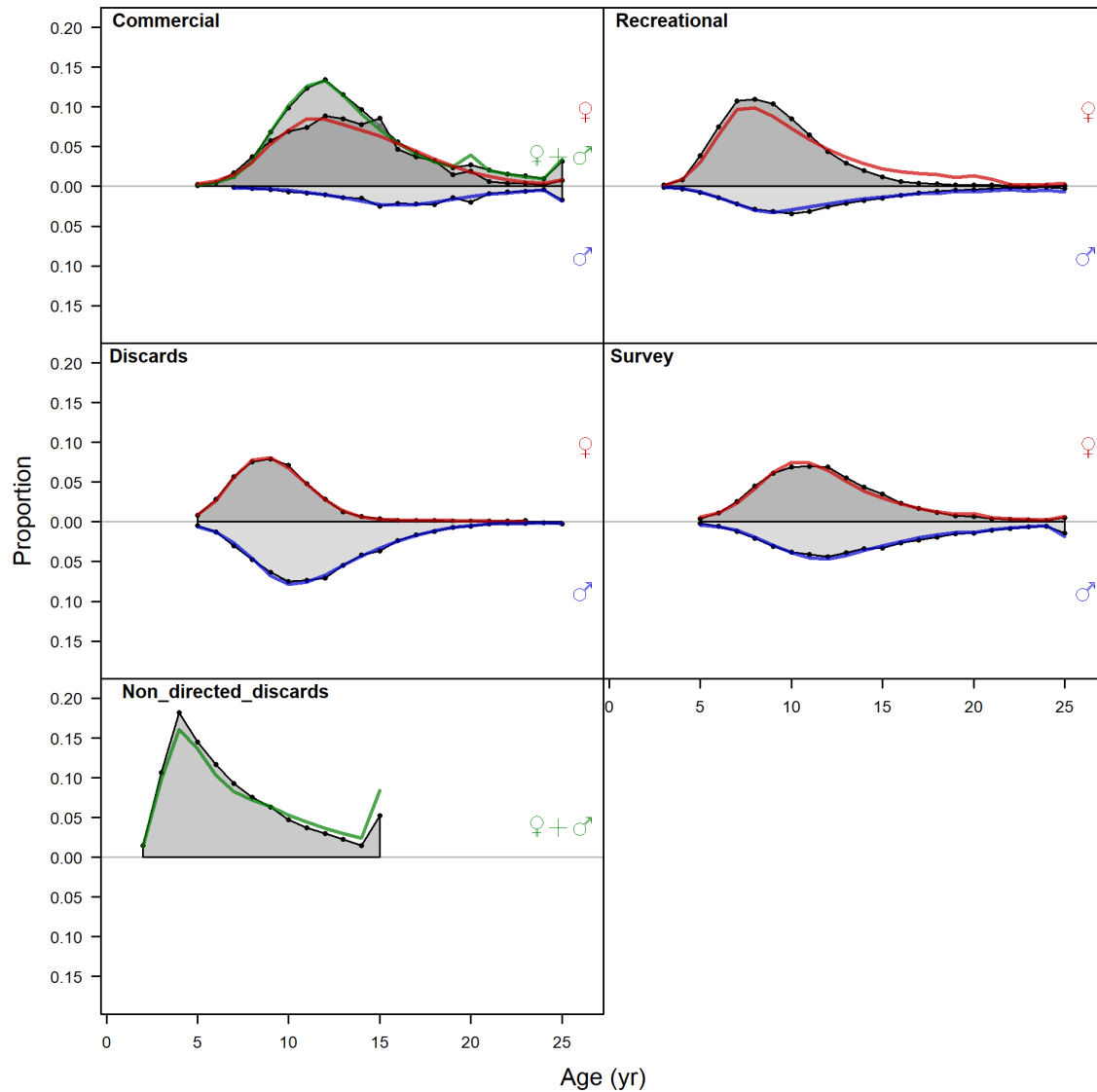
Figure 17. Aggregate fit to all age data by model fleet in the coastwide short model; sex-specific distributions for the commercial fishery represent only 2017-2023 and are plotted on top of sexes-aggregated distributions spanning 1992-2016 + 2024.

Table 18. Post-iteration sample size diagnostics for age-composition data by model and fleet. Average iterated input denotes the value used for model runs reported here, after iterating the bootstrapped starting points.

	Average iterated input	Harmonic mean effective	Francis weight effective	Maximum Pearson residual
Coastwide short				
Fishery	159	423	174	2.08
Discards[1]	6	221	116	0.67
Non-directed discards[1]	3	51	56	1.61
Recreational[1]	3	104	24	0.63
FISS	164	810	163	2.78
Coastwide long				
Fishery	144	318	148	3.01
Discards[1]	6	213	100	0.65
Non-directed discards[1]	3	38	7	1.30
Recreational[1]	3	131	20	0.61
FISS	97	208	97	4.18
AAF short				
Region 2 fishery	456	600	825	4.91
Region 3 fishery	599	609	733	4.10
Region 4 fishery	58	85	61	2.08
Region 4B fishery[2]	49	130	65	2.43
Discards[1]	6	198	80	0.66
Non-directed discards[1]	3	49	28	0.84
Recreational[1]	3	128	21	0.56
Region 2 FISS	7	77	5	1.06
Region 3 FISS	27	317	28	1.33
Region 4 FISS	86	156	90	2.68
Region 4B FISS[2]	41	147	38	2.56
AAF long				
Region 2 fishery	256	293	563	4.87
Region 3 fishery	319	286	468	3.93
Region 4 fishery	47	65	49	2.50
Region 4B fishery[2]	49	122	58	2.34
Discards[1]	6	157	85	1.22
Non-directed discards[1]	3	39	8	1.30
Recreational[1]	3	103	20	0.65
Region 2 FISS	5	63	5	1.35
Region 3 FISS	23	145	21	1.44
Region 4 FISS	114	158	121	2.80
Region 4B FISS[2]	34	147	35	2.00

[1]Inputs intentionally down-weighted – see text.

[2]Iterated sample size equal to maximum (bootstrapped input).
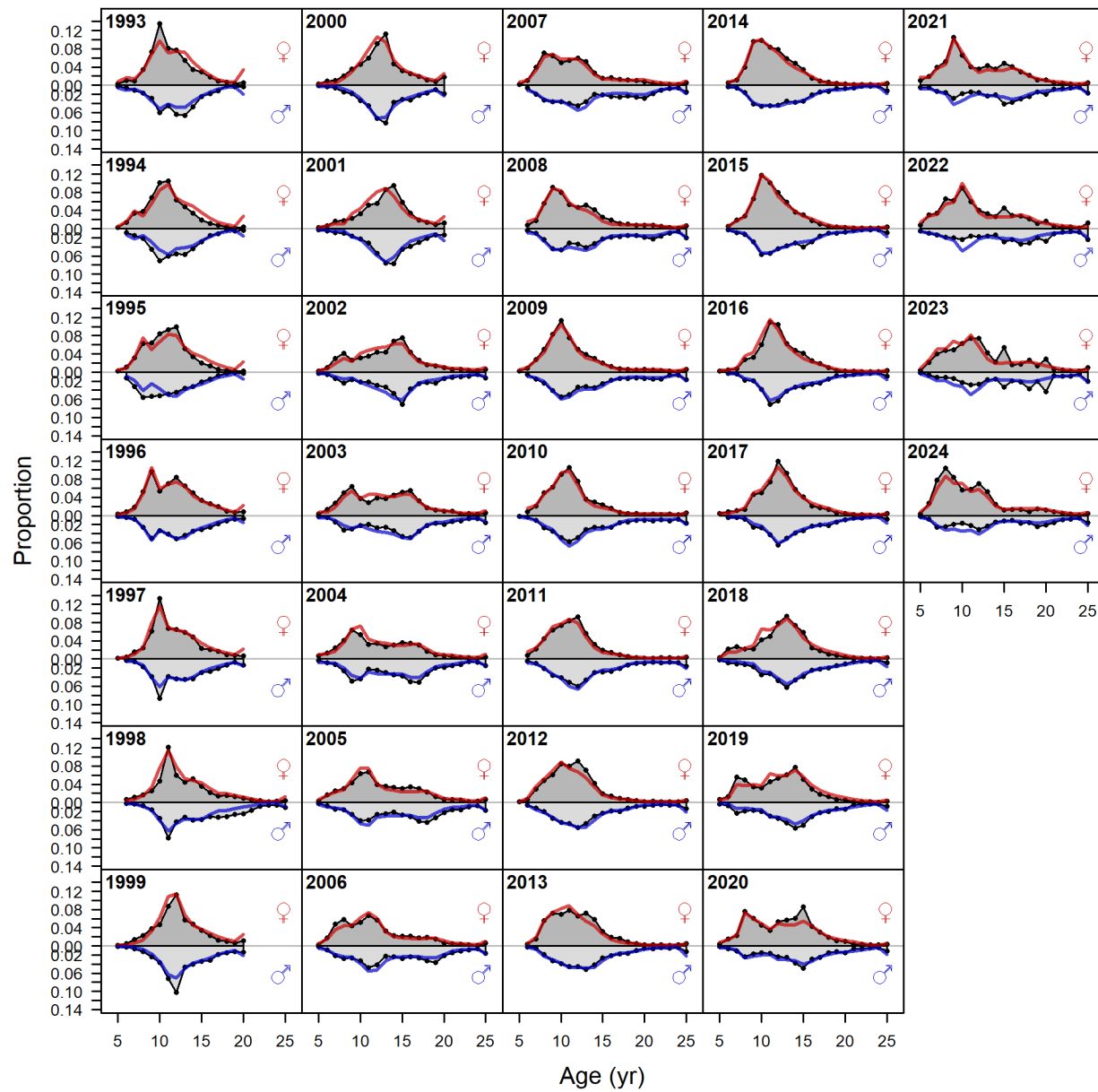
Figure 18. Fit to annual age data from the FISS survey in the coastwide short model.
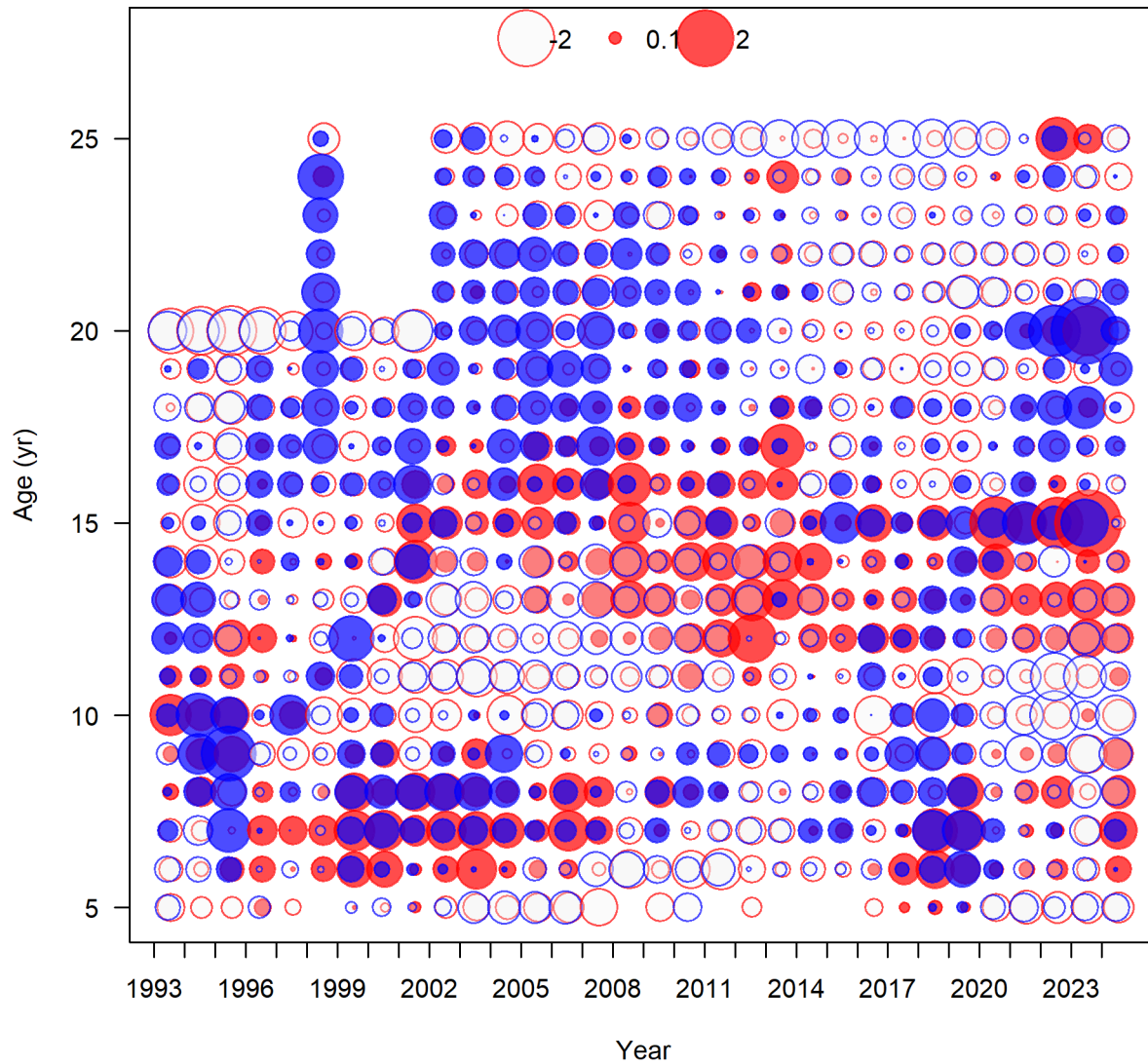
Figure 19. Pearson residuals for fit to annual age data from the FISS survey in the coastwide short model; red circles denote female residuals, and blue circles denote male residuals.
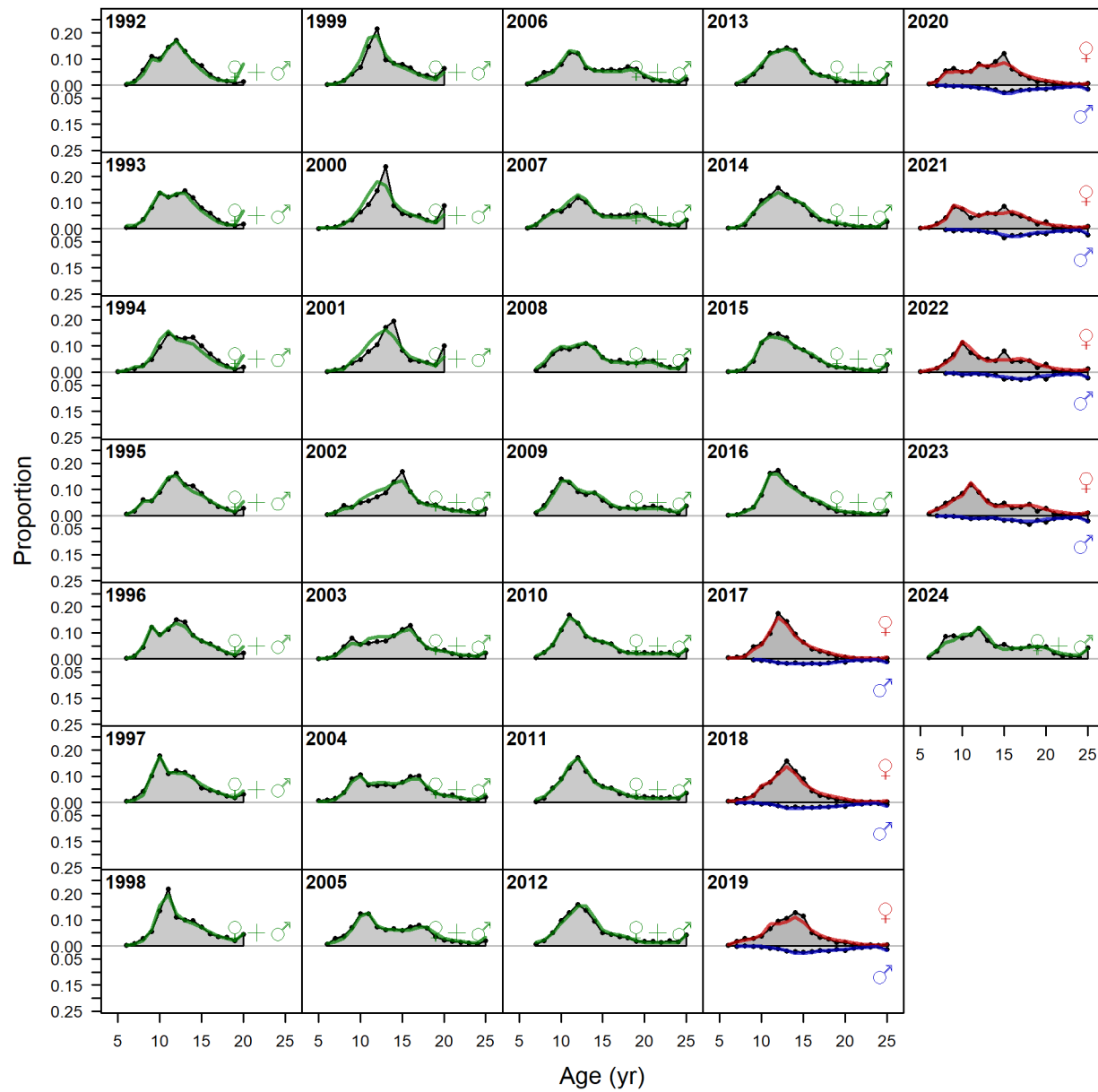
Figure 20. Fit to annual age data from the commercial fishery landings in the coastwide short model.
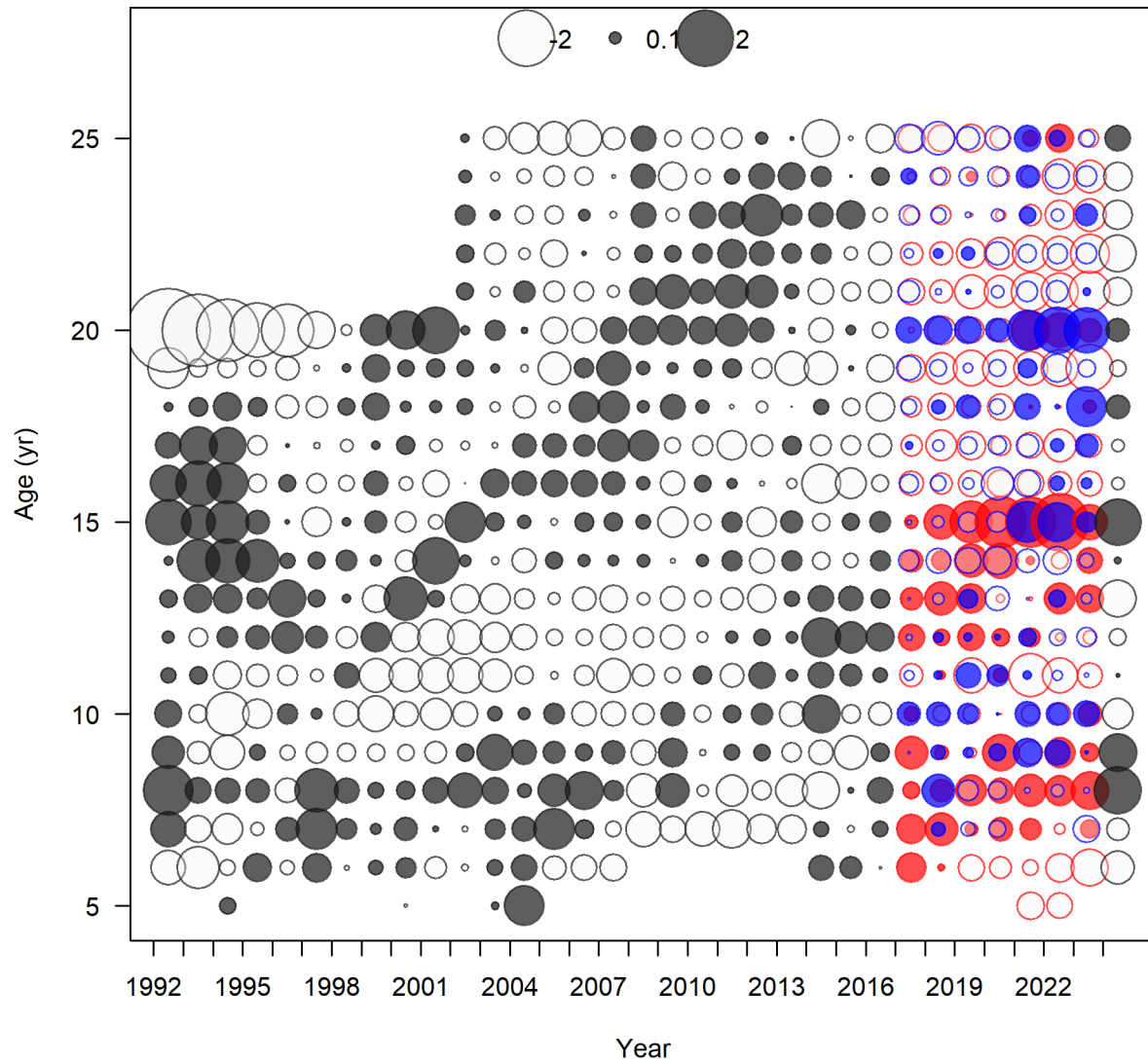
Figure 21. Pearson residuals for the fit to annual age data from the commercial fishery landings in the coastwide short model; grey circles denote unsexed residuals, red circles denote female residuals, and blue circles denote male residuals.
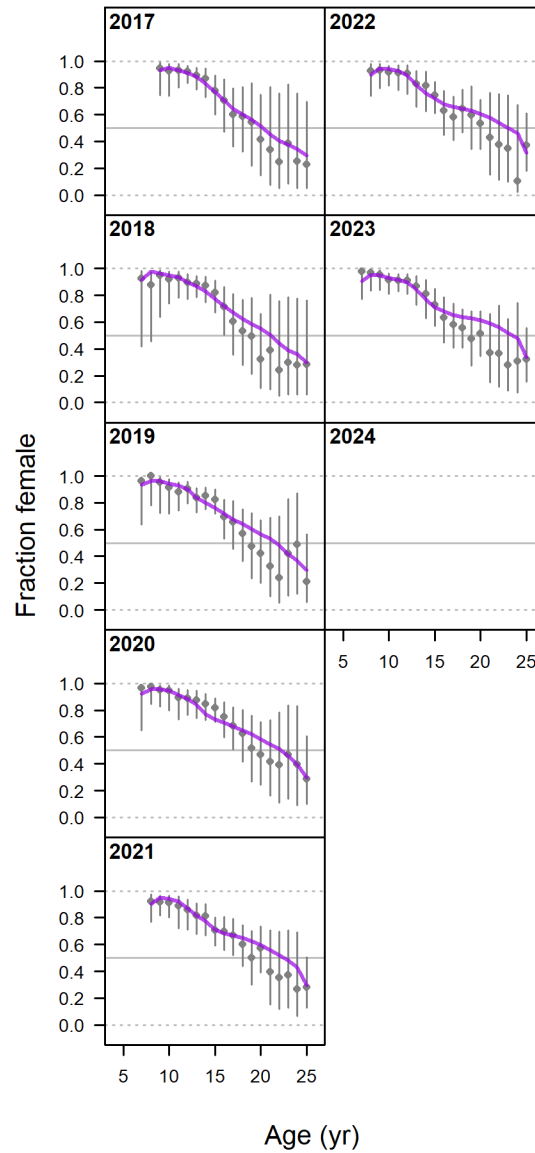
Figure 22. Observed and predicted sex-ratio in the commercial fishery landings from the coastwide short model for years with sex-specific age composition data (2017-2023).
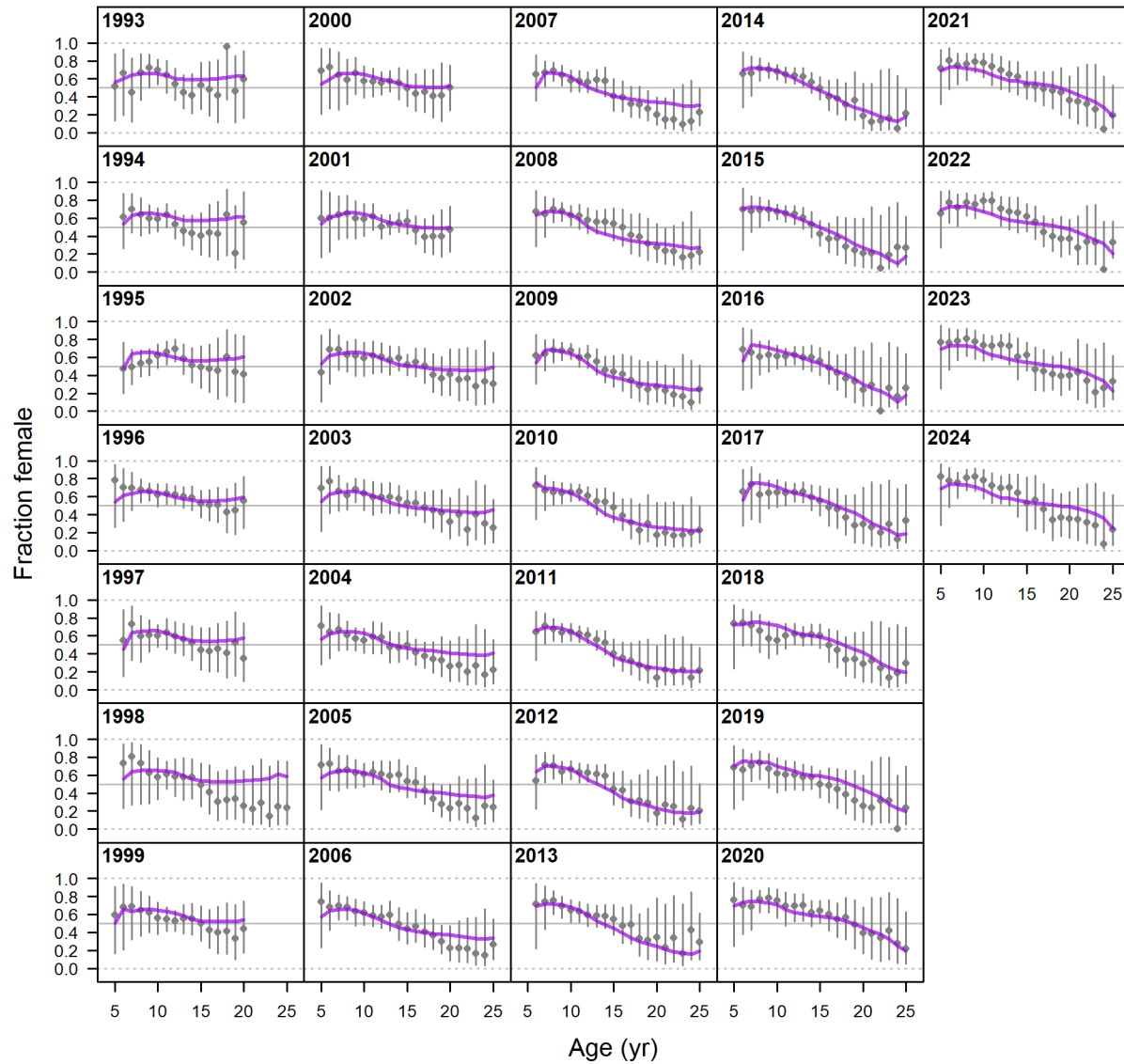
Figure 23. Observed and predicted sex-ratio in the FISS from the coastwide short model.
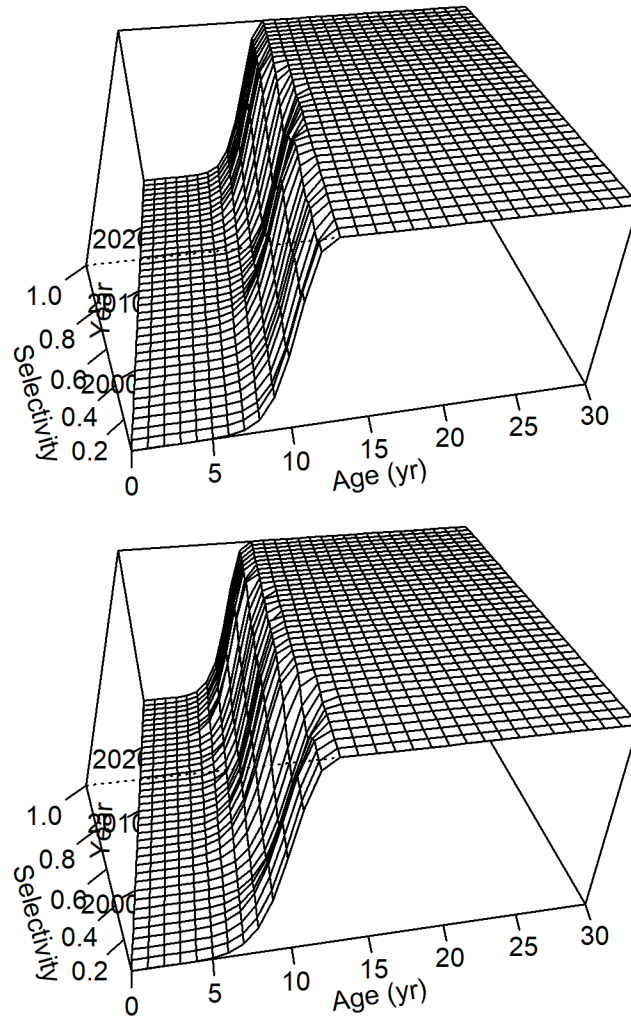
Figure 24. Estimated time-varying female selectivity curves for the commercial fishery landings (upper panel) and the FISS (lower panel).
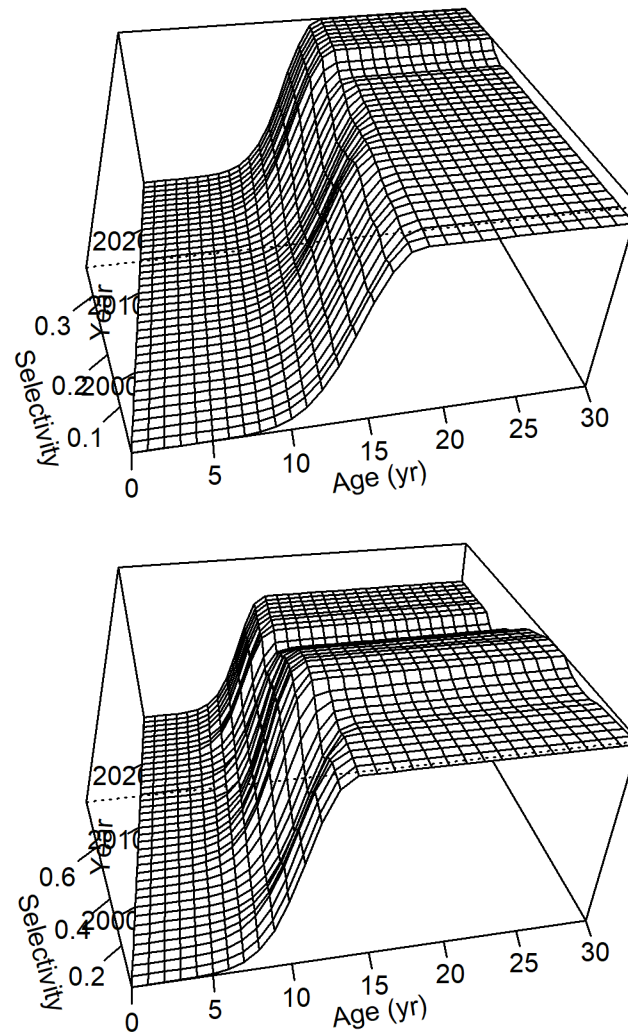
Figure 25. Estimated time-varying male selectivity curves for the commercial fishery landings (upper panel) and the FISS (lower panel).
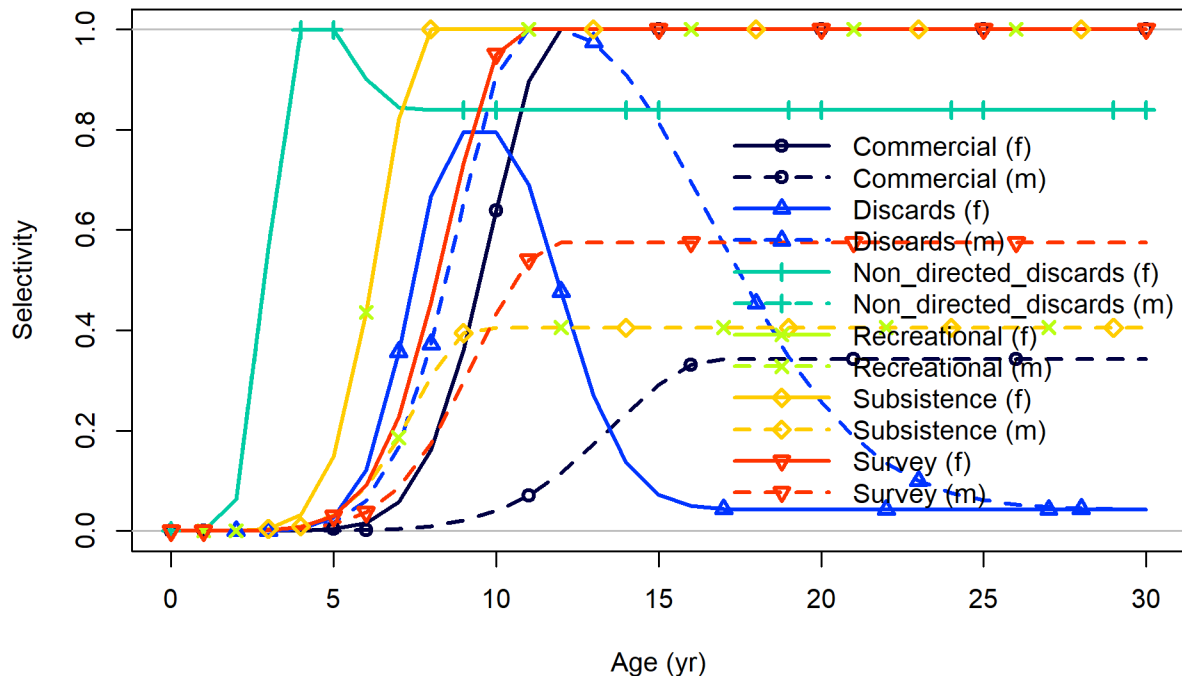
Figure 26. Estimated ending year selectivity curves by sex for the commercial fishery, discard, non-directed discard, recreational and FISS fleets in the coastwide short model.

Male *M* was estimated to be slightly higher (0.164) than the fixed value assumed for females of 0.15 (Table 19); this represented a slight increase from the value estimated in the 2022 and earlier assessments. The large negative estimated initial recruitment offset is consistent with the start year occurring after a very long time-series of fishing. The lower *M* fixed in the coastwide short model corresponded to lower recruitment and female spawning biomass estimates (Table 19) than the other three models, as has been the case for all recent assessments.

Summary of strengths and weaknesses for the coastwide short model:

Strengths:

- Lowest technical overhead (complexity) of the four models in the ensemble
- Fit the fishery and FISS indices very well
- Fit the survey age data (males and females) relatively well
- Parameter estimates are derived from the most recent time-period
- Internally consistent data weighting
- Similar weighting of commercial fishery and FISS age composition data

Weaknesses:

- Basis for fixed female *M* is unclear
- Does not include uncertainty in female *M* (see sensitivity analyses below)
- Does not include extensive historical data

- May lose Region-specific trends and biological patterns due to aggregation
- Does not use environmental information to inform recruitment

Table 19. Select parameter estimates (maximum likelihood value and approximate 95% confidence interval) and important recent population estimates by model and Biological Region (where applicable).

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Biological				
Female *M*	0.150 (*Fixed*)	0.221 (0.185-0.257)	0.220 (0.204-0.236)	0.186 (0.169-0.204)
Male *M*	0.164 (0.155-0.172)	0.198 (0.181-0.216)	0.179 (0.169-0.189)	0.163 (0.154-0.171)
Log(R_0)	11.43 (11.19-11.67)	11.91 (11.51-12.32)	12.30 (12.06-12.54)	11.56 (11.32-11.79)
Initial log(R_0) offset	-1.512 (-1.746--1.278)	NA	-0.193 (-0.411-0.019)	NA
Environmental Link (β)	NA	0.456 (0.238-0.675)	NA	0.430 (0.225-0.636)
Survey Log(q) Δ1984 (transition to circle hooks)	NA	0.933 (0.485-1.381)	NA	R2: 1.344 (0.756-1.513) R3: 1.876 (1.631-2.120)
Fishery Log(q) Δ1984	NA	0.823 (0.647-0.999)	NA	R2: 0.562 (0.373-751) R3: 0.942 (0.751-1.133) R4: 0.850 (0.645-1.055) R4B: 0.381 (0.187-0.575)
2012 Age-0 recruitment (Millions)	67 (48-94)	164 (96-282)	195 (139-273)	115 (86-153)
2025 SB (Million lb)	139 (111-167)	156 (105-208)	226 (165-287)	153 (119-187)

Coastwide long

Both the fishery and FISS indices of abundance were fit well (Figure 27), with breaks in catchability to accommodate the change from "J" to circle hooks (1984) which were very large in both series (Table 19). In aggregate, the predicted age compositions matched the observed data well (Figure 28); however, there were notable differences among years within the time-series. Fits to the FISS were quite poor in the early portion of the time series when the spatial coverage was very limited (Figure 29), but improved where the data became more spatially comprehensive in the mid-1990s, and quite good in the most recent years (Figure 30). Fishery data fit reasonably well for the entire time-series (Figure 31-32), with patterns in the residuals corresponding to relatively small differences with observed distributions. The small contribution

of males to the fishery landed catch is quite clear from the seven years that have sex-specific information (Figure 32). Harmonic mean effective sample sizes were much larger than adjusted inputs when Francis weights were close to 1.0; commercial fishery data were weighted slightly more heavily than FISS data, largely reflecting the spatial coverage of the early FISS years (Table 18).
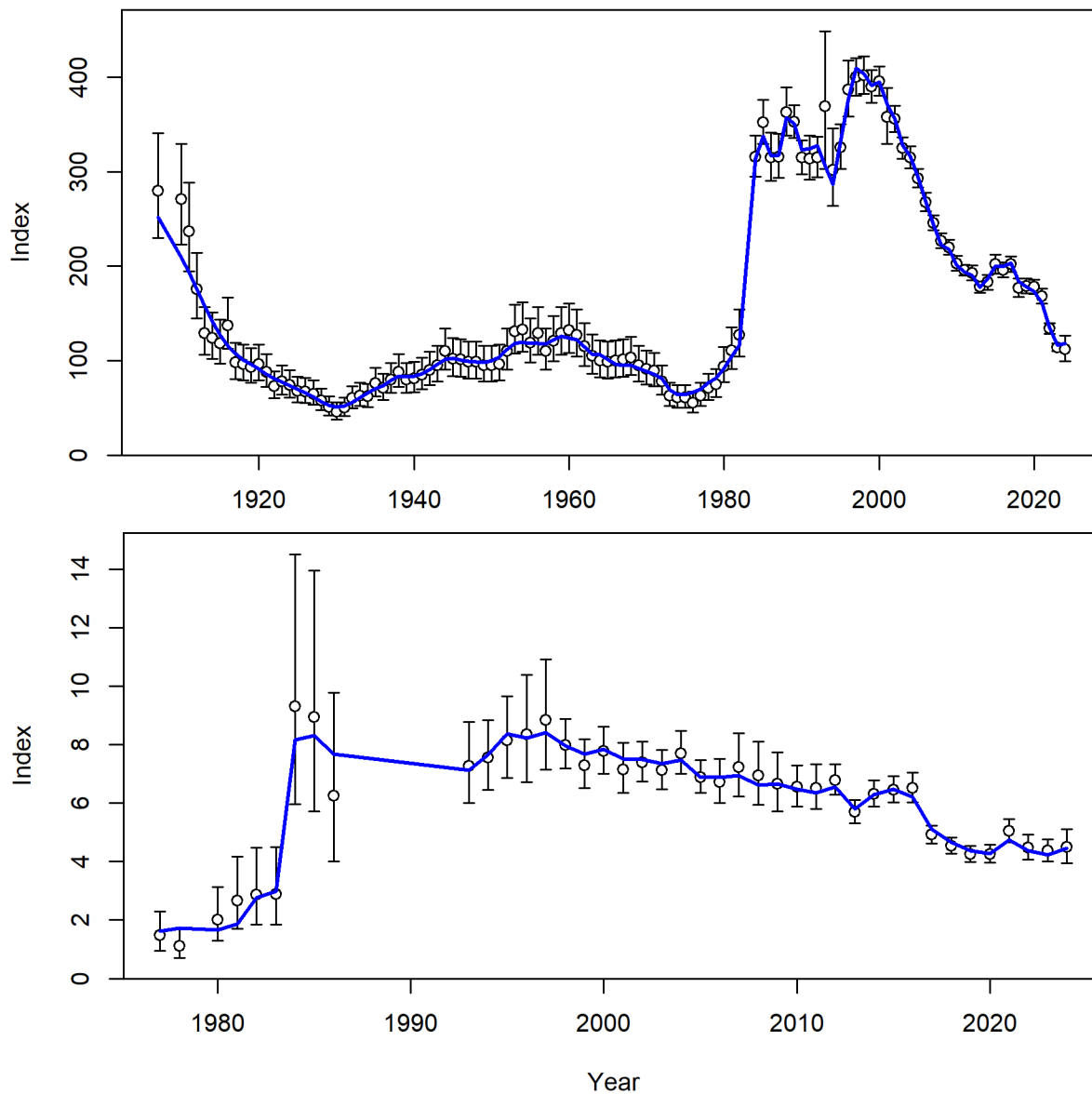
Figure 27. Fit to fishery (upper panel) and FISS (lower panel) indices in the coastwide long model.
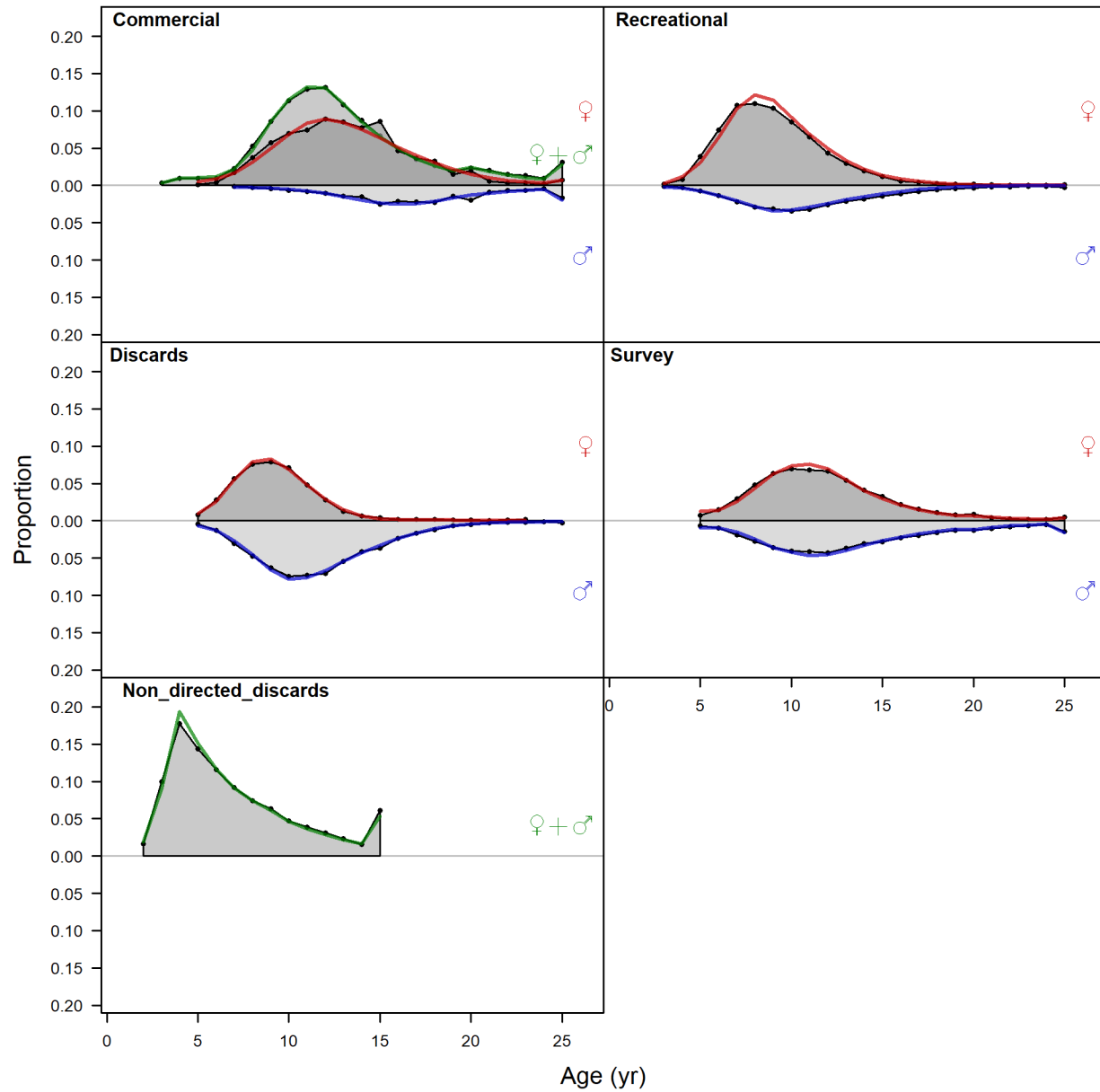
Figure 28. Aggregate fit to all age data by model fleet in the coastwide long model..
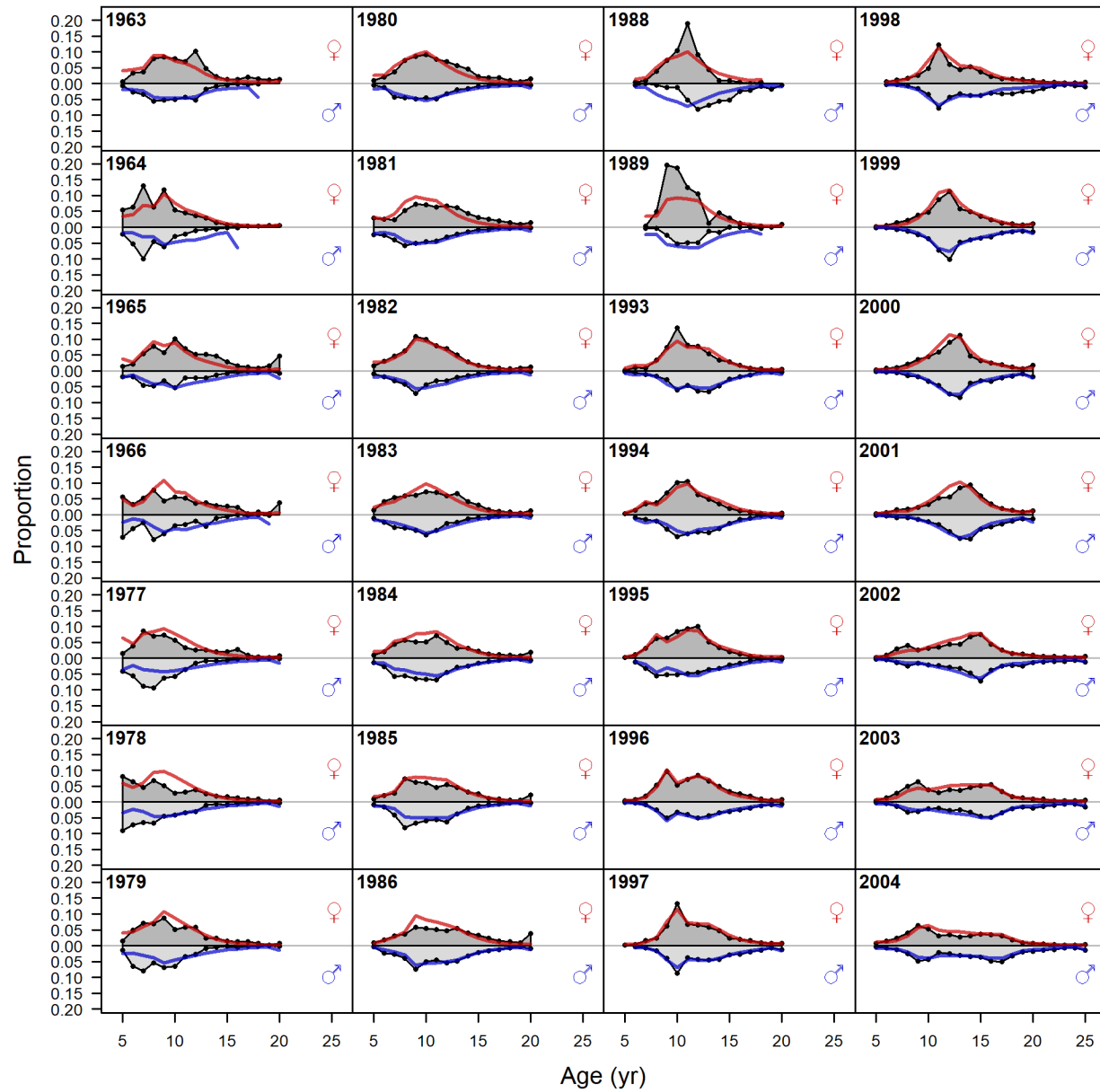
Figure 29. Fit to early years of FISS age data in the coastwide long model.
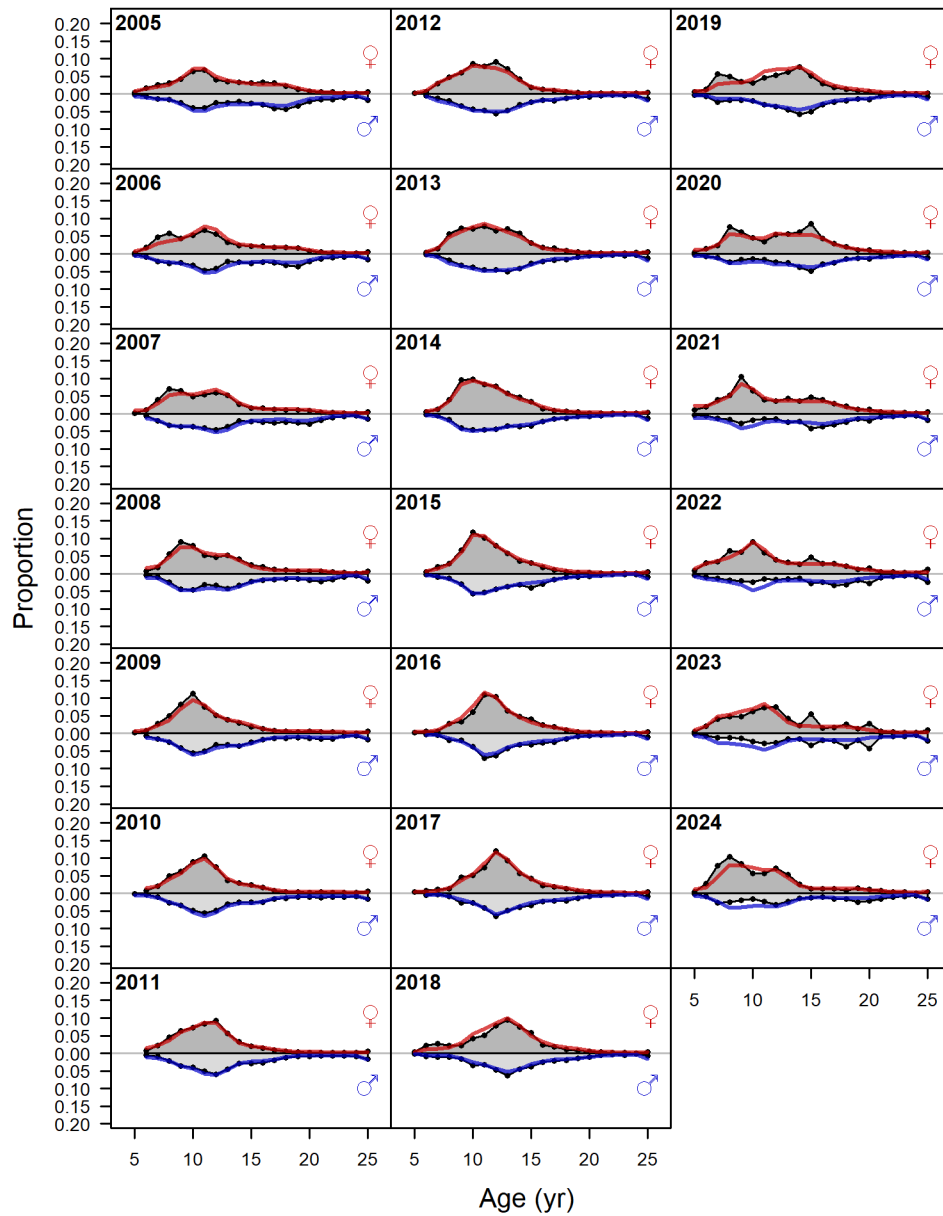
Figure 30. Fit to later years of FISS age data in the coastwide long model.
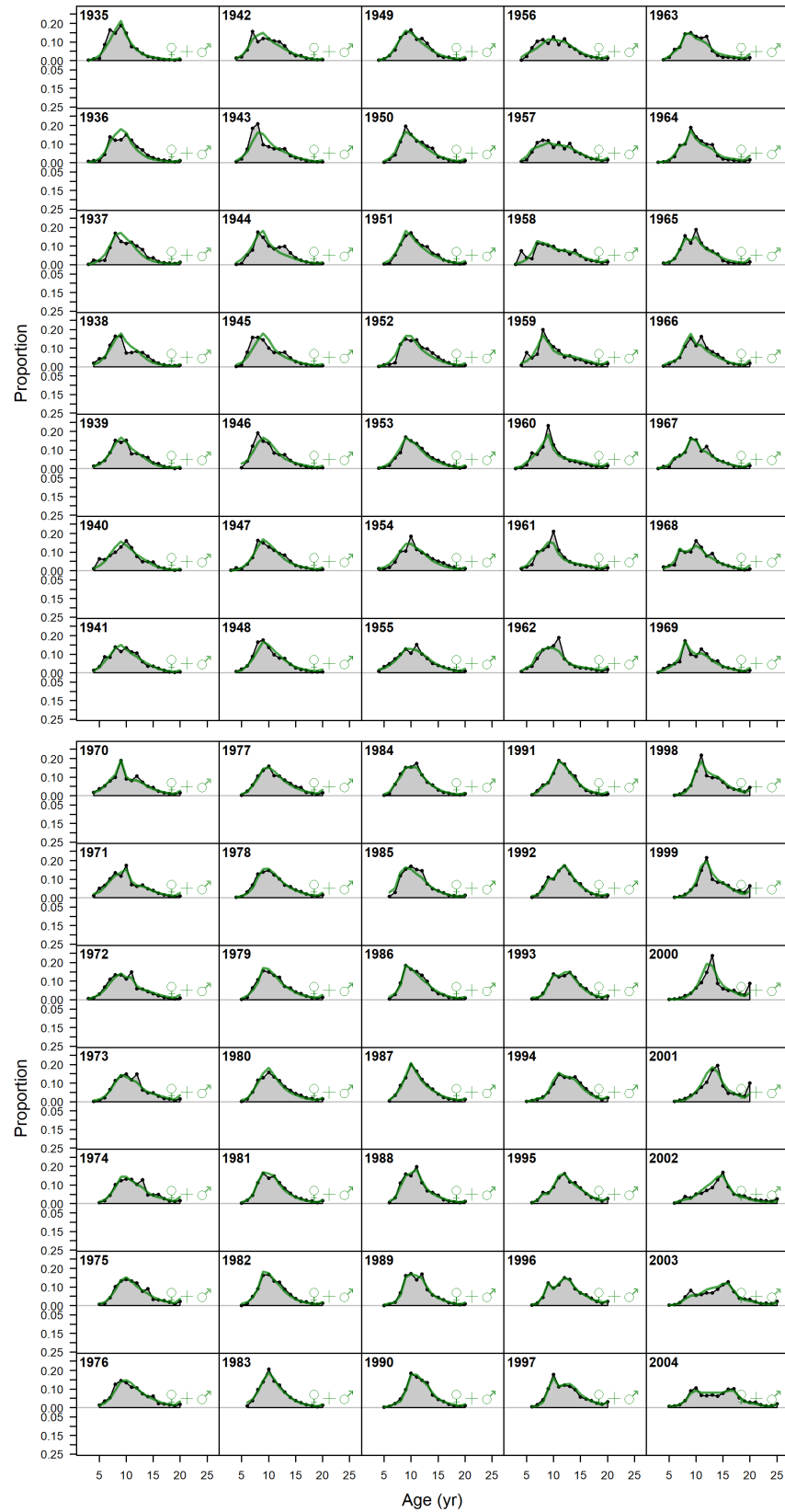
Figure 31. Fit to early years of fishery age data in the coastwide long model.
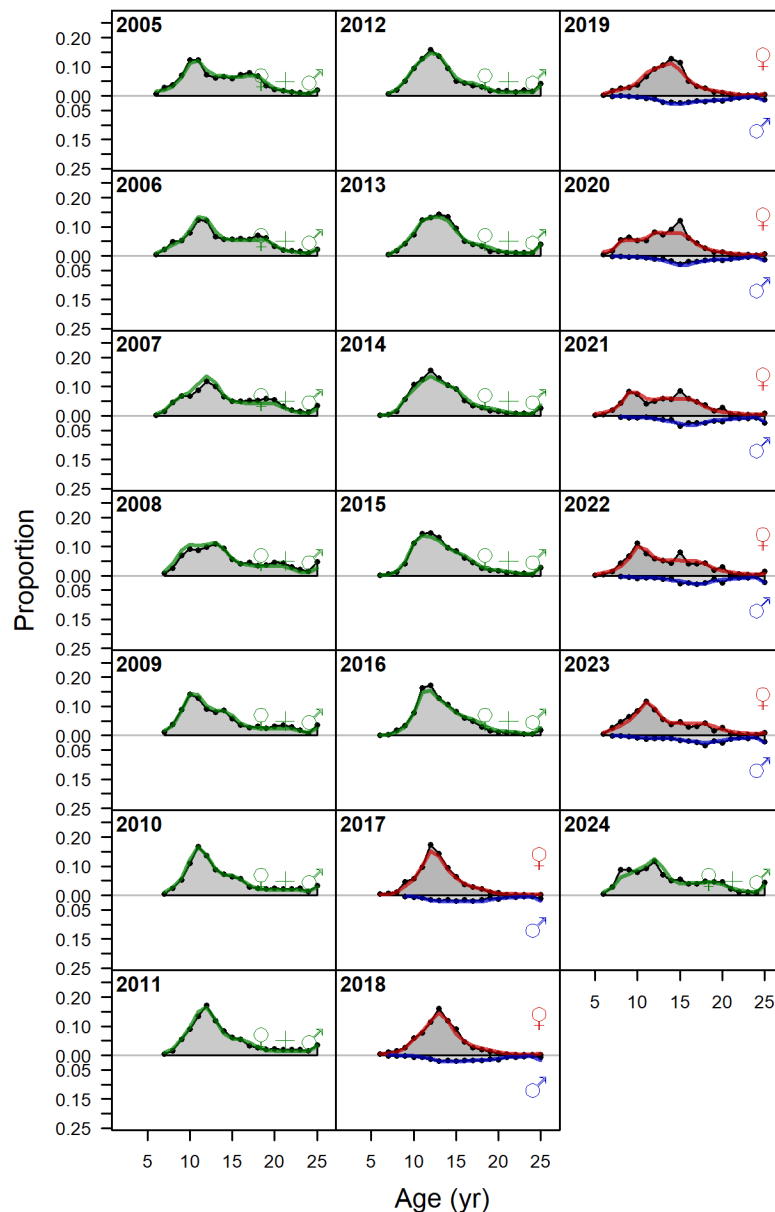
Figure 32. Fit to later years of fishery age data in the coastwide long model.

Fishery selectivity generally showed a pattern toward selecting fewer younger fish in the latter half of the time series, and a similar trend was estimated for the FISS (Figure 33). The apparent deviation toward lower selectivity of males around 2020 for the FISS may reflect the abrupt change in spatial coverage in that year due to logistical challenges and a reduced design. The overall shift toward lower selectivity for younger fish may be consistent with changes in both the age-structure of the stock, the trends in size-at-age interacting with age-based selectivity and the spatial distribution creating changes in availability. Fishery catchability was estimated to have a large (unconstrained) increase associated with the change from "J" to circle hooks (Table 19, Figure 34). Older halibut were more represented in the non-directed fishery discards age data

prior to 1992, and therefore the estimated selectivity was nearly asymptotic. Recreational and discard selectivity estimates were relatively similar to those from the coastwide short model.

Female natural mortality in the coastwide long model was estimated to be higher (0.221) than for males (0.198) although the 95% intervals overlap (Table 19). The environmental link parameter ($β$) was estimated to be positive (0.456), with no density below a value of 0.0, thus suggesting a strong and significant relationship between average recruitment and the phase of the PDO (based on the updated PDO index described above; Table 19). However, the time series of estimated recruitments (Figure 35) and deviates from the PDO-informed stock-recruitment relationship (Figure 36) still show some temporal patterns, suggesting the potential for unmodelled effects on the stock-recruitment relationship might still be present. Specifically, the poor PDO period from the 1940s to the 1970s and the positive phase from the 1970s to the early 2000s generally correspond to negative and positive deviations even with the relationship included (Figure 36).

Summary of strengths and weaknesses for the coastwide long model:

Strengths:

- Includes uncertainty in female natural mortality
- Includes extensive historical data
- Uses environmental information to inform recruitment
- Modest technical overhead (complexity)
- Fits the fishery and survey indices well
- Fits both the survey and fishery age data well
- Internally consistent data weighting

Weaknesses:

- May lose Region-specific trends and biological patterns due to aggregation
- Relies heavily on only fishery trends over the historical period
- Implicitly assumes stationarity in some processes (e.g., the stock-recruitment function after accounting for the PDO, M) over the long historical period
- Implicitly assumes that availability to the fishery did not change over the historical period, despite known patterns in geographical expansion prior to the 1960s
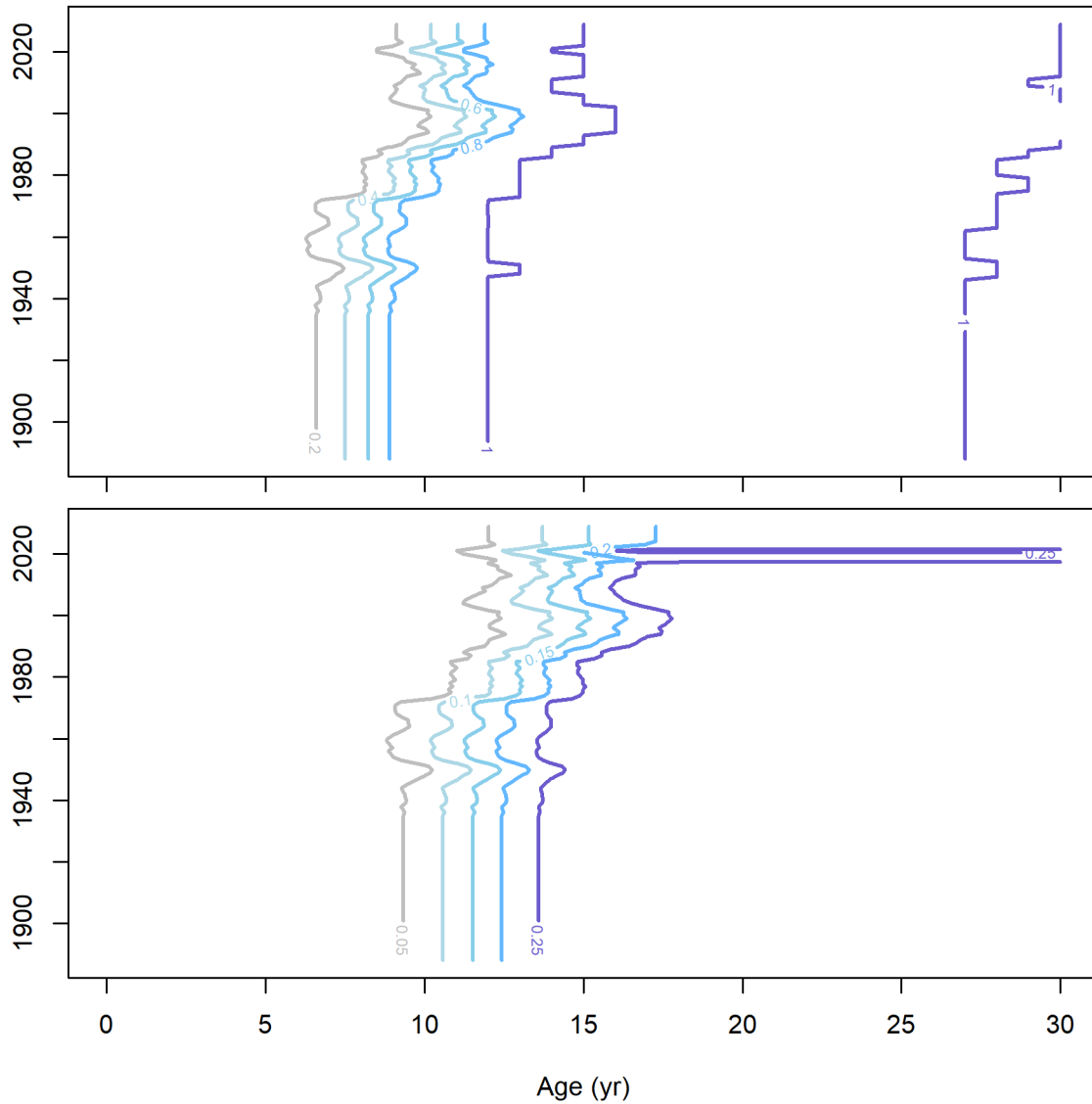
Figure 33. Estimated selectivity for females in the commercial fishery landings (upper panel) and survey (lower panel) in the coastwide long model.
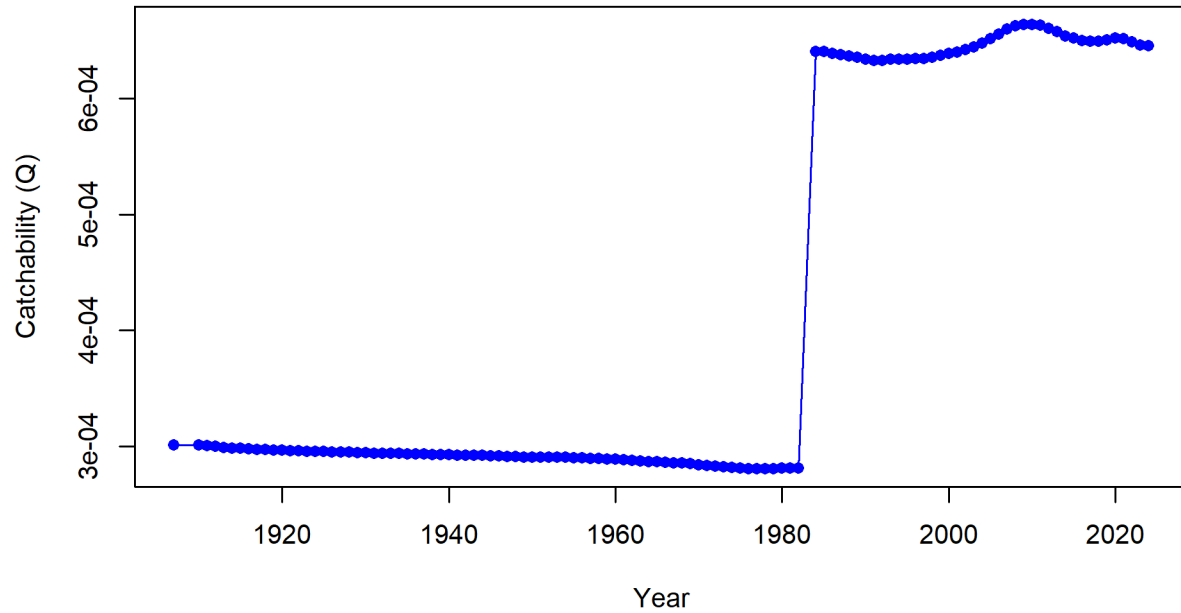
Figure 34. Time-varying fishery catchability in the coastwide long model. The change corresponding to the transition to circle hooks in 1984 is unconstrained.
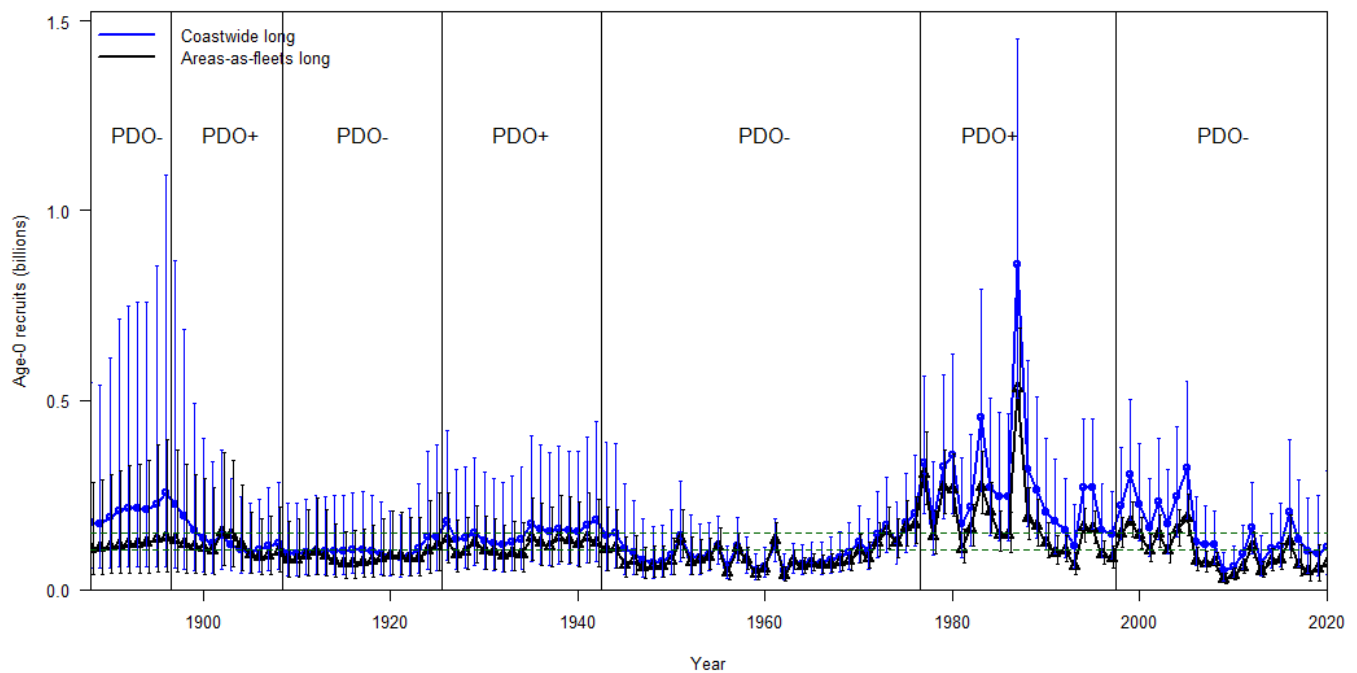
Figure 35. Estimated recruitments and assumed PDO regimes from the coastwide long and AAF long models (right panel); horizontal lines indicate equilibrium values in the absence of the PDO.
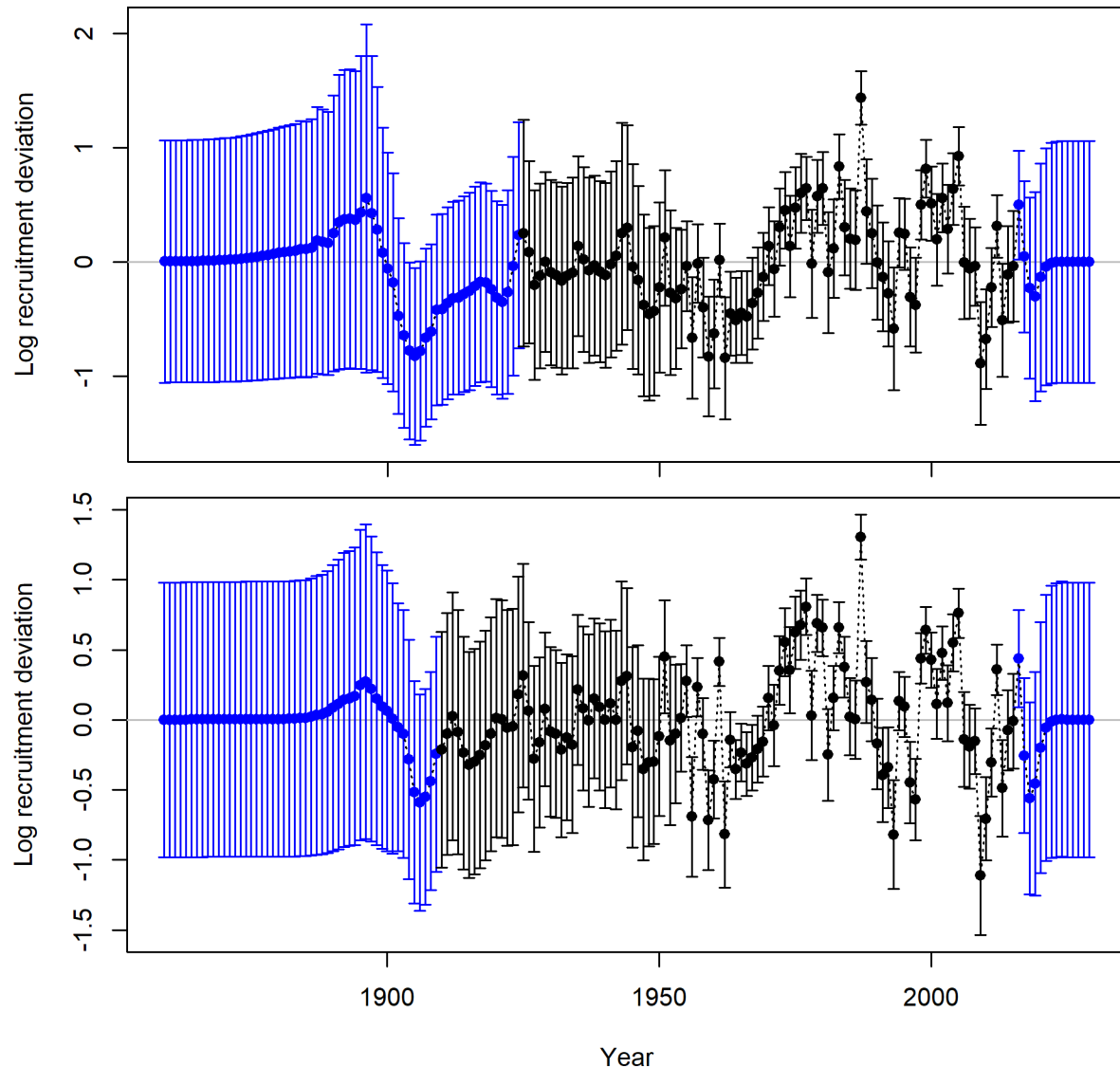
Figure 36. Estimated recruitment deviations in the coastwide long (upper panel) and AAF long (lower panel) models; horizontal lines indicate expected values based on the stock-recruitment functions as modified by the estimated PDO relationships.

AAF short

The AAF short model fit the observed trends in all fishery and FISS indices relatively well (Figure 37-38). Fit to the aggregate age data for each fleet clearly illustrated the differences in age structure among the data from each biological region and among fishery sectors (Figure 39). The biggest differences between the age of female and male halibut observed from the FISS occurred in Region 3, and generally Regions 4 and 4B were predicted (and observed) to have the greatest fraction of older halibut, a majority of which were males. The fit to the annual FISS age data generally captured these patterns, with the worst fit occurring for the data from Region 2 (Figure 40); the model weighting suggested a low effective sample size for the Region 2 FISS data consistent with these patterns in lack of fit (Table 18). Considerable exploration was made toward improving the fit to the Region 2 FISS data and addressing the clear residual patterns (see sensitivity analyses below; Figure 40); however, so satisfactory

replacement approach was identified. The fit to the age composition data from Region 4 clearly shows the very small proportion of males in the landings (Figure 41), with a much greater proportion observed in Region 4B (Figure 42). Although showing a reasonably good aggregate fit, predicted annual commercial fishery landings in Biological Regions 4 and 4B did not capture the strong peaks created by the 1987 year-class in the late 1990s and early 2000s, suggesting that this large year class may have moved toward Regions 2 and 3 as those fish grew older and therefore the fit represents a compromise between fitting the younger and older observations from that cohort. This type of spatial dynamic is not fully approximated by an Areas-As-Fleets approach and would require a fully spatial model to model more accurately. Both of these Regions were weighted similarly after iterative tuning (Table 18).

The estimate of female natural mortality in the AAF short model (0.220) was slightly lower than in the coastwide long model and males were estimated to have a much lower value (0.179; Table 19). The lack of overlap on the 95% intervals indicates the clearly different explanation in this model for the observed sex-ratios, albeit restricted to the most recent portion of the time-series. This result likely indicates the trade-off between the assumption of asymptotic selectivity in the coastwide model and domed selectivity for most Regions in the AAF models. The AAF short model estimated a negative but somewhat smaller initial offset to recruitment as the coastwide short model. Due to the higher estimated M, the AAF short model estimated a higher absolute level of recent recruitment and spawning biomass than the coastwide short model (Table 19).

Summary of strengths and weaknesses for the AAF short model:

Strengths:

- Parameter estimates are derived from the most recent time-period
- Avoids aggregating data over Biological Regions with differing trends and biological patterns
- Fits the Regional fishery and FISS indices well
- Fits Regions 2 and 3 fishery age data well
- Internally consistent data weighting
- Propagates uncertainty in female and male M estimates

Weaknesses:

- Does not include environmental information to inform recruitment
- Increased technical overhead (complexity)
- Residual patterns in Region 4 and 4B fishery and survey age data
- Fits Region 2 FISS age data poorly
- Does not include extensive historical data
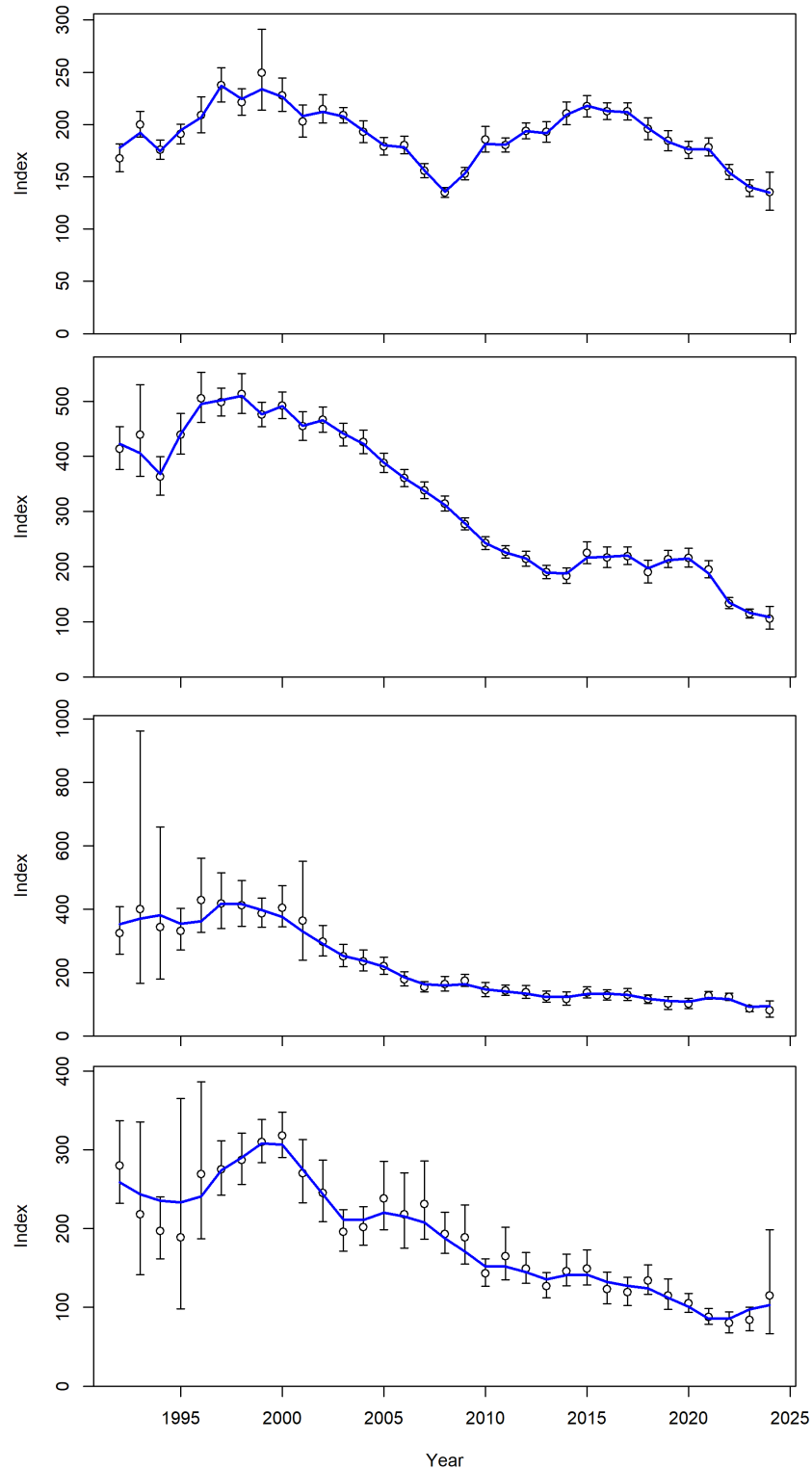
Figure 37. Fit to fishery trends in Biological Regions 2, 3, 4, and 4B (top to bottom) in the AAF short model.
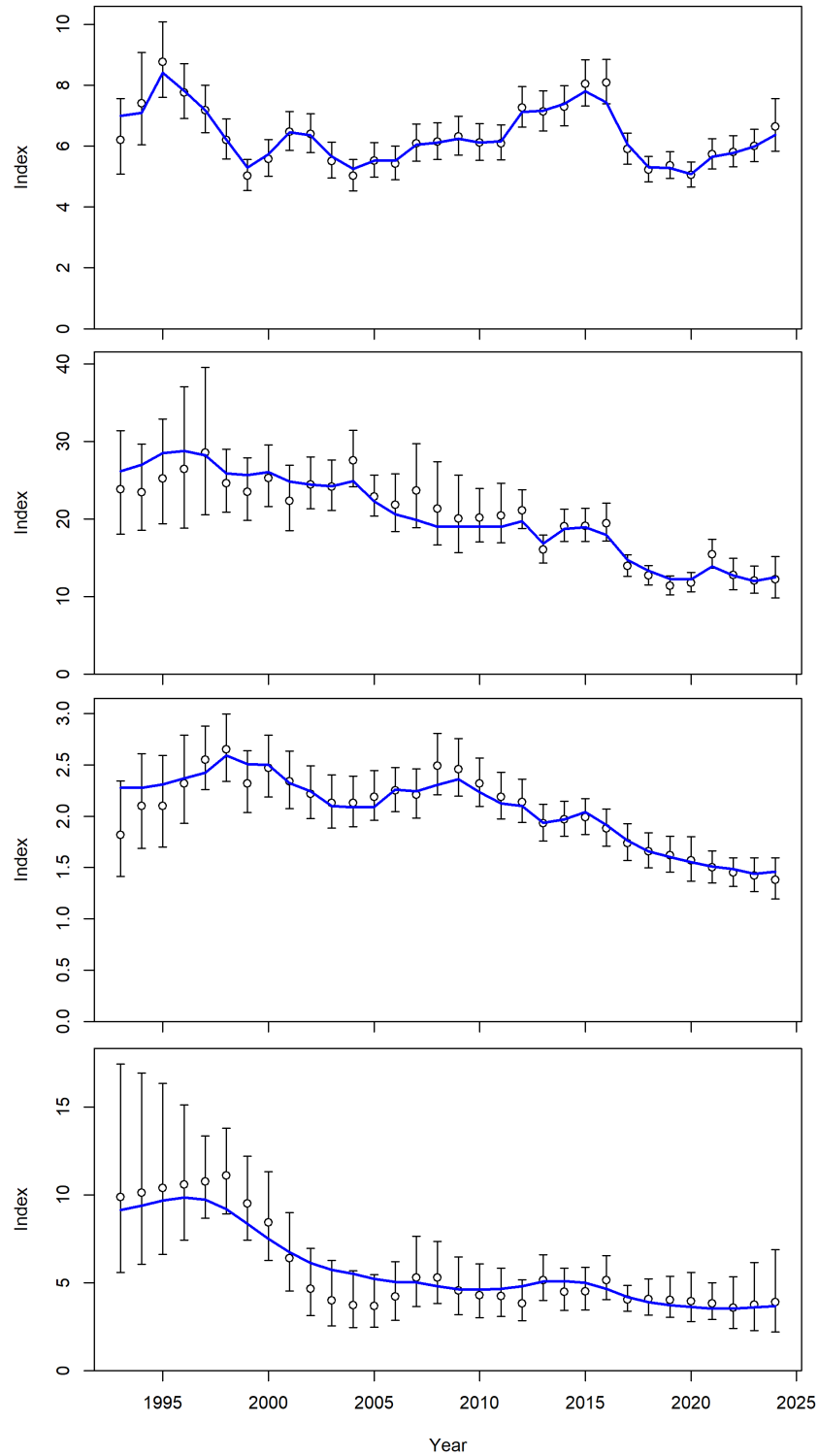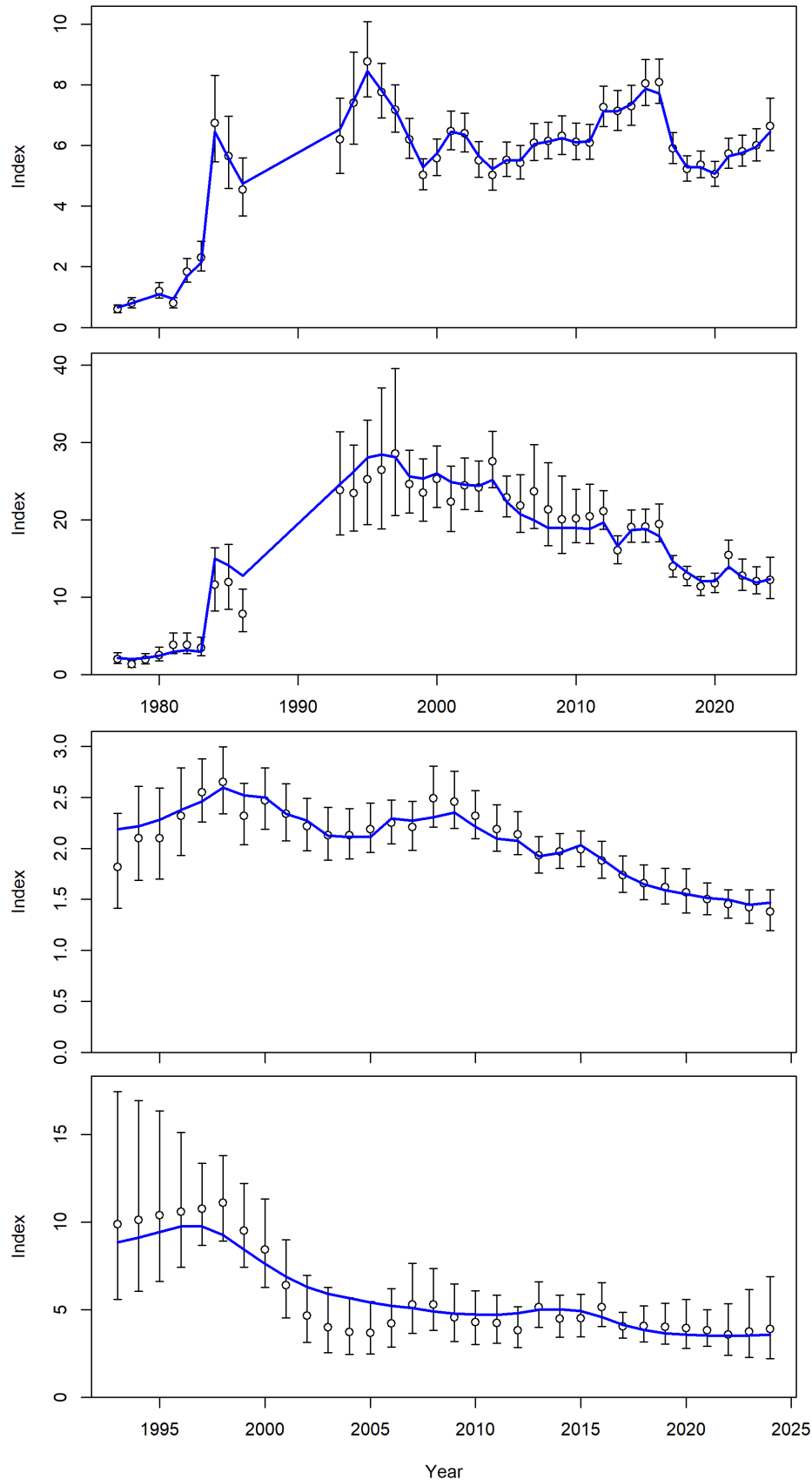
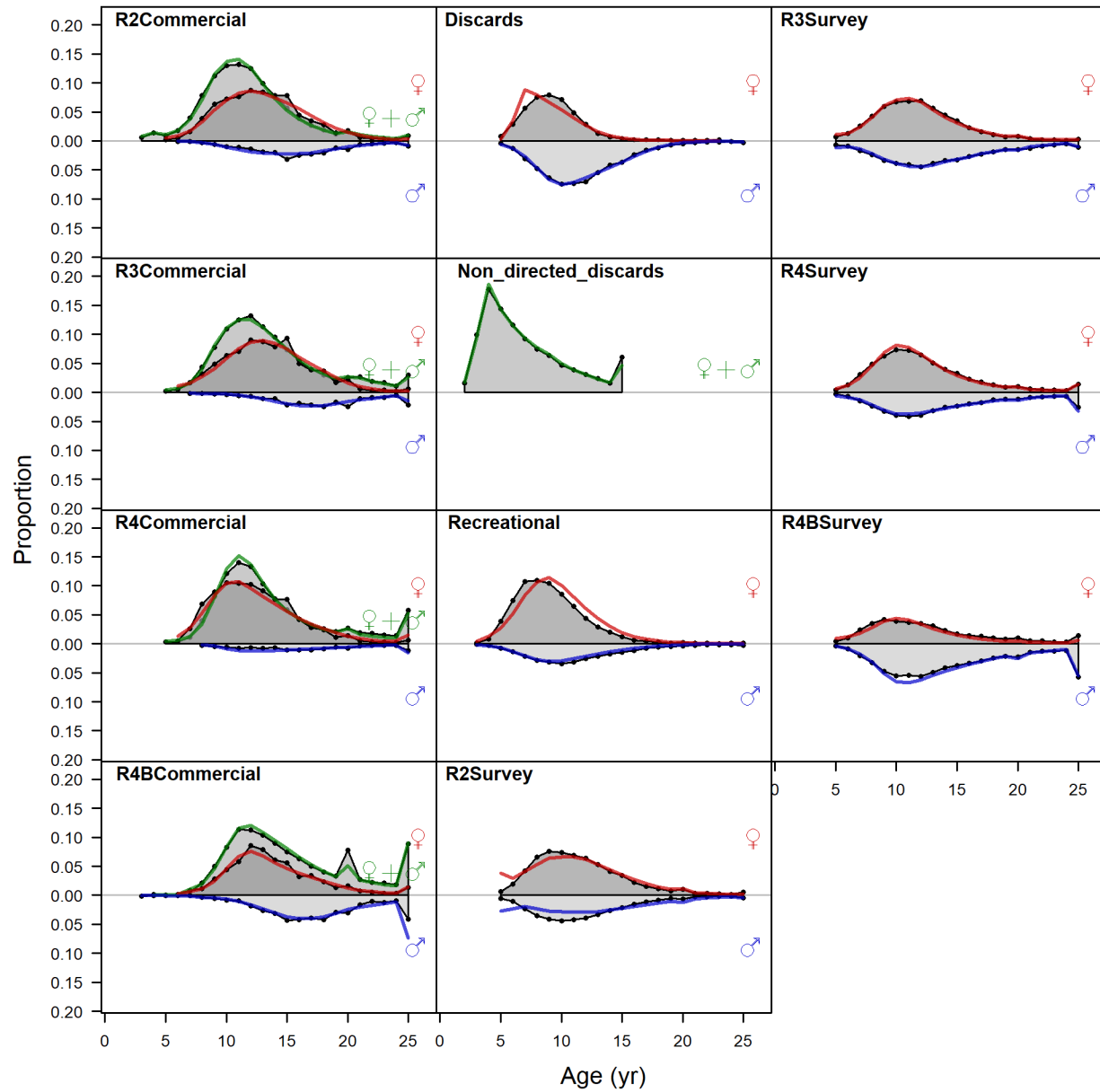**Figure 38.** Fit to survey trends in Biological Regions 2, 3, 4, and 4B (top to bottom) in the AAF short model.
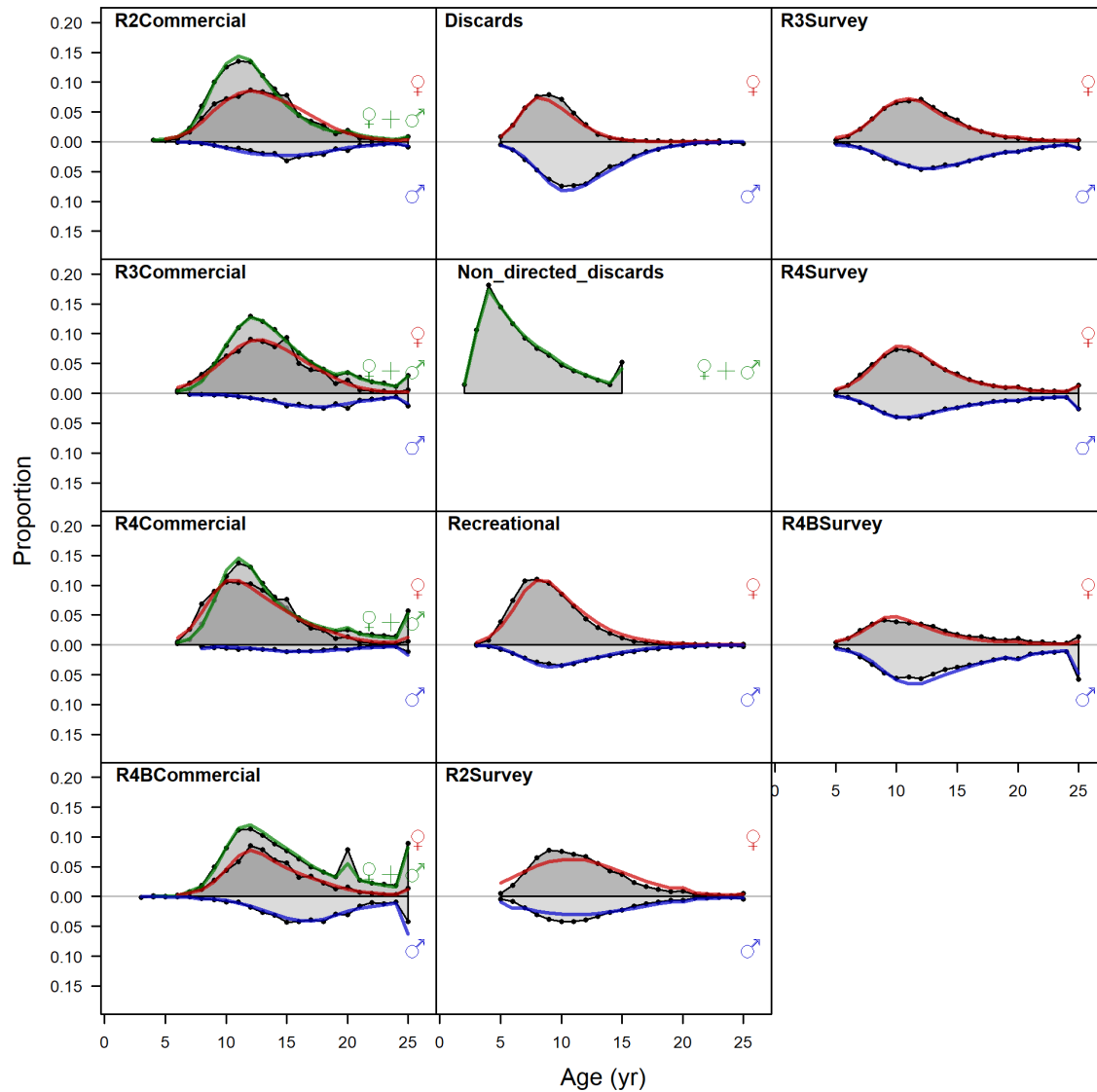
**Figure 39.** Aggregate fit to age data for each model fleet in the AAF short model.
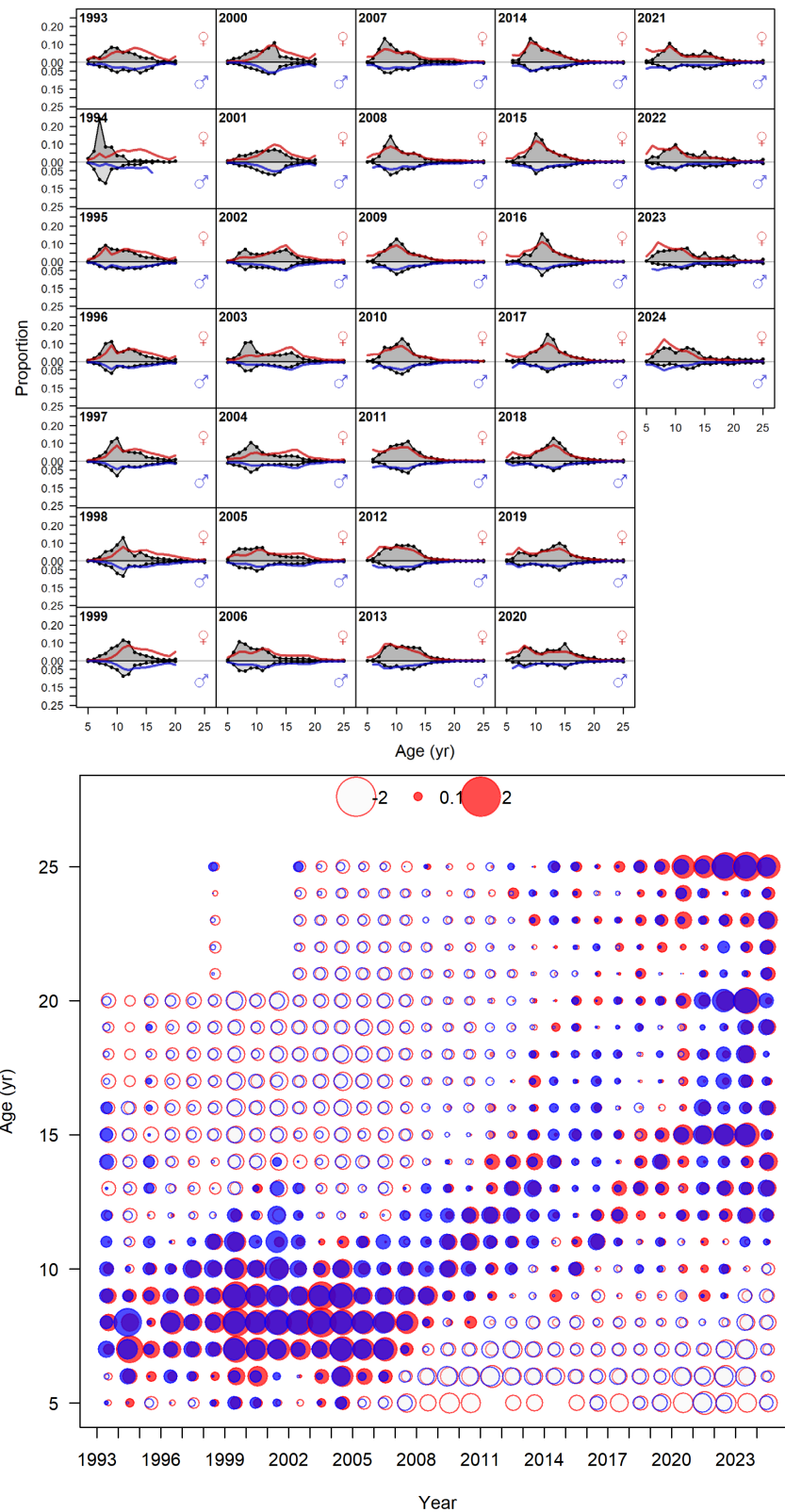
Figure 40. Fit to age data (upper panel) and Pearson residuals (lower panel) from the Region 2 FISS in the AAF short model; red circles denote female residuals, and blue circles denote male residuals.
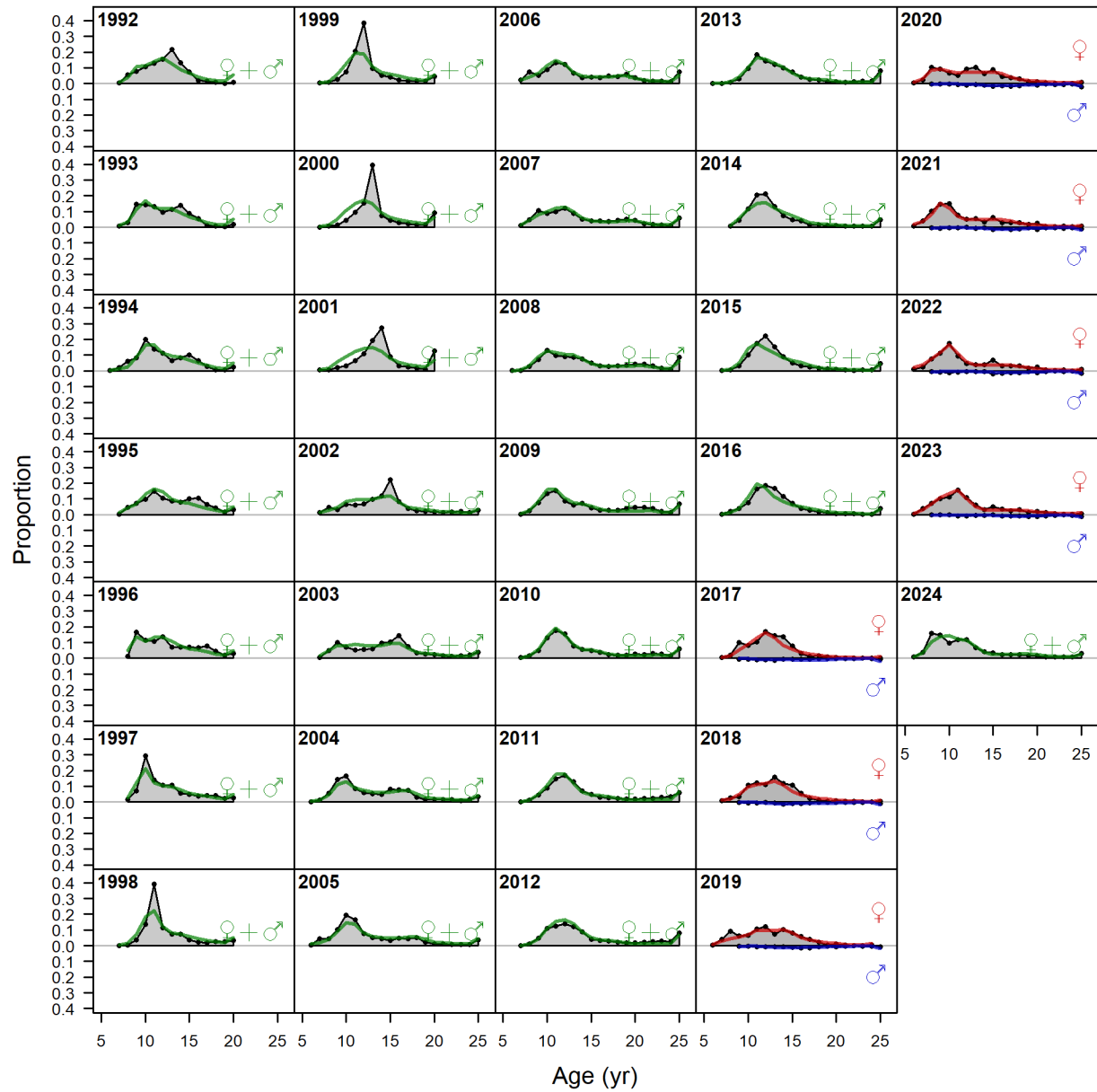
Figure 41. Fit to age data from the Region 4 commercial fishery landings in the AAF short model.
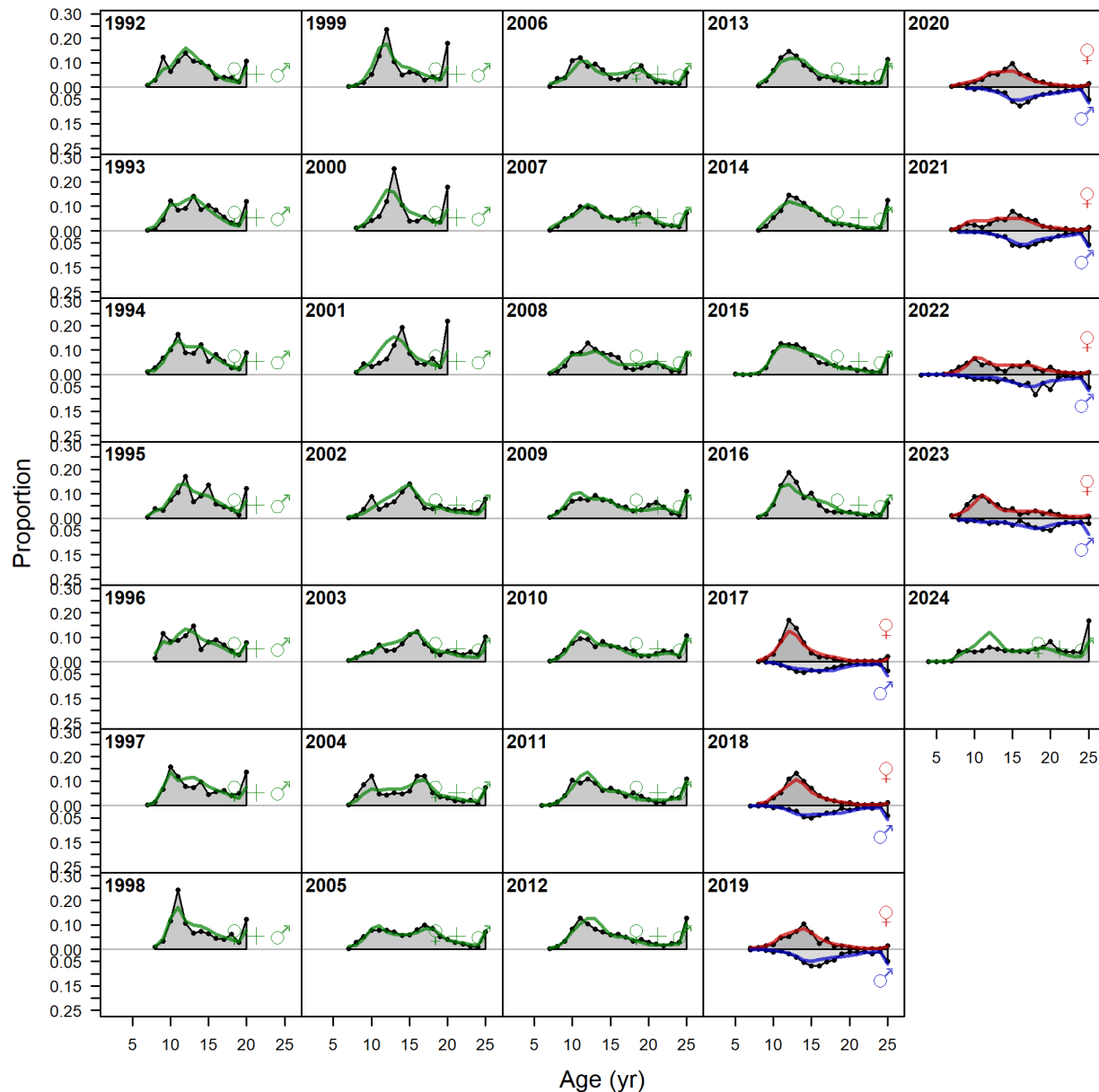
Figure 42. Fit to age data from the Region 4B commercial fishery landings in the AAF short model.

AAF long

Like the AAF short model, the AAF long model fit both the fishery and FISS trends well (Figure 43-44). Aggregate fits to the FISS age composition data showed similar patterns to those observed in the AAF short model (Figure 45). The fit to the FISS age data improved over the time series, but the Region 2 and 3 FISS age data was heavily down-weighted in order to achieve internally consistent weighting (Table 18). This corresponded to poor fits to the Region 2 age data over much of the time series (Figure 46-47). Lack of fit to the Region 3 FISS data occurred primarily in the early part of the time-series (Figure 48-49). Among the fishery fleets, the Region 4 data were most heavily down-weighted from the bootstrapped input sample sizes (Table 18).
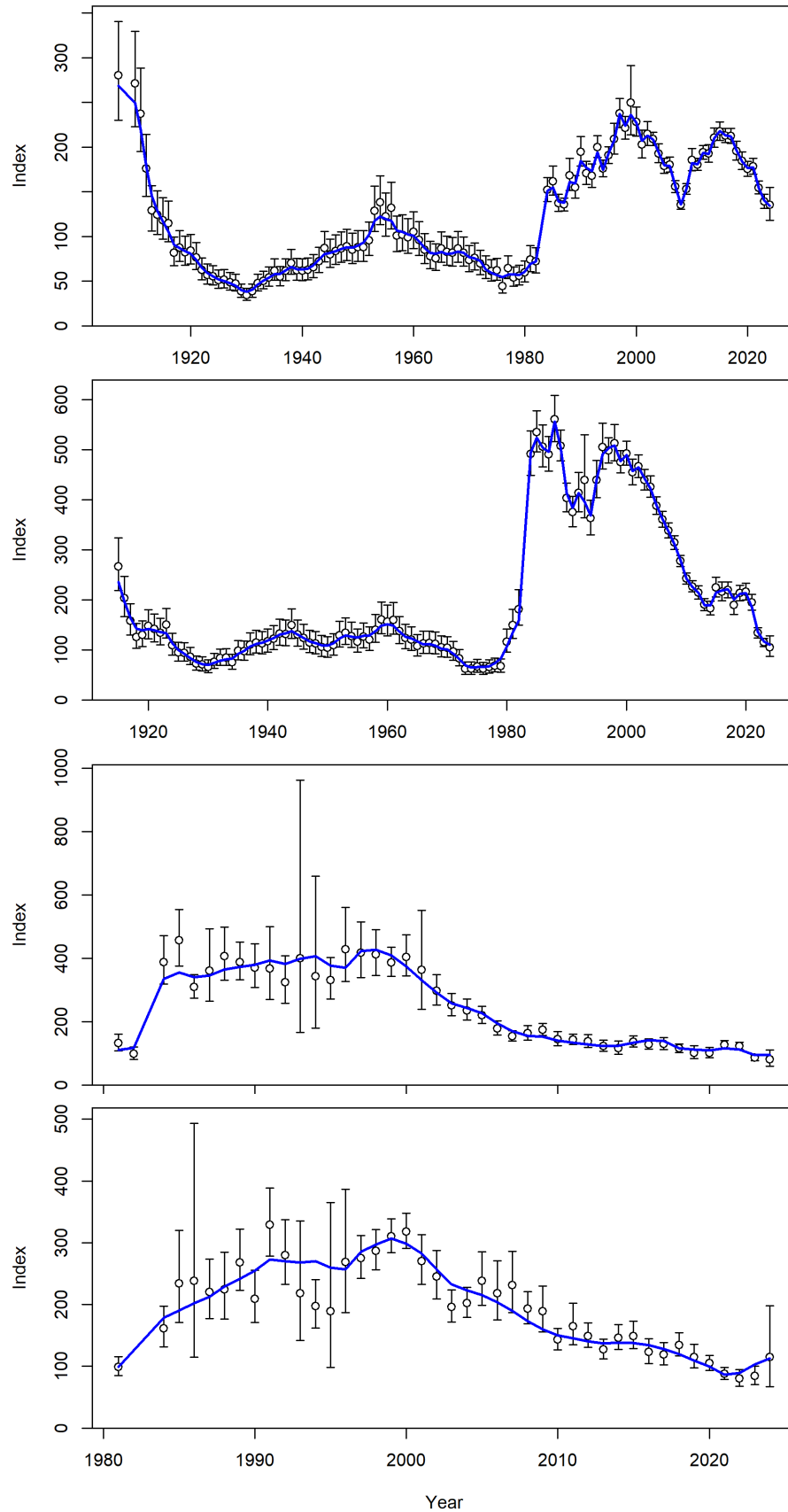
Figure 43. Fit to fishery trends in Biological Regions 2, 3, 4, and 4B (top to bottom) in the AAF long model.

Figure 44. Fit to FISS trends in Biological Regions 2, 3, 4, and 4B (top to bottom) in the AAF long model.

Figure 45. Aggregate fit to age data for each model fleet in the AAF long model.
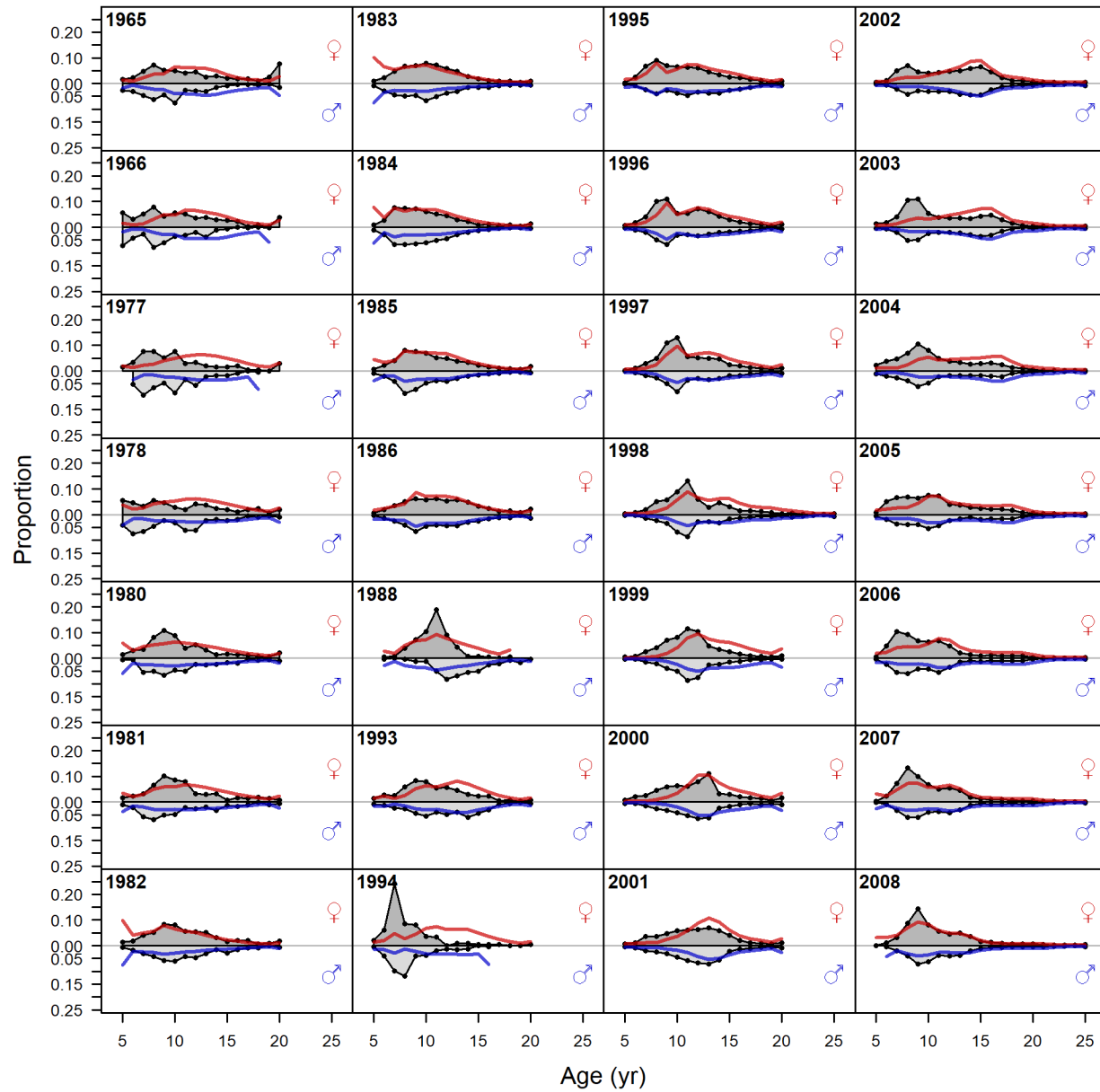
Figure 46. Fit to 1965-2008 age data from the Region 2 FISS in the AAF long model.
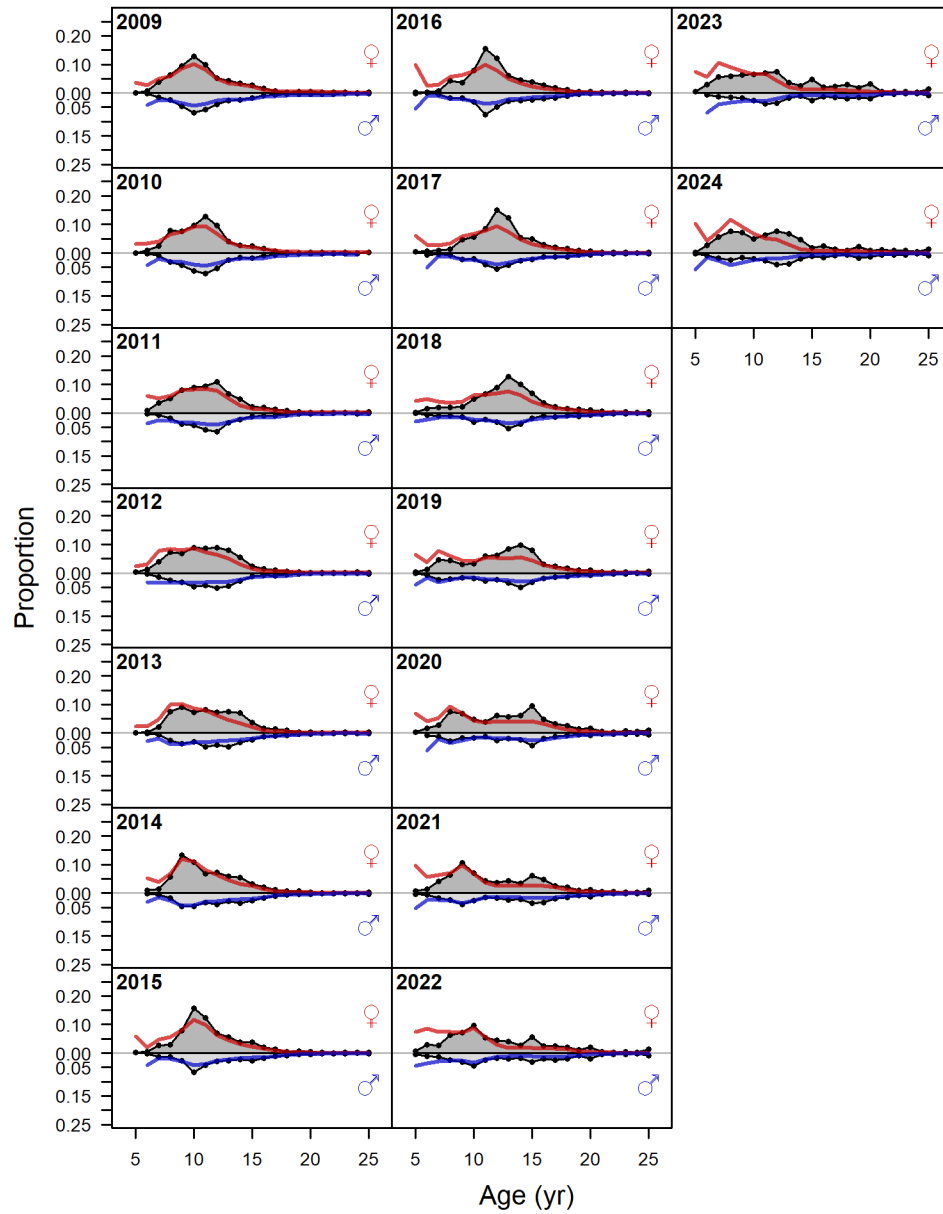
Figure 47. Fit to recent (2009+) age data from the Region 2 FISS in the AAF long model.
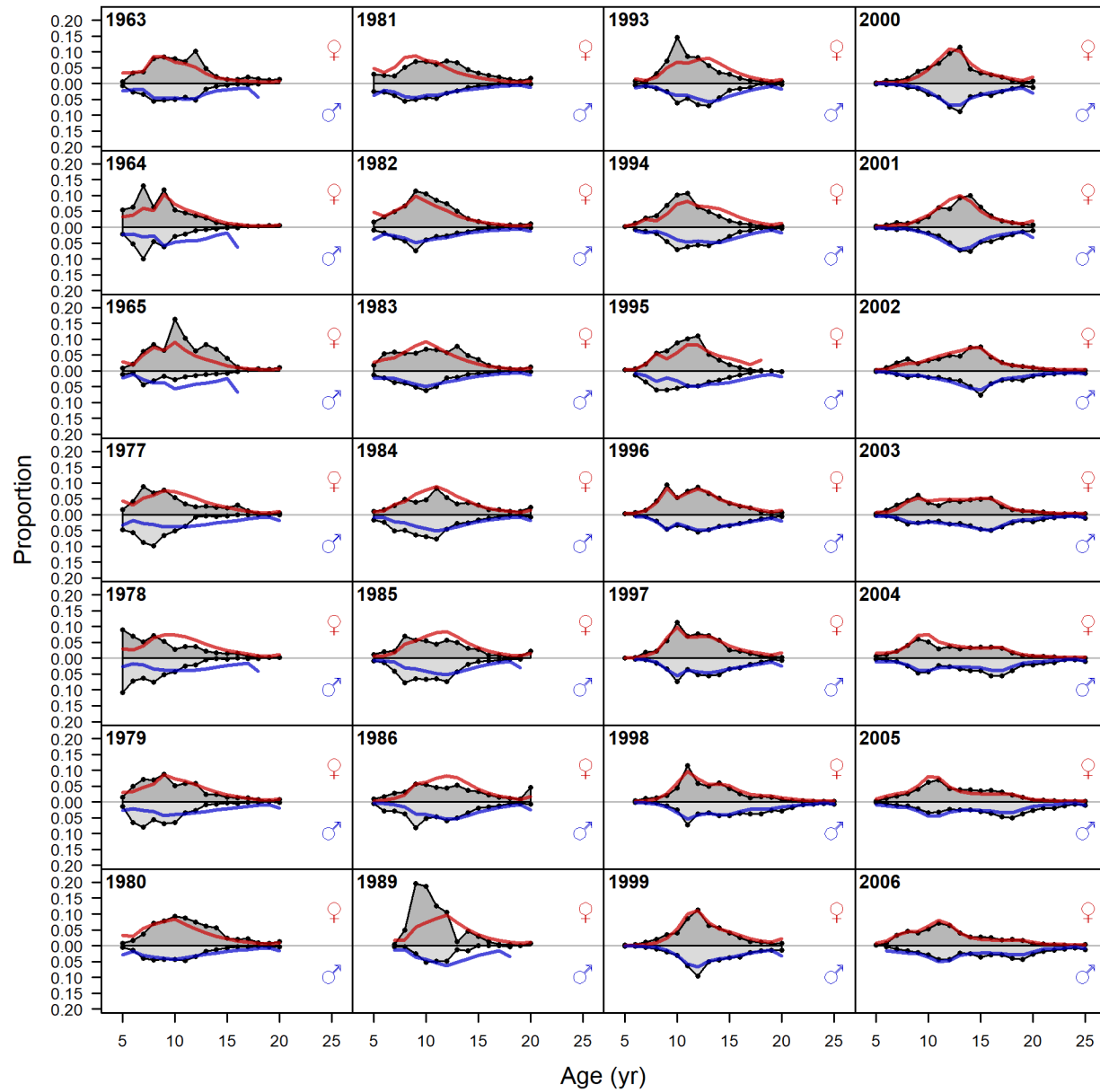
Figure 48. Fit to early age data from the Biological Region 3 FISS in the AAF long model.
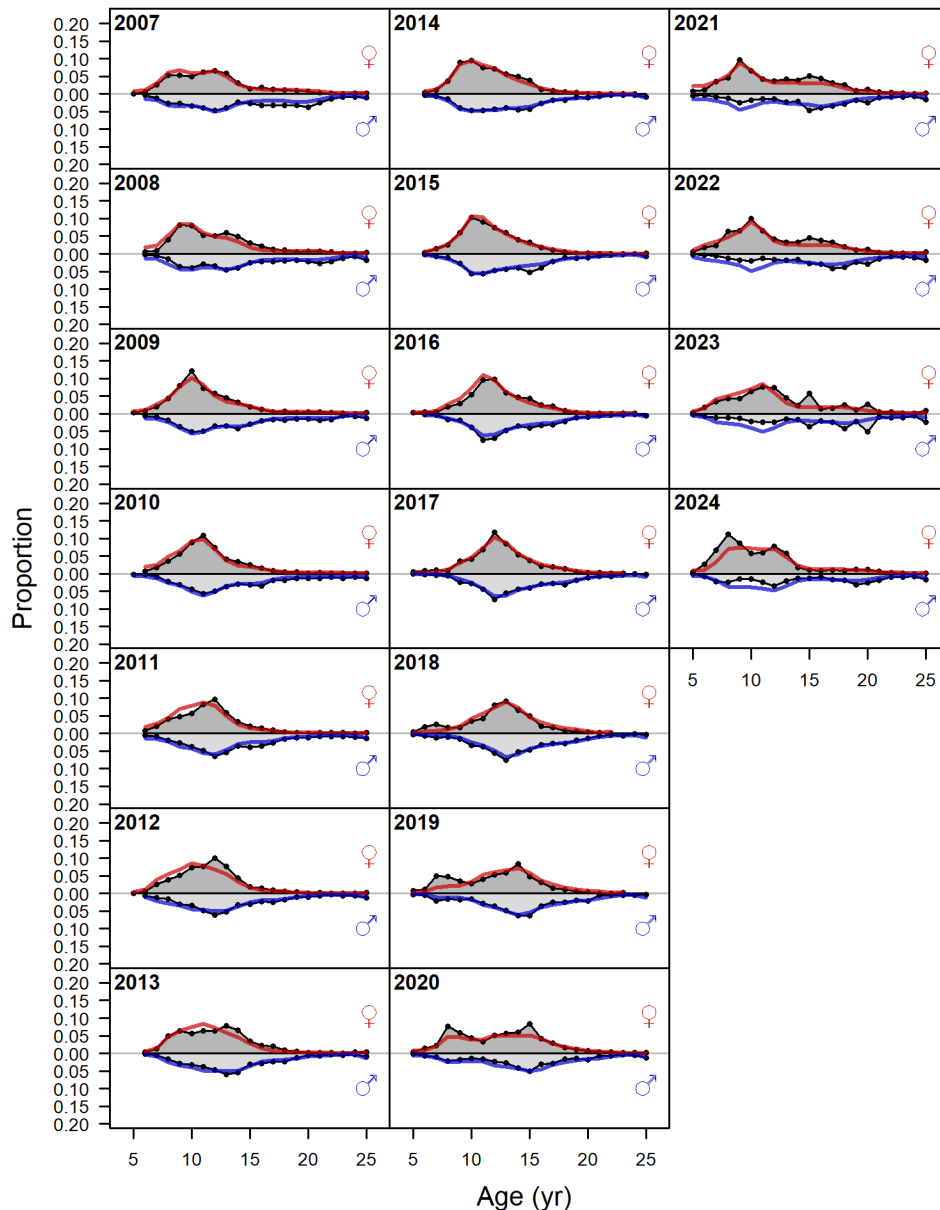
Figure 49. Fit to later age data from the Biological Region 3 FISS in the AAF long model.

Similar to the AAF short model, FISS selectivity was estimated to be asymptotic for Biological Regions 4 and 4B. Peak male selectivity in the commercial fishery landings was also estimated to be asymptotic. All fleets with data extending past the transition from J to circle hooks (1984) showed a strong offset in the unconstrained deviation in catchability for that year (Table 19). Discard and recreational selectivity estimates were similar in the AAF long model to those estimated in the coastwide long model. Non-directed discard selectivity was estimated to be domed, again illustrating the trade-off between domed fleets in the AAF models and asymptotic selectivity over the entire time-series in the coastwide models. This likely interacts with the estimation of natural mortality, producing slightly lower values in the AAF long model (0.186 for females, and 0.163 for males) than in the coastwide long model (Table 19). The environmental link coefficient was estimated to be slightly weaker (0.430) than in the coastwide

long model, although the 95% interval still did not approach zero indicating a highly significant relationship (Table 19). The AAF long model produced intermediate estimates of recent recruitment and female spawning biomass compared to the other three model (Table 19). This result is consistent with the intermediate estimates of male and female *M* from this model.

Summary of strengths and weaknesses for the AAF long model:

Strengths:

- Includes uncertainty in female and male *M*
- Includes extensive historical data
- Uses environmental information to inform recruitment
- Fits the fishery and survey indices well
- Fits both the Regions 2, 3 and 4B fishery age data well
- Fits Region 4 and 4B FISS age data well
- Internally consistent data weighting

Weaknesses:

- Highest technical overhead (complexity) of the four models
- Most challenging model to check and ensure reliable convergence
- Relies heavily on only fishery trends over the historical period
- Implicitly assumes stationarity in some processes (e.g., the stock-recruitment function, *M*) over the long historical period
- Fit Biological Regions 2 and 3 survey age data poorly

Sources of uncertainty

The four models evaluated here represent, within the set itself, significant sources of uncertainty in how to treat the data (partitioning by fleets or aggregating to a single series), as well as how to treat the time-series (emphasizing the recent dynamics or including more historical information). Further, the differing assumptions of fixed vs. estimated female natural mortality rate and the treatment of environmental covariates to the stock-recruitment relationship are also embedded in the differences observed among the four model results. These factors lead to important differences in both scale and trend. In aggregate, the four models together reflected much more uncertainty than would any single model. However, it is notable that the data remain generally informative of a similar population scale and recent trend for both spawning biomass and recruitment.

Sensitivity analyses

Many alternative model configurations were evaluated during model development, but only a subset of these is reported here. Several of the bridging steps from the 2024 models to the 2025 preliminary models described here also represent sensitivity analyses. The focus of the analyses described below was on model behavior and understanding; sensitivity analyses specifically intended to highlight the importance of ongoing research (e.g., whale depredation, maturity

ogives, etc.) are produced each year as part of the final stock assessment (Stewart and Hicks 2025).

The large differences in the scale of the spawning biomass in the historical period between the two long time series models represent importantly differing assumptions about the connectivity of the stock via spatial availability (Figure 50). Specifically, domed selectivity for Biological Regions 2 and 3 in the long AAF model implicitly assumes that older fish (located in northern and western areas) were historically less available and therefore not mobile enough to be readily available to those fisheries. Conversely, in the coastwide long model the assumption of asymptotic selectivity implies a high degree of availability and therefore connectivity between all geographic components in the population. Sensitivity analyses in the 2015 assessment indicted that these two models could be made much more similar by adjusting the degree of domed selectivity (Stewart and Martell 2016). The use of both models encompasses the range of uncertainty that exists over this aspect of the historical population dynamics, thus the primary sensitivity in the stock assessment is included in the ensemble results.
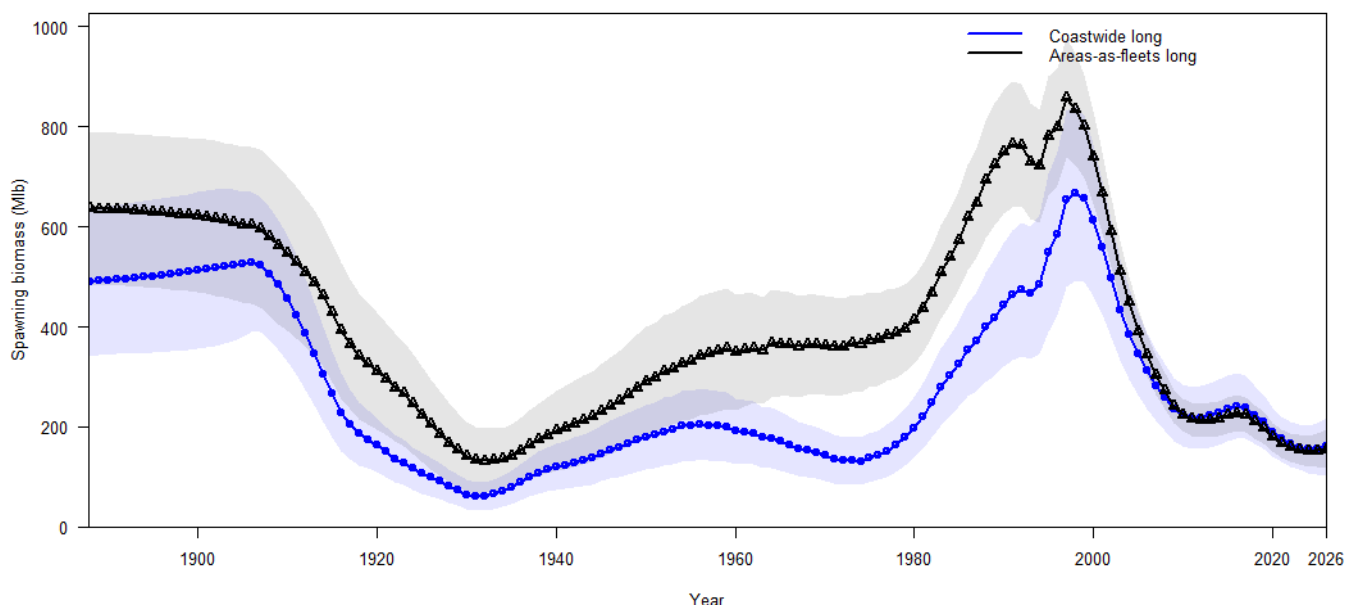
Figure 50. Comparison of the spawning biomass for the long coastwide and AAF models.

The treatment of the PDO in the two long time-series models was explored extensively as part of the 2022 full stock assessment. As described above, the current approach classifies the PDO into a series of binary 'regimes', and then estimates a coefficient describing the effect of these regimes on the equilibrium recruitment used in the stock-recruitment relationship. There is still considerable variability remaining in the annual recruitment deviations and the overall effect on the estimates of recruitment from the use of this covariate primarily occur at the end of the time-series when there is little other information to inform recruitment estimates. Due to increasing evidence that the environmental and oceanographic conditions associated with the PDO may be changing (Litzow et al. 2020), it is possible that at some point the use of the PDO as a covariate will no longer provide an improvement to the Pacific halibut models. To explore how

this might affect model results, a sensitivity analysis of the two long time-series models was conducted by removing the PDO entirely. Results indicate an increased biomass in the very early part of the modelled period, but either little change (AAF long) or a slight increase in spawning biomass (CW long) in the most recent years (Figure 51).
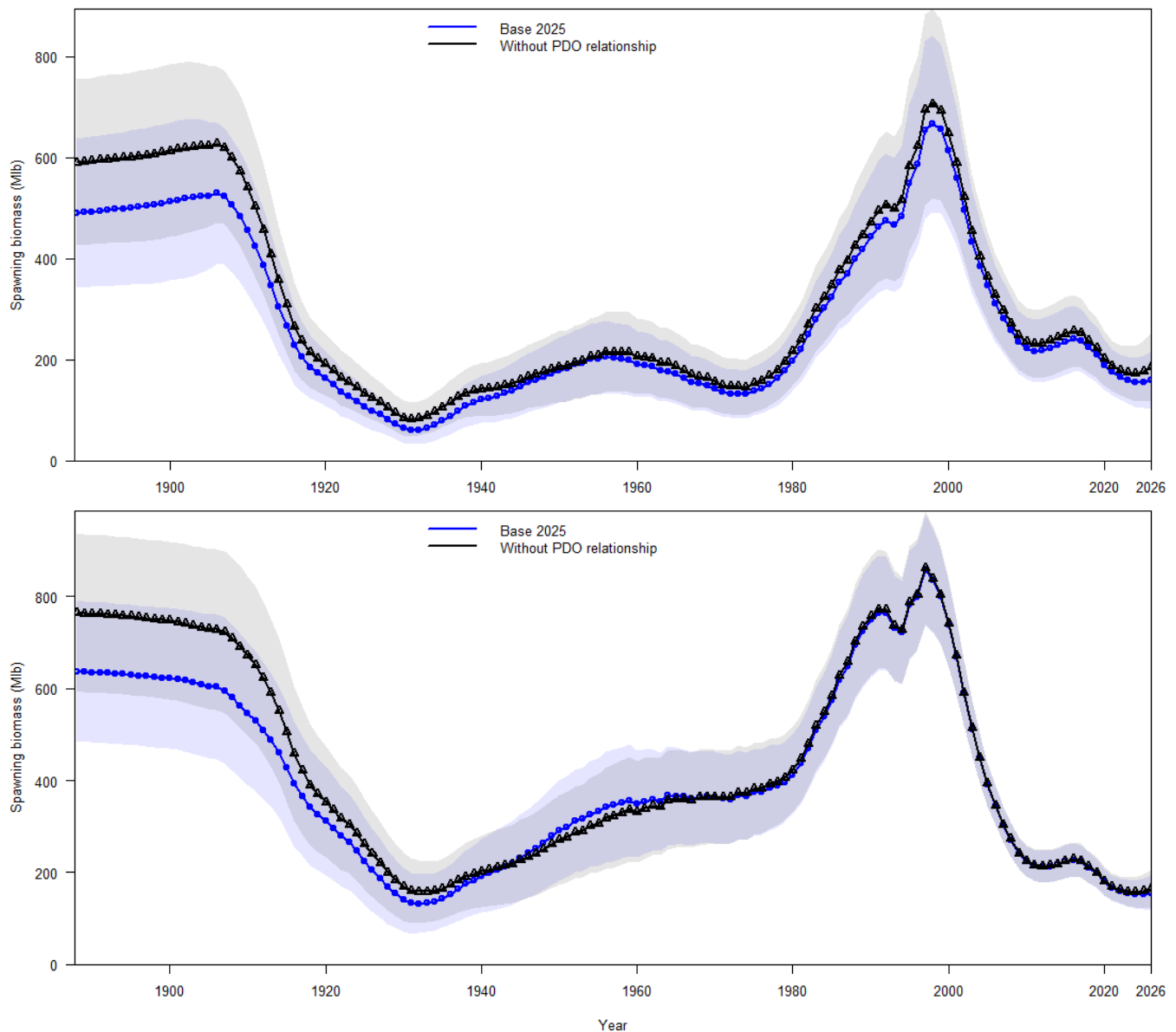
Figure 51. Comparison of the spawning biomass for the long coastwide (top panel) and long AAF (bottom panel) models with and without the PDO relationship included.

To further explore the sensitivity of the stock-recruitment relationship in the Pacific halibut models a sensitivity analysis to the value for steepness (*h*) was also performed. Each of the four models uses a fixed value of 0.75 as the base case. Previous assessments and other supporting analyses have found that this choice provides for modest feedback between spawning biomass and subsequent recruitment but does not have a strong effect on the modelled dynamics. In contrast, a fixed steepness is known to have a very important effect on reference points that rely

on the stock-recruitment relationship (e.g. MSY; Mangel et al. 2013). For this reason, the MSE operating models used by the IPHC include additional variability in steepness beyond what is used in the stock assessment. A sensitivity to higher and lower values of steepness showed that the coastwide short model (Figure 52) and the AAF short model (Figure 53) spawning biomass estimates were largely unchanged for alternative values of steepness. Recruitments tended to be estimated slightly higher at the end of the time-series for lower values of steepness, indicating there may be a small effect on forecasts. The coastwide long model showed the greatest sensitivity to steepness with a slightly larger spawning biomass at lower steepness, albeit with a very similar trend over time (Figure 54). The spawning biomass from the AAF long model only differed at the beginning of the modelled period and was almost identical in the most recent years (Figure 55). As currently configured, this sensitivity analysis indicated that the assumed value for steepness was not critically important to the stock assessment results over the range of values considered plausible for flatfish (Myers et al. 1999).
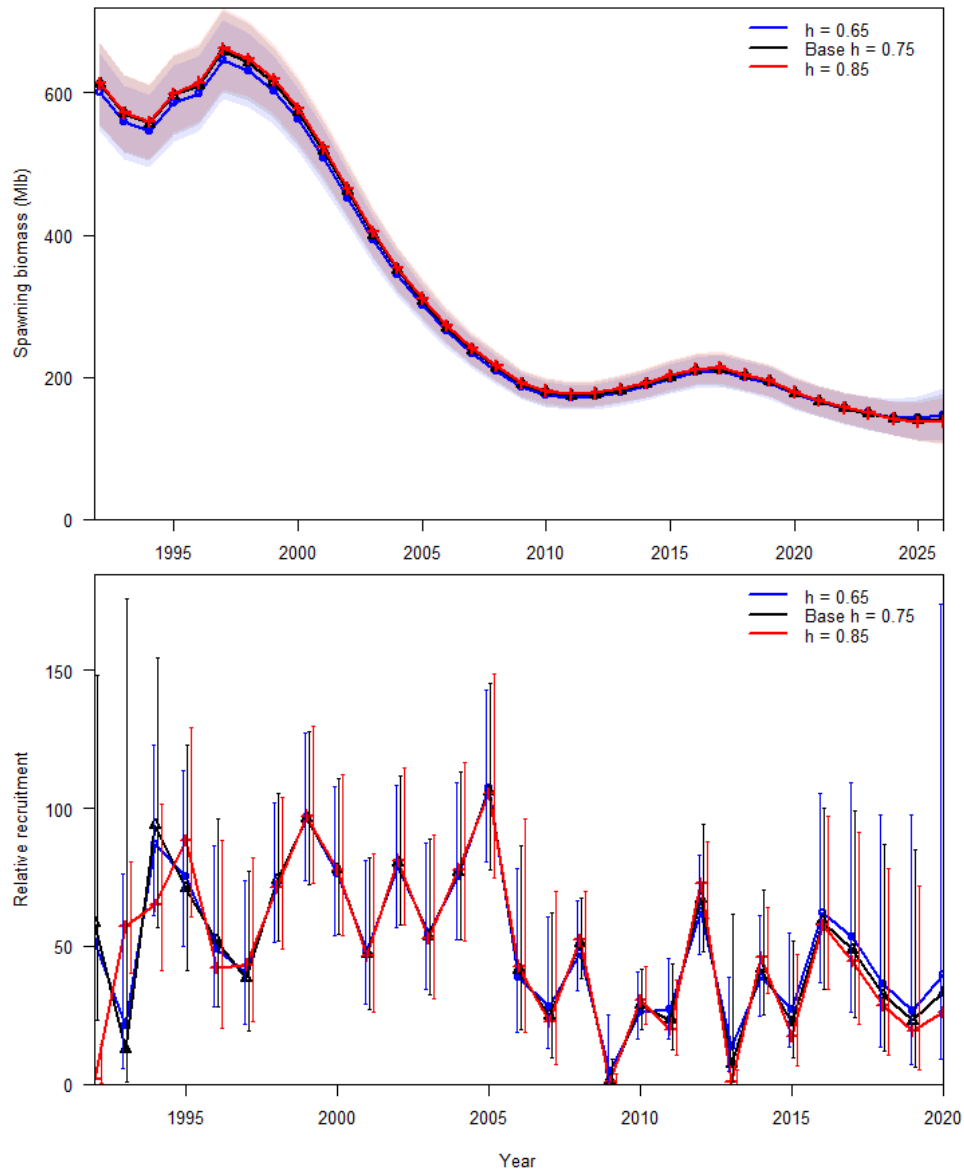
Figure 52. Sensitivity of the CW short model to alternative values of steepness.
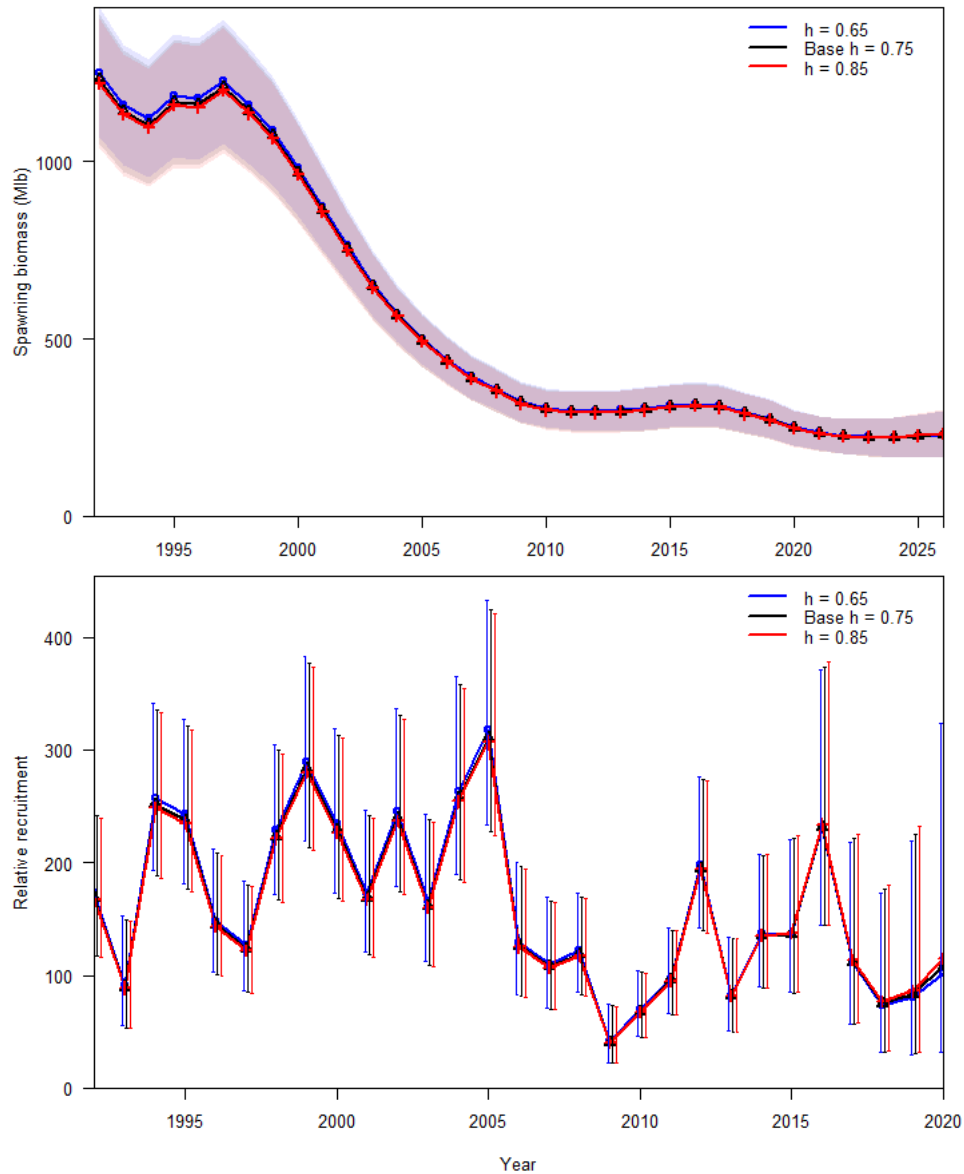
Figure 53. Sensitivity of the AAF short model to alternative values of steepness.
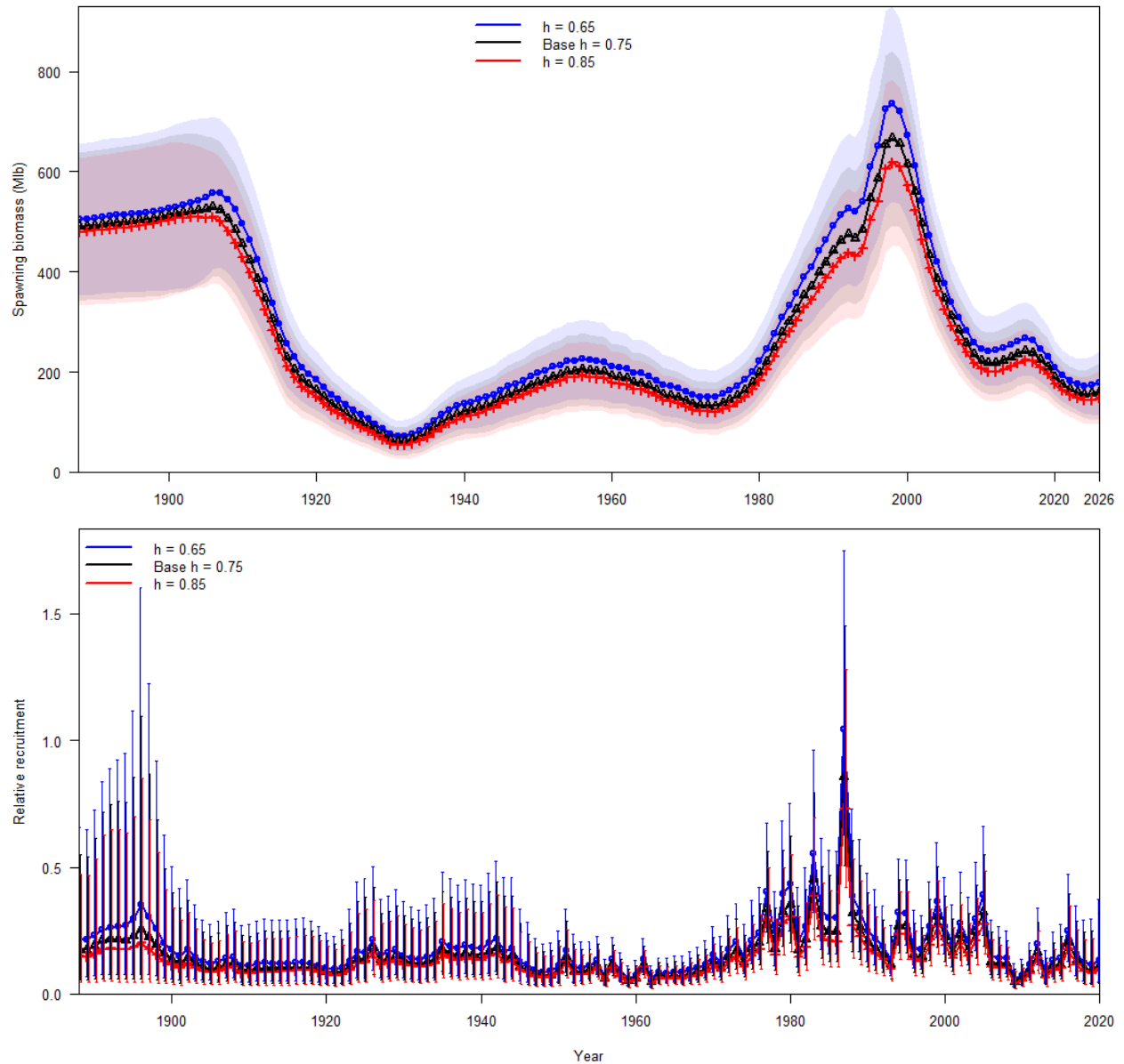
Figure 54. Sensitivity of the CW long model to alternative values of steepness.
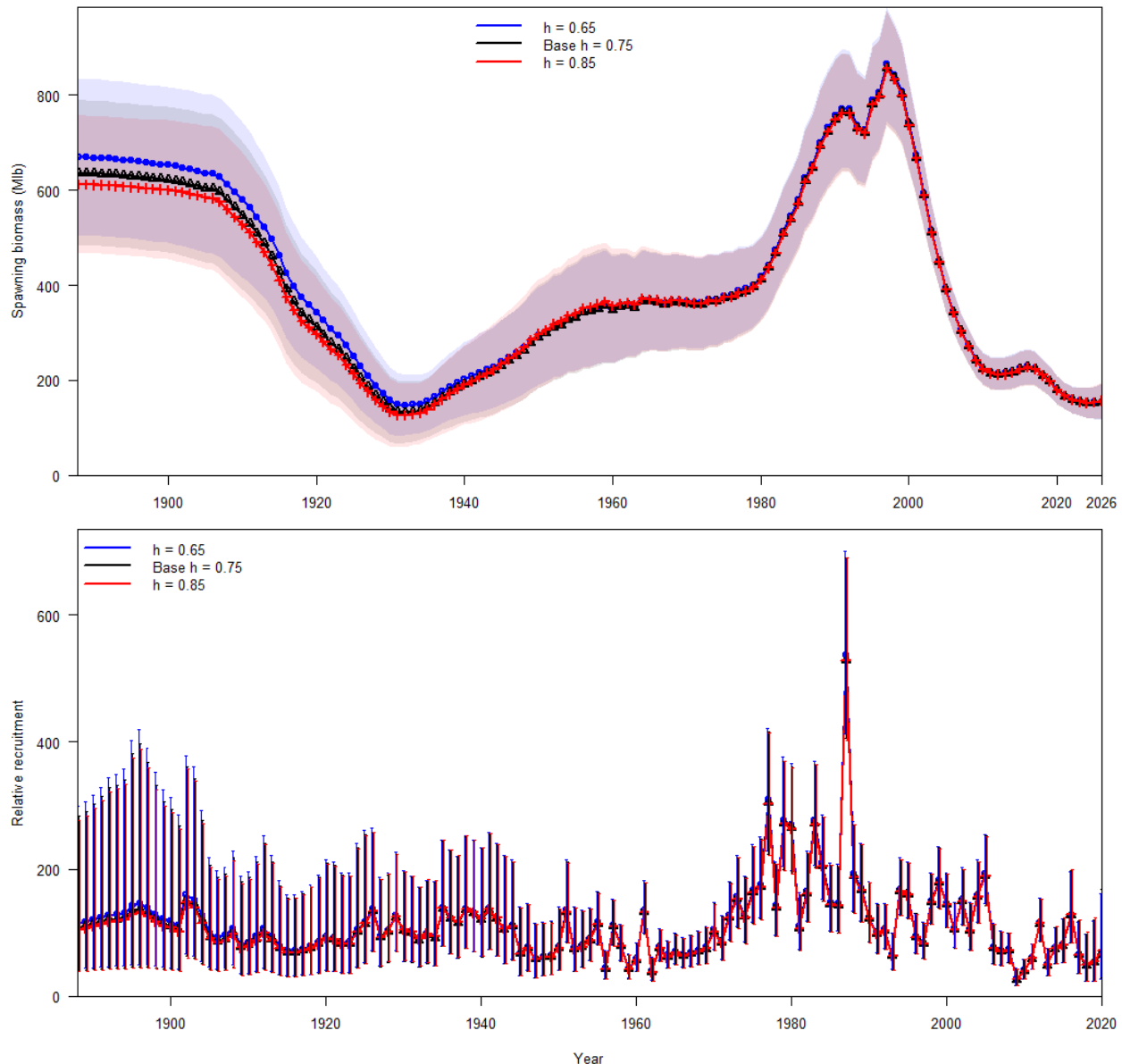
Figure 55. Sensitivity of the AAF long model to alternative values of steepness.

The next sensitivity analysis focused on the fixed value of natural mortality used in the coastwide short model. Previous stock assessments have shown that the scale of the estimated spawning biomass and recruitment is very sensitive to natural mortality, and the coastwide short model is the only model where this value is not estimated, a topic of substantial exploration in the 2022 assessment (Stewart and Hicks 2022). Models were fit assuming a fixed value higher (0.18) and lower than the base case (0.15) natural mortality for female Pacific halibut (the value for males is estimated). Results were consistent with previous assessments showing larger biomass estimates for higher values of female natural mortality, but little difference in estimated spawning biomass trends or relative recruitment strengths (Figure 56). Extensive exploration of this model and the potential for estimating this parameter did not indicate a model configuration that produced a reliable value - all tended to favor much higher estimates at whatever upper bound

was specified. The estimation of relative male:female selectivity parameters reduction of time-varying processes, as well as different (non-iterated) values for the standard deviation of recruitment variability all produced similar behaviour. At this time it was concluded that natural mortality was not able to be reliably estimated in the coastwide short model. As discussed in the 2022 stock assessment, it would be possible to include uncertainty in the fixed value of natural mortality used in this model via inclusion and appropriate weighting of alternative values in the stock assessment ensemble.
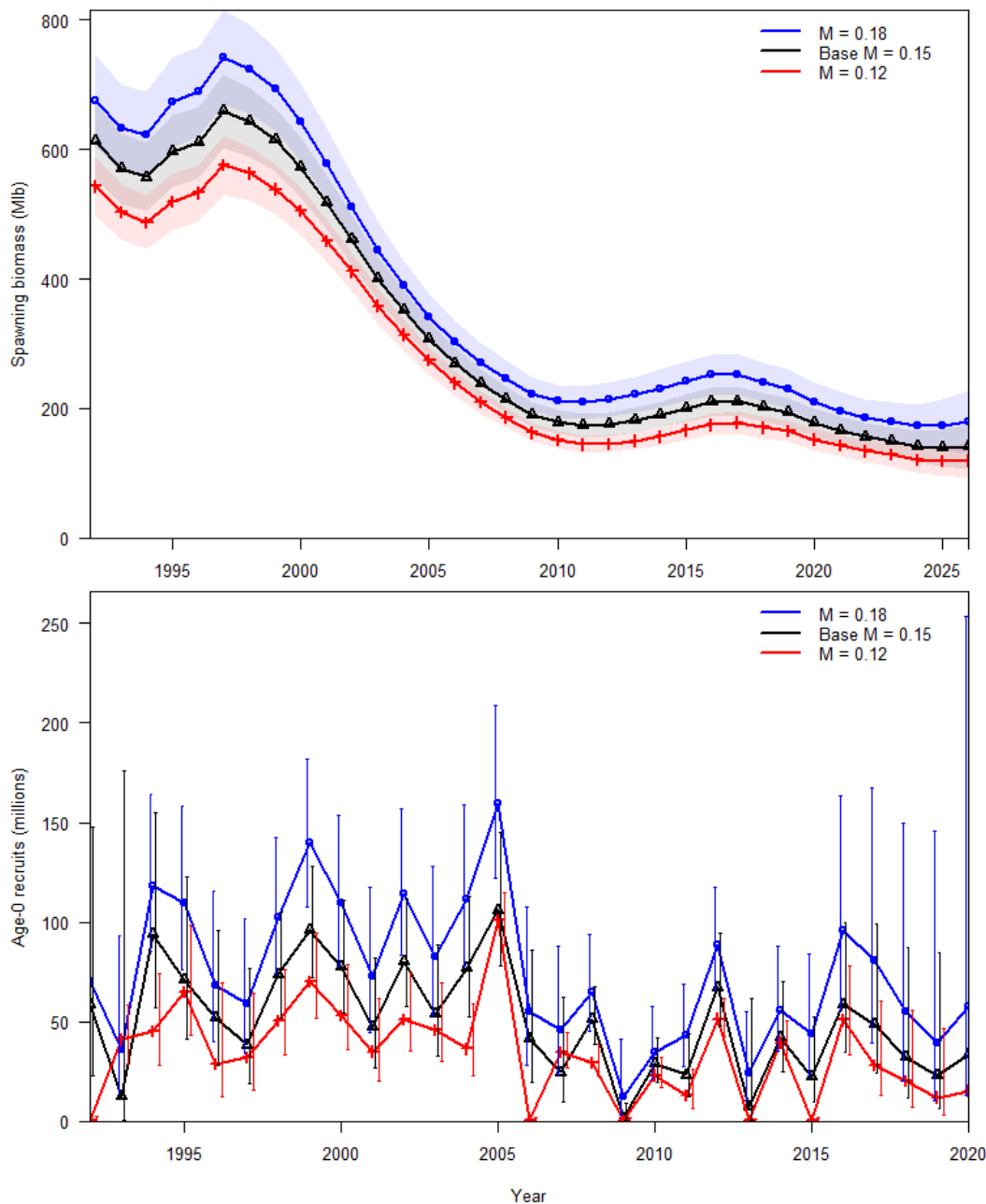
Figure 56. Sensitivity of the coastwide short model to alternative fixed values of female *M*.

Sensitivity to the revised maturity ogive used as the base case for this preliminary stock assessment was included as a step in the bridging analysis described above. The modelled ogive was truncated below the youngest age for which a female Pacific halibut has been observed (age 7). This choice was made to avoid assuming even a small fraction of the much more numerous younger ages was mature without clear evidence suggesting this might be the case. To explore how sensitive the models might be to this choice, alternative models were run without truncating the ogive and allowing a small fraction of the age-5 and age-6 females to be mature (Figure 8). All four models estimated a larger spawning biomass with the updated maturity ogive compared to the historical curve shifted toward older fish (Figure 57-60). The non-truncated ogive had little effect on model results, with only a very slight increase in the estimated spawning biomass.
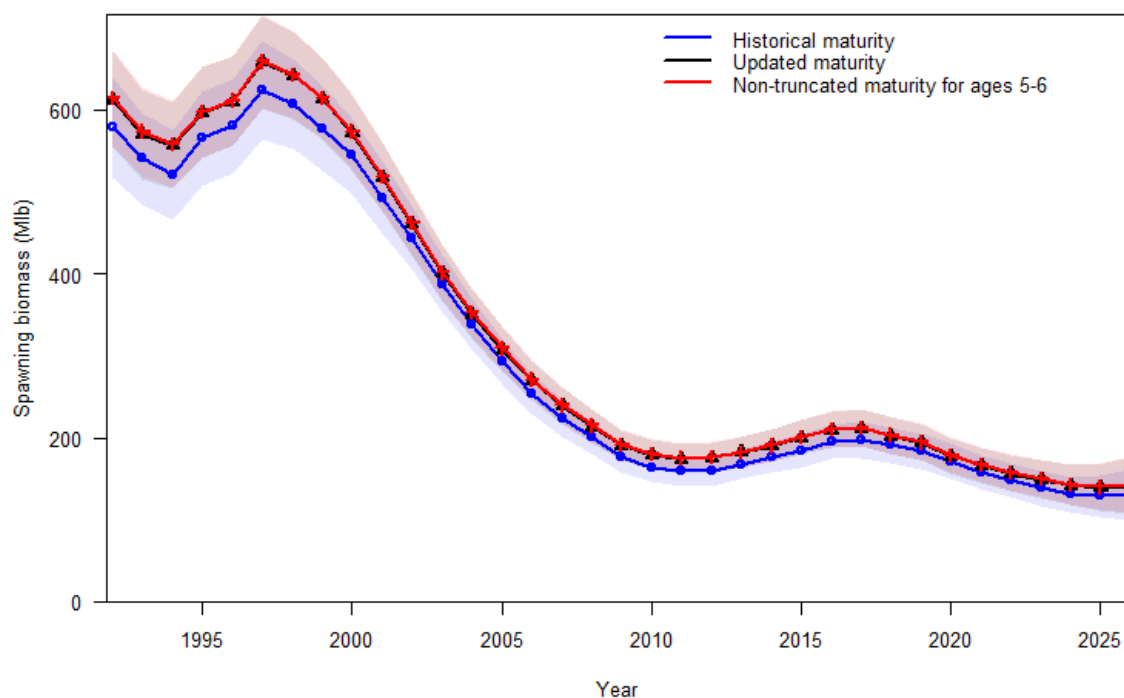
Figure 57. Sensitivity of the coastwide short model to the historical, updated, and non-truncated updated maturity ogives.
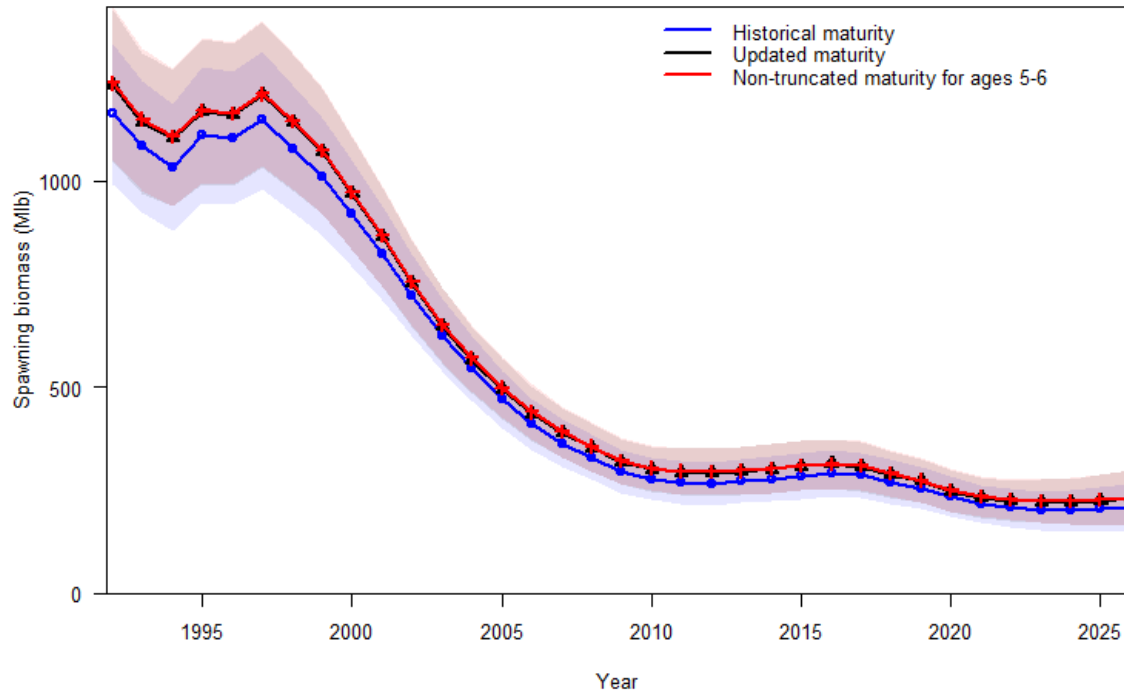
Figure 58. Sensitivity of the AAF short model to the historical, updated, and non-truncated updated maturity ogives.
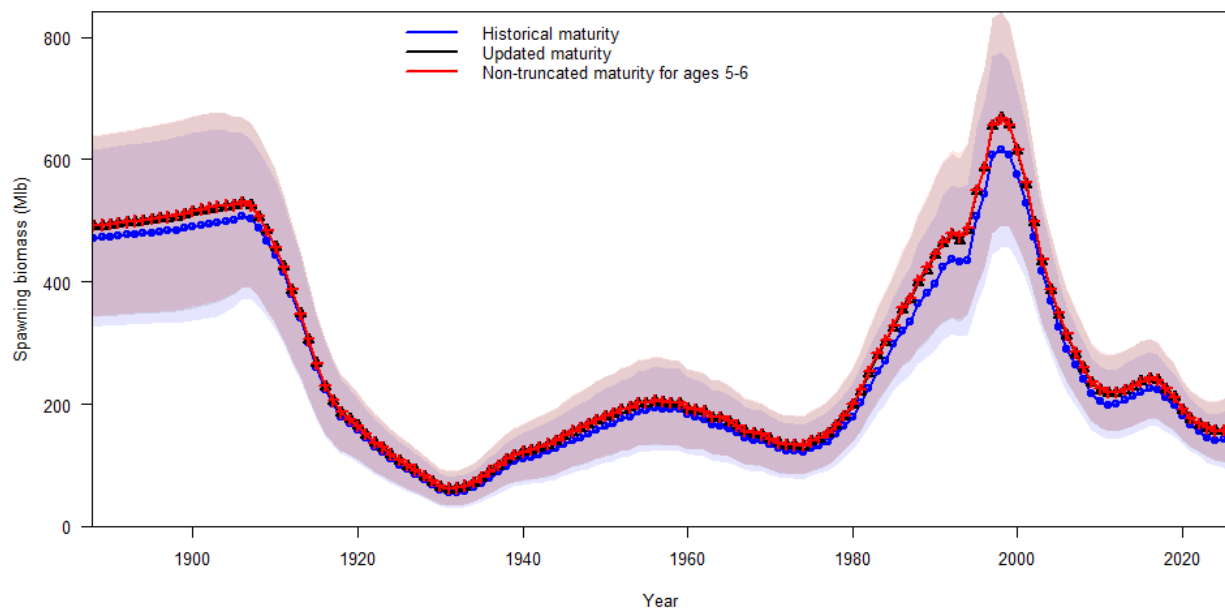
Figure 59. Sensitivity of the coastwide long model to the historical, updated, and non-truncated updated maturity ogives.
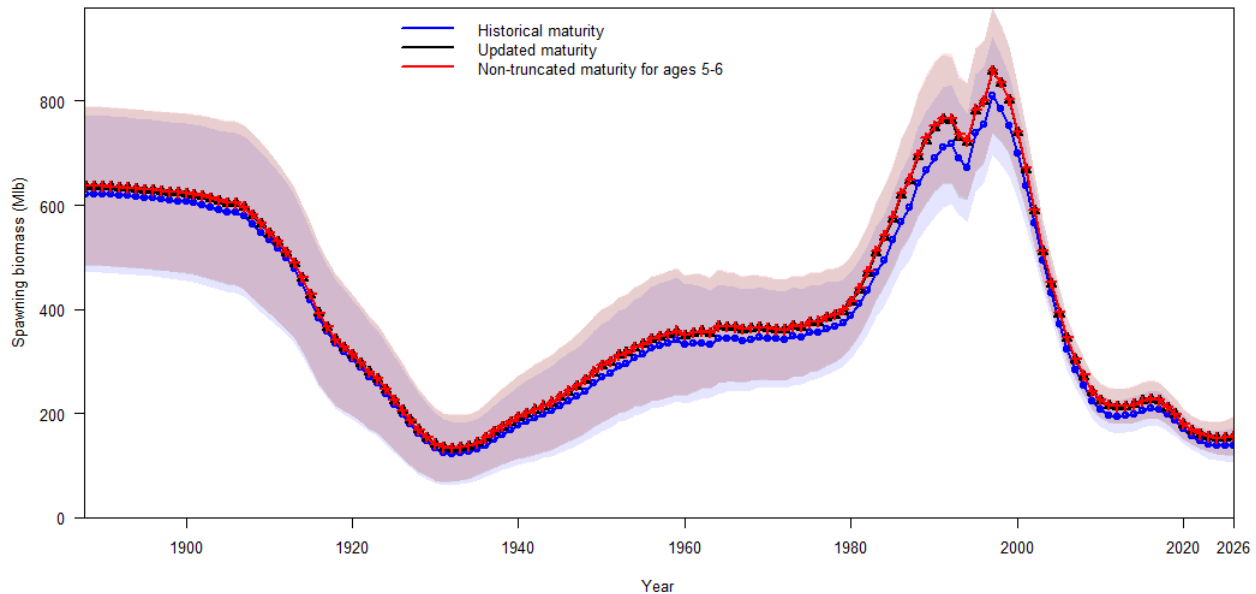
Figure 60. Sensitivity of the AAF long model to the historical, updated, and non-truncated updated maturity ogives.

Additional sensitivity analyses were explored but did not or were not intended to produce reliable models for consideration. The first of these represented an effort to address the lack of fit to Region 2 FISS age composition data. Lack of fit in the coastwide short model showed a clear pattern of large positive residuals at younger ages in the early time-series (until the mid-2000s) and large positive residuals for the older ages in the latter part of the time series (Figure 40). Models were fit to each part (early vs late) of the FISS age composition data separately (but only one of the two periods at the same time) and achieved much improved fit with differing selectivity. However, a similar fit was not produced even when selectivity was allowed to change greatly over this same period. This indicated that there was a catchability component: it appears that spatial availability, particularly for Region 2, may have shifted over time to a degree that cannot be fully captured with changes in selectivity alone. Future modelling could consider allowing time varying catchability (but this would greatly reduce any information in the survey index) or further explore explicitly spatial models.

An additional sensitivity explored a question often raised during public interactions: Could the recent low recruitment since 2006 be explained by increased whale depredation? One hypothesis is that the estimated reduced recruitment may be a function of increased whale depredation on these year classes as they are entering the Pacific halibut fisheries. To explore this hypothesis, the commercial landings and discards were inflated by 50% in each of the four models beginning in 2010 (around the time the 2006 cohort were first entering the catch and being discarded). Models were run with only this change, retaining all other data, but re-estimating all model parameters. The results showed a slightly larger estimated spawning biomass and virtually no change in the time-series of relative recruitment (Figure 61). This result is not unexpected, as the relative recruitment strengths are largely dictated by the

compositional data which were unchanged in this sensitivity analysis. A similar analysis, with a three-fold increase in commercial catch produced a much larger spawning biomass but little change in relative recruitment. This sensitivity provides a response to stakeholder concerns and supports the conclusion that recent low productivity is not a direct result of unobserved mortality due to whale depredation on the directed commercial halibut fishery.
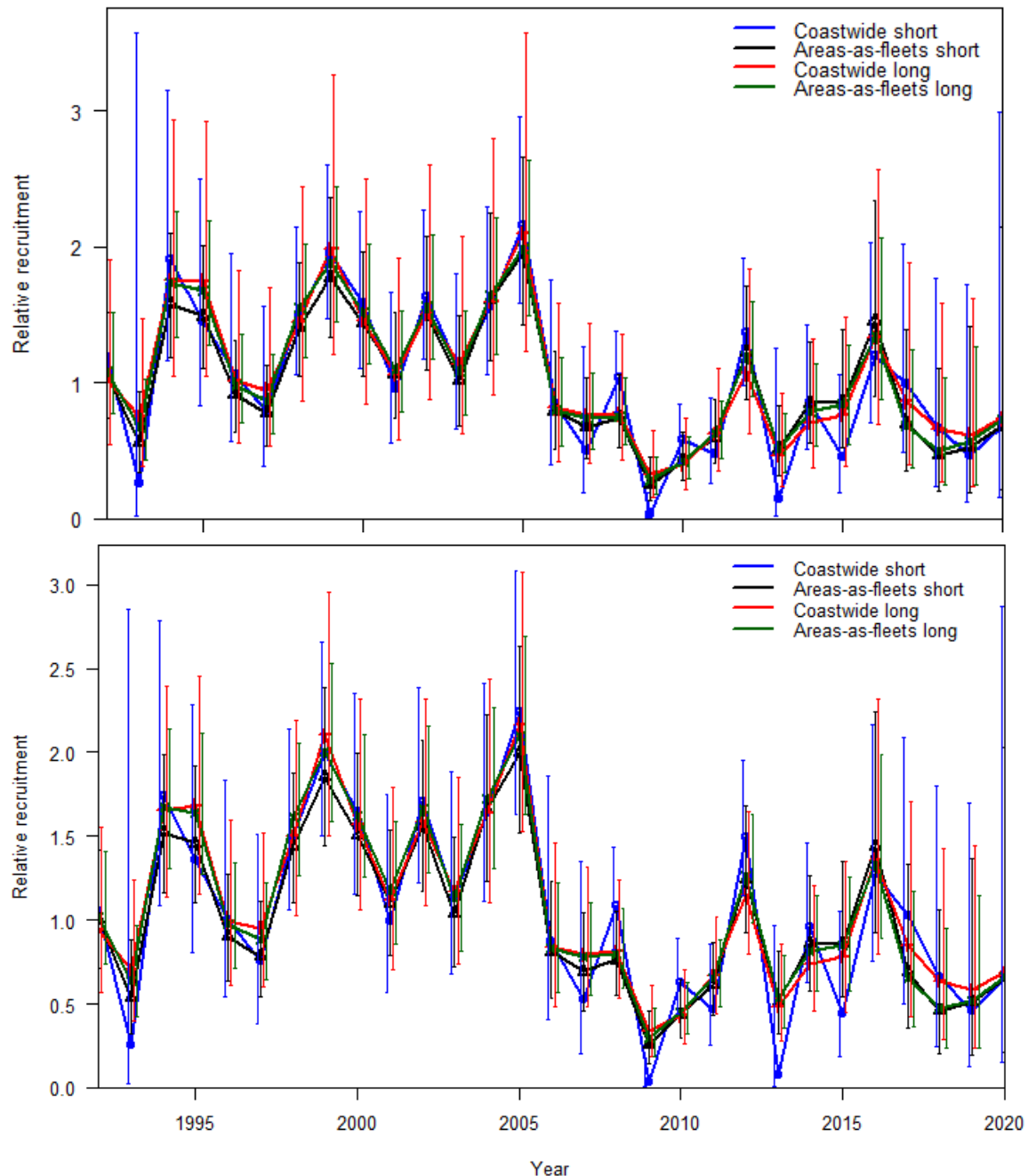

Figure 61. Relative recruitment estimates (divided by the mean of each model) for the preliminary stock assessment (upper panel) and an alternative model assuming whale depredation on the commercial catch (landings + discards) of 50% (lower panel).

Retrospective analyses

The halibut model used from 2006 until 2011 was plagued by a very strong retrospective pattern, both in the scale of the most recent stock size estimates as well as the trend in those estimates (Stewart and Martell 2014; Stewart et al. 2013a). The solution to this problem was additional flexibility for process error (temporal variability) in the selectivity curves for both the fishery and survey representing not just gear (or 'contact') selectivity but also spatial availability.

Retrospective analyses were conducted for these preliminary 2025 models by sequentially removing the terminal eight years of data from the model (a seven-year retrospective, since the terminal year currently contains no information other than mortality projections). Limiting this approach to the most recent eight years of data allows the models to be informed by at least one year of commercial fishery sex-ratio data (2017).

The coastwide short model showed very little retrospective change as the terminal years of data were removed (Figure 62). The AAF short model showed a trend toward higher biomass estimates with a similar trend as data were sequentially removed (Figure 63). This indicates an updating of information informing scale in this model with the most recent observations. Somewhat differently, the coastwide long model showed some increase for some of the most recent years but did not show a strong increase across the entire time-series (Figure 64). Finally, the AAF long model showed a positive retrospective pattern that had changes in both the scale and recent trend (Figure 65). These patterns were more pronounced than those observed in the 2022 stock assessment, but much less pronounced than those found in the 2019 assessment (Stewart and Hicks 2019b). To explore whether the changes in scale could be related to estimates of natural mortality, each retrospective estimate of this parameter from the three models estimating it were compared; however, there were no clear trends (Figures 66).

A further retrospective analysis is based on comparing the spawning biomass estimates among the actual stock assessments conducted since 2012. This type of 'across assessment' retrospective looks at the performance of the stock assessment ensemble as new data and model changes have evolved over time and best reflects the changes actually incorporated into management supporting information. The terminal spawning biomass estimated from most of these assessments are nearly identical to the time series from the preliminary 2025 analysis (Figures 67). However, as has been the focus of much discussion in the 2023 and 2024 stock assessments, the terminal estimates from those analyses both showed a downward revision from the previous year (Stewart and Hicks 2024, 2025). Supplementary and bridging model runs in both of those assessments indicated that the commercial fishery data were providing most of the downward trend; when those data were removed the FISS and other information was very consistent with the previous year's results. This could be due to changes in the fishery, loss of information and/or bias in the FISS, or other unmodelled processes.
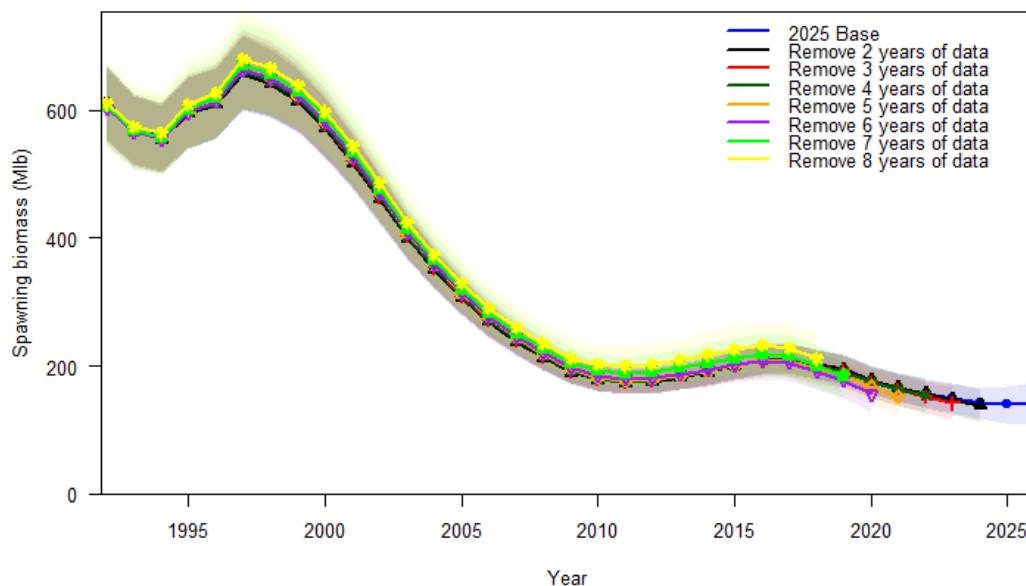
Figure 62. Seven-year retrospective analysis of spawning biomass (there are no data available for 2025 at this time so two years are removed for the first comparison) based on the coastwide short model.
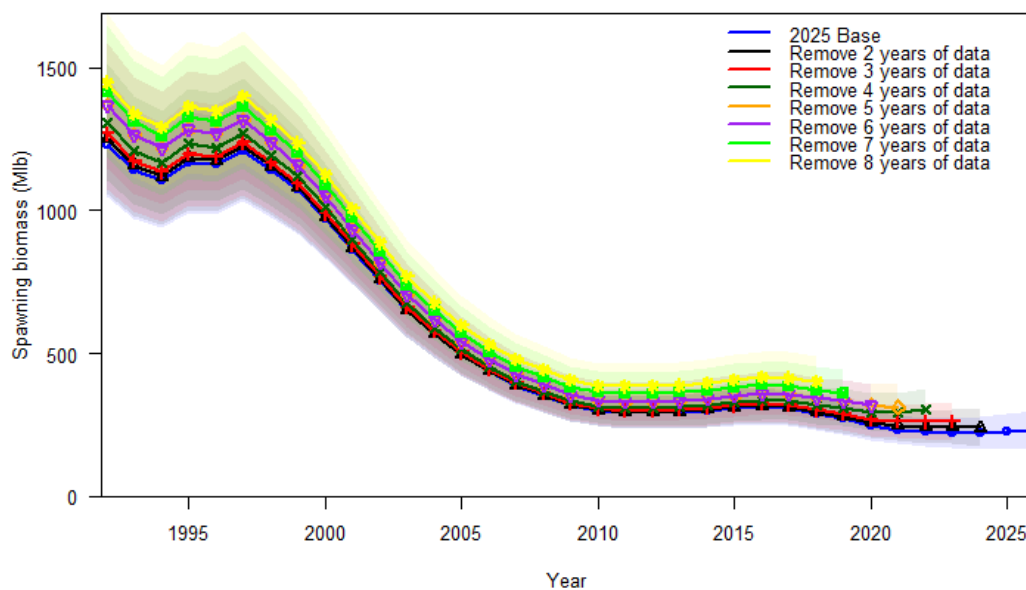

Figure 63. Seven-year retrospective analysis of spawning biomass (there are no data available for 2025 at this time so two years are removed for the first comparison) based on the AAF short model.
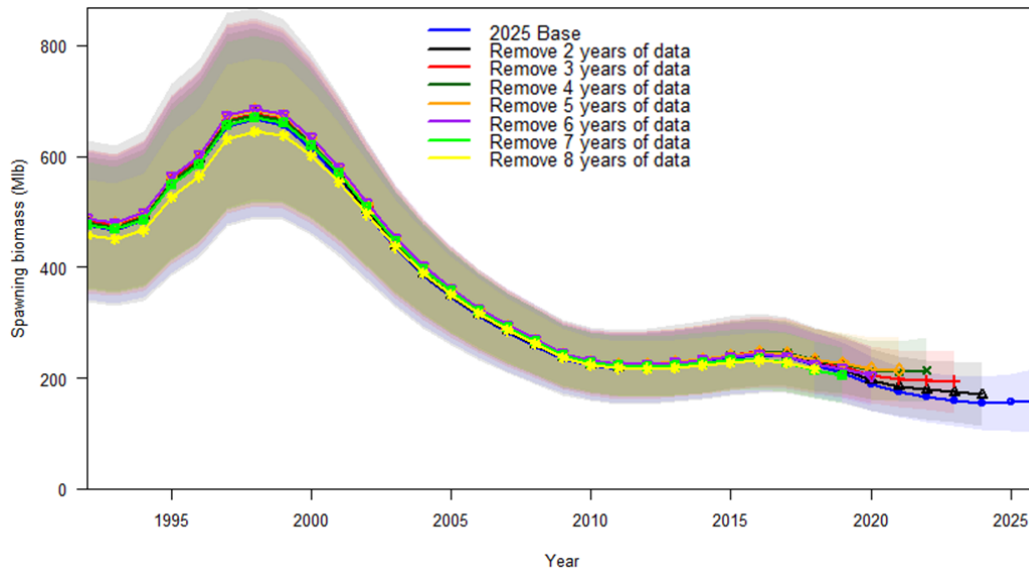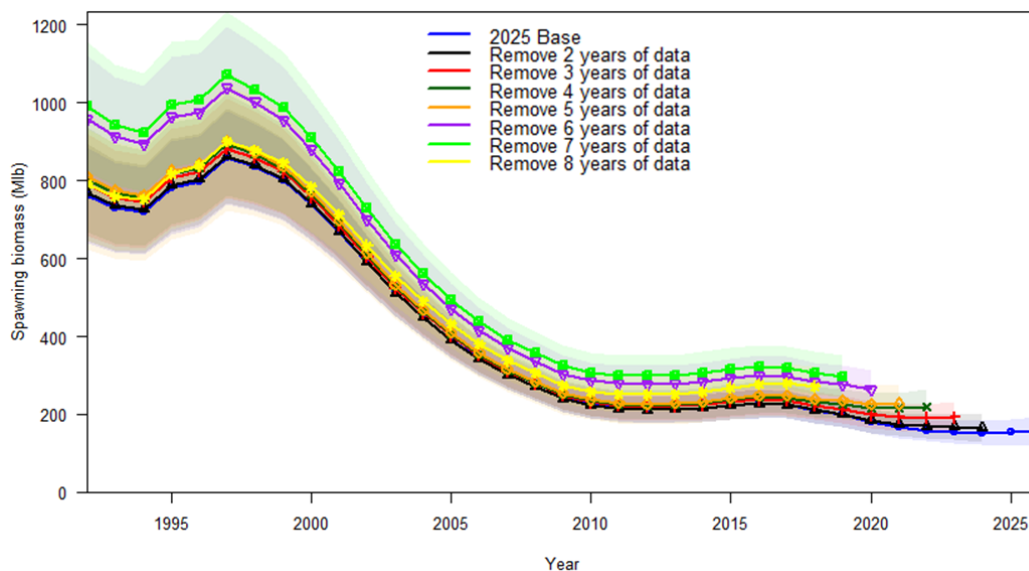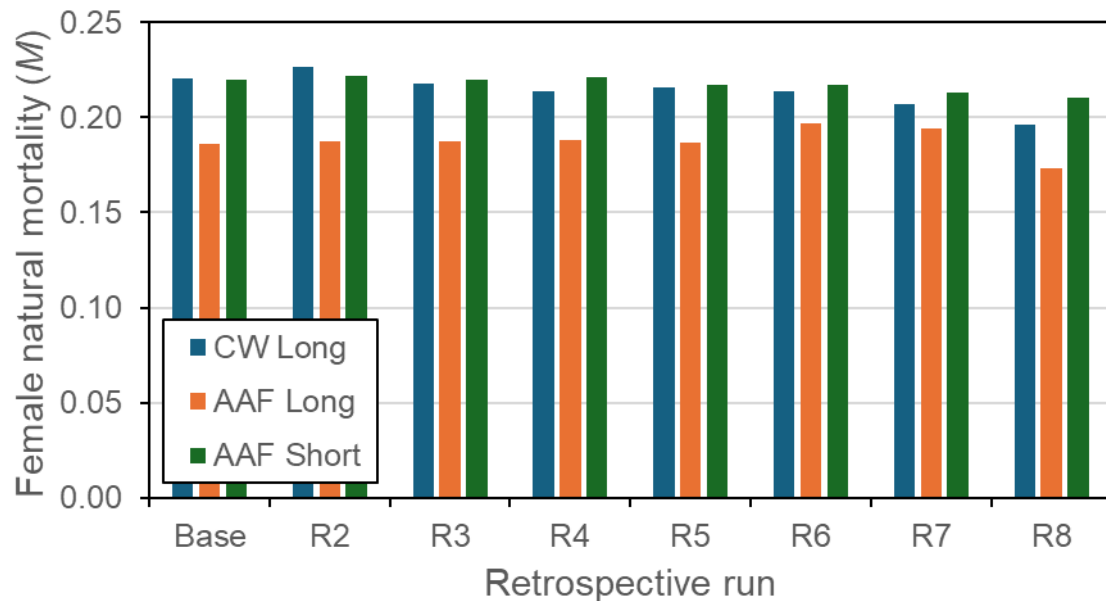
Figure 64. Seven-year retrospective analysis of spawning biomass (there are no data available for 2025 at this time so two years are removed for the first comparison) based on the coastwide long model. Time-series is truncated in 1992 so that differences in the terminal years are more visible.
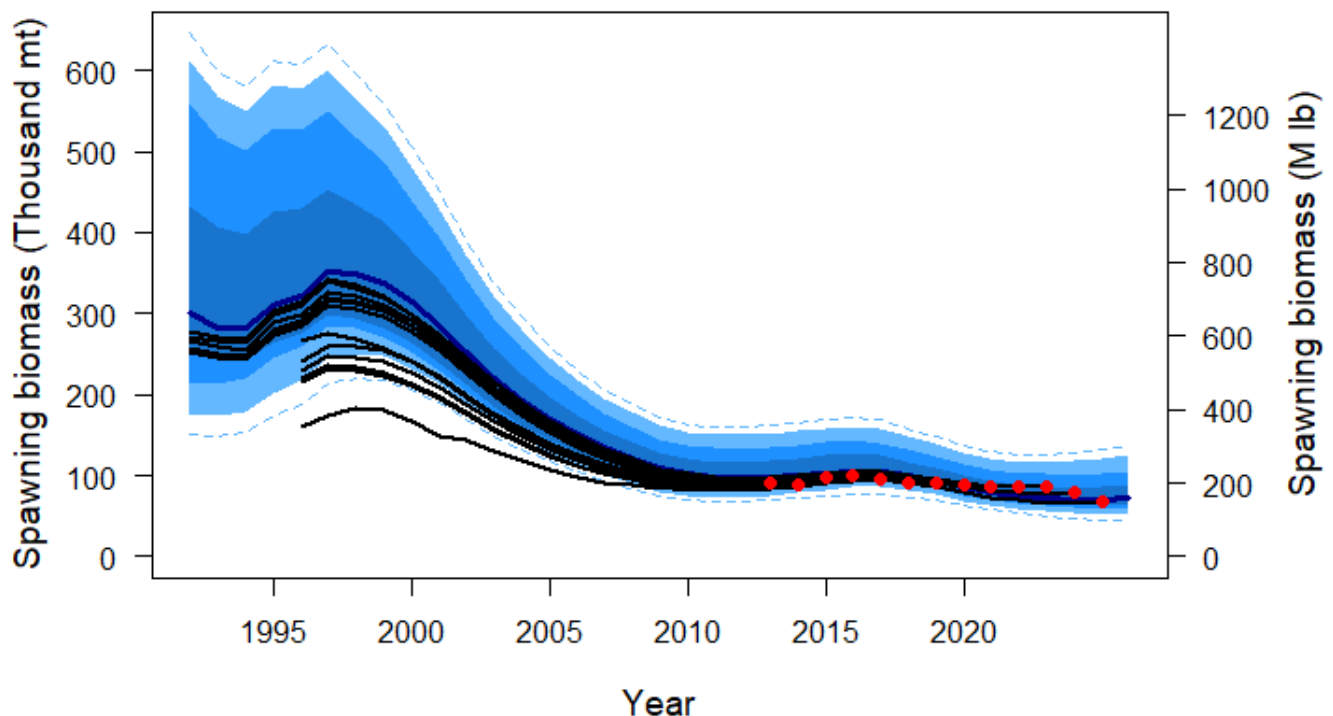

Figure 65. Seven-year retrospective analysis of spawning biomass (there are no data available for 2025 at this time so two years are removed for the first comparison) based on the AAF long model. Time-series is truncated in 1992 so that differences in the terminal years are more visible.

Figures 66. Estimates of female natural mortality (M) over the 7-year retrospective analyses for the three models where this parameter is estimated. The base model includes all data through 2024 (there is no data from 2025 at this preliminary stage); each of the other estimates represent models with two (R2) to seven (R8) years of data removed.

Figures 67. Retrospective analysis of spawning biomass across stock assessments conducted from 2012 to 2024. Red points indicate the terminal estimate from each stock assessment; shaded region indicates the uncertainty around the median ensemble estimate (solid blue line) from the preliminary 2025 stock assessment.

Bayesian analysis

The 2019 stock assessment included a substantial evaluation of Bayesian integration for the short coastwide model (Stewart and Hicks 2019b). This effort did not produce substantially different results from the maximum likelihood and asymptotical variance methods (Fournier et al. 2012) routinely employed. However, there are a number of potential benefits to using an explicitly Bayesian approach, including better characterization of uncertainty (Magnusson et al. 2012) and a more directly interpretable characterization of the probability distributions. There is also the potential for differences in the results of Bayesian analyses due to the right-skewed nature of some distributions for key parameter and management-related quantities in complex fisheries models (Stewart et al. 2013b).

In aggregate, the 2019 results suggested that the asymptotic distributions were a reasonable approximation for the full posterior distributions in these models, and also that the process of regularizing the selectivity parameters and removing some deviations to improve integration did not having an appreciable effect on the solution. This is generally consistent with studies of process error where overparameterizing (adding the capability for variation when it wasn't present) was generally found to be unbiased, and therefore preferable to underparameterizing when temporal variability was present (e.g., Martell and Stewart 2014; Stewart and Monnahan 2017).

Additional Bayesian analysis was not included in this (or the 2022) assessments. However, if a multi-year assessment approach was to become part of a future management procedure for the IPHC more time could be devoted to exploring Bayesian models.

Other uncertainty considerations

There are many important sources of uncertainty not captured in the four models included in this ensemble. These include myriad alternative structural assumptions such as spatially-explicit population dynamics, connection with Russian waters, alternative stock-recruitment functions, time-varying mortality, different data weighting approaches, and many others. There are also several tractable sources of projection uncertainty that are not in the current approach, including uncertainty in projected weight-at-age (although the sensitivity of this was investigated at SRB request in 2016 and found to be low) and uncertainty in the realized mortality associated with limits set by the Commission.

Within the modelled time-series there are also data-related uncertainties that could be addressed via a range of alternative approaches. Uncertainty in the time series of mortality for these models is not currently captured, as they are treated as inputs and assumed to be known without error. In previous assessments, sensitivity analyses have been conducted to the degree of discard mortality in the commercial fishery, potential effects of unobserved whale depredation, as well as to the magnitude of total bycatch mortality. In concept, these types of uncertainties could be explicitly included in the models; however, full estimation of catch in statistical catch-at-age models generally requires other stabilizing assumptions, so direct integration of this uncertainty may still prove challenging. Additional sources of uncertainty and avenues for development are identified in the Research Priorities section below.

The ensemble

Model-integrated quantities are used as the primary stock assessment output for management use, as well as the basis for decision table probabilities (Stewart and Hicks 2025). All quantities of management interest are integrated for the recent time period (1992+), for which all four sets of model results are available. These quantities include: spawning biomass, relative spawning biomass, and the Spawning Potential Ratio (SPR; summarized as fishing intensity, $F_{XX\%}$, where the *XX%* represents *SPR*). Decision table quantities are divided into four categories: stock trend (which is the only set of metrics that are independent of any harvest strategy related assumptions), stock status, fishery trend, and fishery status. Integration is performed for all these quantities using the basic approach outlined below.

Methods

The basic approach to model integration remains unchanged from the 2015 and subsequent analyses. A sample of random draws is created from the output from each of the models included in the ensemble. For the spawning biomass time-series, the estimates and associated standard deviations for female spawning biomass from each of the four models were extracted from the report file. A vector of length n is created for each model (m), where the relative weight (w_m) is simply the relative fraction of the total draws across all models comprised by n_m:

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

This approach allows for easily adjusted weighting of models. Routine reporting of results uses $\sum_m n_m$ for all models equal to twenty million; this has been found to produce negligible Monte-Carlo error even in the tails of extremely skewed distributions, creating robust and stable reporting of all quantities of interest with a smooth distribution. Although this choice could potentially be optimized for each statistic of interest, current integration code (in *R*) does not represent a constraining step in the analysis.

The harvest strategy employs a control rule that reduces the coastwide SPR target linearly from the interim 'reference level' at $SB_{30\%}$ to zero at $SB_{20\%}$. Since the 2019 assessment this calculation uses a dynamic estimate of 'unfished' biomass calculated for each year of the time-series. This calculation replays the entire time-series, without the fishing mortality, assuming the same parameter values (including recruitment deviations) but accounting for the different level of spawning biomass projected for each year and its effect on subsequent expected (pre-deviation) recruitment in each year. Since 2020 the dynamic unfished biomass calculation has been included simultaneously with variance calculations of all model parameters and outputs and (importantly) includes the covariance in the estimated and unfished dynamic spawning biomass in the variance of the IPHC's reference points and other outputs.

Evaluation of weighting based on predictive skill

All Pacific halibut assessments since 2014 have relied on equal weighting of all four models. However, weighting based on several potential approaches has been considered since the 2015 stock assessment (Stewart and Martell 2016). Briefly, these have included:

AIC – but this is known to be highly dependent on data weighting, and can only be applied in cases where the same data sets are being fit by all models under consideration

Strength of retrospective patterns – perhaps relative to a 'null' distribution for a statistic like Mohn's rho (Mohn 1999) based on simulation (Hurtado-Ferro et al. 2015); while helpful to diagnose model performance, it does not necessarily indicate a 'good' model, as evidenced by the fact that a static prediction will have no retrospective pattern at all.

Fit to the FISS index – without an AIC-type correction, there is no penalty for overparameterized models

Expert opinion – this is subjective, and the tendency has been to revert to equal weighting in the absence of strong evidence to the contrary.

Mean Absolute Standardized Error (MASE; Hyndman and Koehler 2006) – a measure of predictive skill.

Most recently in the 2022 full stock assessment, the MASE statistic was extensively evaluated (Stewart and Hicks 2022). After considerable exploration and review the SRB recommended against moving forward with MASE-based model weighing (IPHC 2022). Model weighting has not yet been revised or explored further for 2025.

Preliminary results for 2025

Comparison of the spawning biomass estimates from the four stock assessment models comprising the ensemble shows that the 95% intervals from any single model are substantially narrower than the aggregate (Figure 68). All four models indicate a similar overall trajectory, including the small increase in biomass over 2011-2016 and subsequent decrease as the effects of reduced recruitment subsequent to 2006 (Figure 61; upper panel) graduate through to the spawning biomass. The AAF long model provides the largest estimate of the beginning of year 2026 spawning biomass; however, that distribution still contains the estimates from the other three models (Figure 69).The differing estimates of natural mortality in each of the four models result in recruitment (at age-0) of differing scales; however when divided by the average for each model trends in recruitment are very similar across all models (Figure 70).
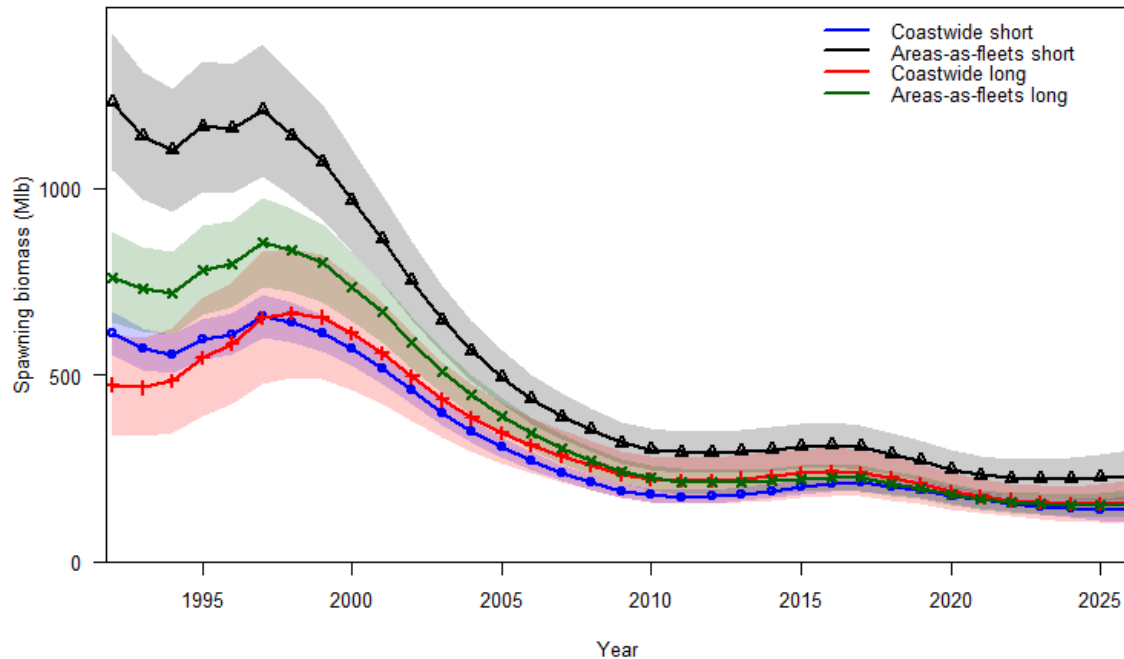
Figure 68. Comparison of spawning biomass time series (shaded regions indicate asymptotic approximations to the 95% confidence interval) from each of the preliminary models contributing to the 2025 preliminary ensemble.
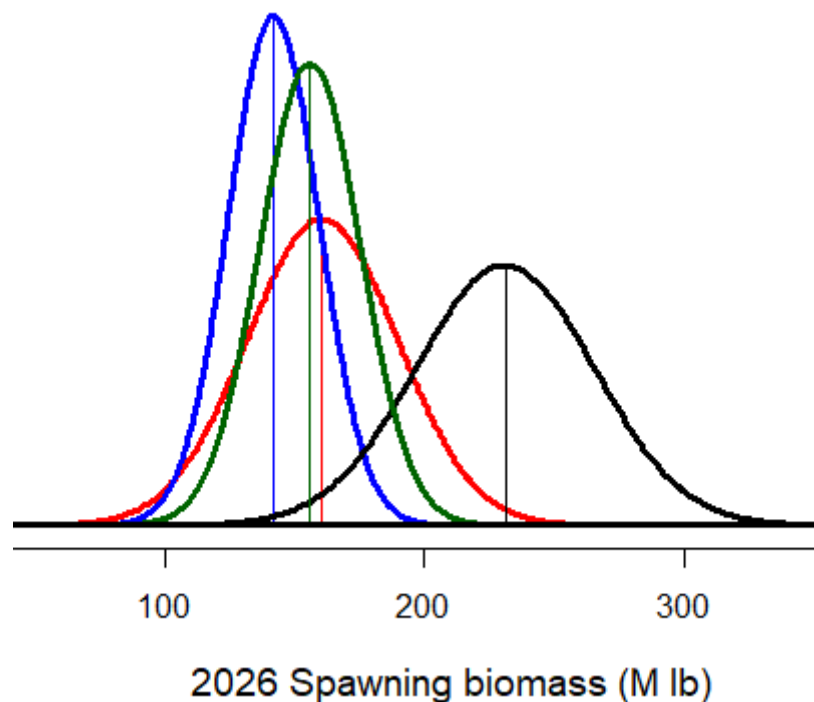

Figure 69. Comparison of terminal (2026) spawning biomass estimates (pdfs) from each of the preliminary models contributing to the 2025 preliminary ensemble. Vertical lines represent the median value from each model.
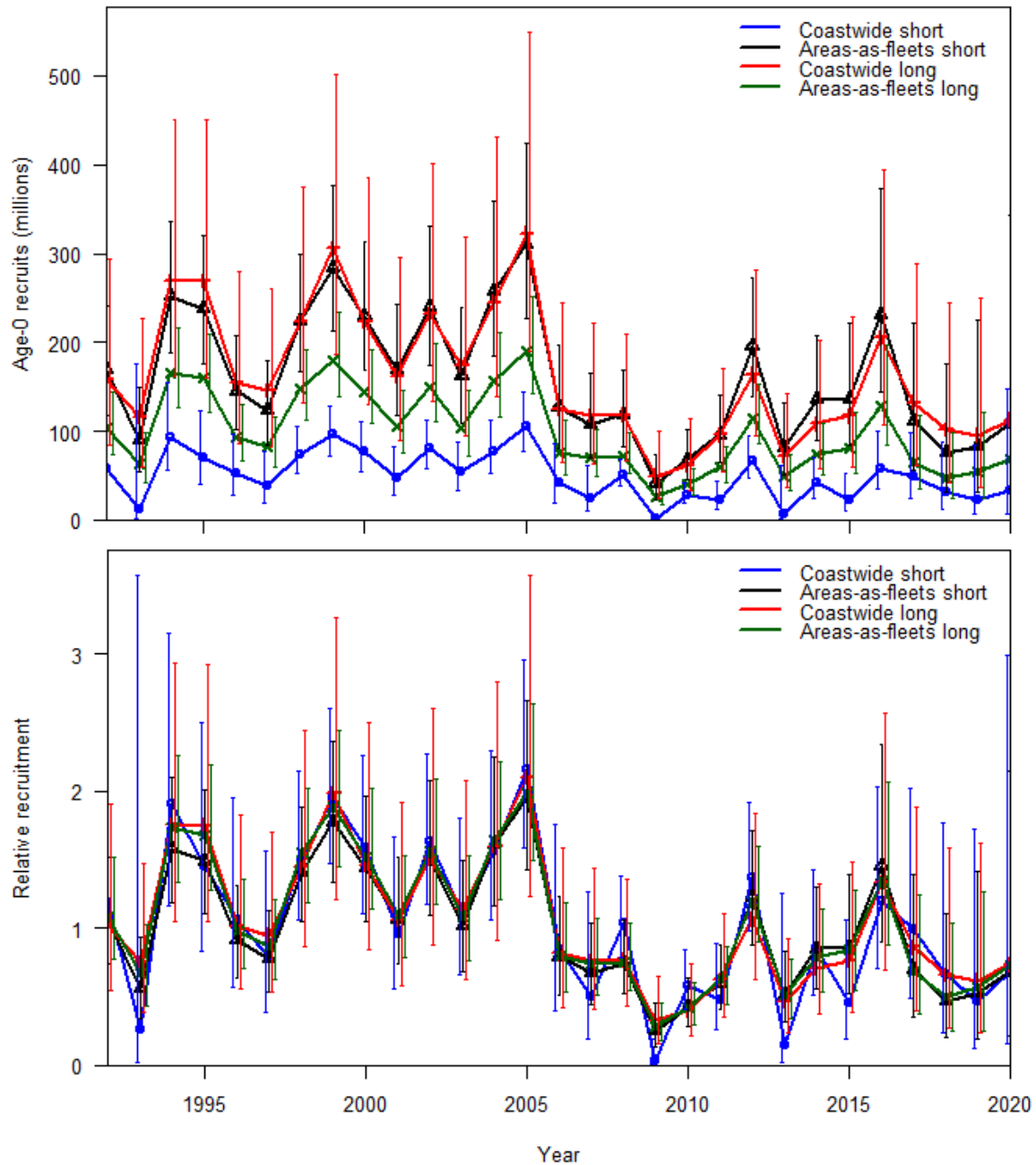
Figure 70. Comparison of recruitment time series (upper panel; vertical lines indicate asymptotic approximations to the 95% confidence interval) and relative recruitment series (each standardized to its mean; lower panel) from each of the preliminary models contributing to the 2025 ensemble.

Future development

Several extensions to this preliminary assessment will be possible for the final 2025 analysis. These include:

- Responses to suggestions and comments generated from SRB026 and SRB027.
- Addition of all 2025 data, extending existing time series (mortality, indices, ages, etc.).
- The sex-ratio of the 2024 commercial fisheries landings based on the IPHC's genetic assay will be available by late summer.

In addition to the list of research priorities (longer list below), there are several potential avenues for development within and among the four models included in the ensemble.

The updated bootstrapping performed for this assessment provides a strong basis for objective interannual and among fleet weighting of age composition data. Both alternative likelihoods, including those already evaluated to some degree for this assessment over the last several years (e.g., the Dirichlet-multinomial, logistic normal) and alternative calculations of composition residuals (e.g., One-Step-Ahead (OSA) residuals; Thygesen et al. 2017; Trijoulet et al. 2023) are strong candidates for further investigation. A considerable effort exploring the properties of OSA residuals was made as part of this 2025 stock assessment, and a draft manuscript has been produced. Incorporation of that approach may be possible for the next Pacific halibut assessment.

Other avenues for development include changes to the ensemble approach itself. The 2019 assessment explored expanding the number of models included in the ensemble to better capture the uncertainty in M that was missed through using a fixed value in the two (at that time) short time-series models. By estimating M for the short AAF model in the 2022 stock assessment, the integration of uncertainty was improved. Upcoming assessments may need to explore whether the fixed value of 0.15 in the coastwide short model is still appropriate given the increasing weight of evidence that M for Pacific halibut is higher.

As ensemble changes are evaluated, both weighting and technical efficiency should be considered. Technical costs of adding additional models to the ensemble include additional time spent running these additional models rather than exploring other sensitivities and identifying clear effects of newly available data during the very short assessment analysis period each fall. Pragmatically, there may be relatively little to be gained from increasing the ensemble in this manner beyond slightly smoother integrated distributions. As the IPHC's management procedure evolves, to potentially include multi-year assessments, there may be additional latitude for increased model and ensemble complexity.

The current ensemble is based on maximum likelihood estimates and asymptotic approximations to the posterior distributions for model parameters and derived quantities. Bayesian posteriors represent a conceptually more appealing basis for probability distributions, and could better capture the full range and potential asymmetries in the distributions for model quantities (Magnusson et al. 2012; Stewart et al. 2013b). Bayesian integration may also allow for statistically correct treatment of variance parameters (such as the sigmas governing

recruitment variability and selectivity or catchability process error), as would use of true random effects methods. Although it would be technically preferable to regularize and run all four assessment models as Bayesian analyses, at present this is technically infeasible given the tight time-line between data availability and the deadline for the annual stock assessment. The analysis time difference between minimization and full posterior integration, even using the most efficient methods available for the coastwide short model (see section above), is still too large. However, if the IPHC were to move to a more formal management procedure and/or to a multi-year mortality limit-setting process, the stock assessment could be conducted at a pace that would allow much greater reliance on Bayesian models.

Research priorities

The development of the IPHC's research priorities has been closely tied to the needs of the stock assessment and harvest strategy policy analyses, such that the IPHC's research projects will provide data, and hopefully knowledge, about key biological and ecosystem processes that can then be incorporated directly into analyses supporting the management of Pacific halibut. Research priorities for the Pacific halibut stock assessment are delineated into three broad categories: improvements in basic biological understanding (including fishery dynamics), investigation of existing data series and collection of new information, and technical development of models and modelling approaches. The highest priority items in each of these categories are highlighted in the 5-year research plan and are expected to be the primary focus of ongoing efforts. However, it is helpful to maintain a longer list of items to inform future prioritization, to create a record of data and research needs, and to foster opportunistic and/or collaborative work on these topics when possible.

Biological understanding and fishery yield

Key areas for improvement in biological understanding include:

- *Highest priority*: Updating the fecundity-weight relationship and the presence and/or rate of skip spawning.

- *Highest priority*: The relative role of potential factors underlying changes in size-at-age is not currently understood. Delineating between competition, density dependence, environmental effects, size-selective fishing and other factors could allow improved prediction of size-at-age under future conditions.

- Movement rates among Biological Regions at the adult, juvenile and larval stages remain uncertain and likely variable over time. Long-term research to inform these rates could lead to a spatially explicit stock assessment model for future inclusion into the ensemble.

- Improved understanding of recruitment processes and larval dynamics could lead to covariates explaining more or the residual variability about the stock-recruit relationship than is currently accounted for via the binary indicator used for the Pacific Decadal Oscillation.

Data related research

This section represents a list of potential projects relating specifically to existing, and new data sources that could benefit the Pacific halibut stock assessment:

- *Highest priority*: Continued collection of sex-ratio from the commercial landings will provide valuable information for determining relative selectivity of males and females, and therefore the scale of the estimated spawning biomass, and the level of fishing intensity as measured by SPR.

- *Highest priority*: Evaluation of the magnitude of marine mammal depredation and tools to reduce it.

- A space-time model could be used to calculate weighted FISS and/or commercial fishery age-composition data. This might alleviate some of the lack of fit to existing data sets that is occurring not because of model misspecification but because of incomplete spatial coverage in the annual FISS sampling which is accounted for in the generation of the index, but not in the standardization of the composition information.

- The work of Monnahan and Stewart (2015) modelling commercial fishery catch rates could be used to provide a standardized fishery index for the recent time-series that would be analogous to the space-time model used for the FISS.

- There is a vast quantity of archived historical data that is currently inaccessible until organized, electronically entered, and formatted into the IPHC's database with appropriate meta-data. Information on historical fishery landings, effort, and age samples would provide a much clearer (and more reproducible) perception of the historical period.

- Additional efforts could be made to reconstruct estimates of subsistence harvest prior to 1991.

- Discard mortality estimates for the IPHC Regulatory Area 2B recreational fishery are currently unavailable, but there is an estimation system in place. Further work to develop these estimates would be preferable to the use of proxy rates from IPHC Regulatory Area 2C.

- NMFS observer data from the directed Pacific halibut fleet in Alaska could be evaluated for use in updating discard mortality rates and the age-distributions for discard mortality. This may be more feasible if observer coverage is increased and if smaller vessels (< 40 feet LOA, 12.2 m) are observed in the future. Post-stratification and investigation of observed vs. unobserved fishing behavior may be required.

- Historical bycatch length frequencies and mortality estimates should be reanalyzed accounting for sampling rates in target fisheries and evaluating data quality over the historical period.

- There are currently no comprehensive variance estimates for the sources of mortality used in the assessment models. In some cases, variance due to sampling and perhaps even non-sampling sources could be quantified and used as inputs to the models via scaling parameters or even alternative models in the ensemble.

Technical development

There are a variety of technical explorations and improvements that could benefit the stock assessment models and ensemble framework. Larger changes (such as entirely new data sets) naturally fit into full assessment analyses; however, incremental changes may be possible during updated assessments when and if new information or methods become available. Specifically, development is intended to occur in time for initial SRB review (generally in June), with primarily only refinements made for final review (October), such that untested approaches are not being implemented during the annual stock assessment itself. Technical research priorities include:

- *Highest priority*: Maintaining consistency and coordination between MSE, and stock assessment data, modelling and methodology.

- *Highest priority*: Exploration of state-space models for Pacific halibut allowing for direct estimation of the variance in time-varying processes.

- *Highest priority*: Continued exploration into the estimation of M in the short coastwide model.

- Continued refinement of the ensemble of models used in the stock assessment. This may include investigation of alternative approaches to modelling selectivity that would reduce relative down-weighting of certain data sources (see section above), evaluation of additional axis of uncertainty (e.g., steepness, as explored above), or others.

- Exploration of methods for better including uncertainty in directed and non-directed discard mortalities in the assessment (now evaluated only via alternative mortality projection tables or model sensitivity tests) in order to better include these sources uncertainty in the decision table. These could include explicit discard/retention relationships, including uncertainty in discard mortality rates, and allow for some uncertainty directly in the magnitude of mortality for these sources.

- Bayesian methods for fully integrating parameter uncertainty may provide improved uncertainty estimates within the models contributing to the assessment, and a more natural approach for combining the individual models in the ensemble (see section above).

- Alternative model structures, including a growth-explicit statistical catch-at-age approach and a spatially explicit approach may provide avenues for future exploration. Efforts to develop these approaches thus far have been challenging due to the technical complexity and data requirements of both. Previous reviews have indicated that such

efforts may be more tractable in the context of operating models for the MSE, where conditioning to historical data may be much more easily achieved than fully fitting an assessment model to all data sources for use in tactical management decision making.

Acknowledgements

References

Adams, G.D., Holsman, K.K., Barbeaux, S.J., Dorn, M.W., Ianelli, J.N., Spies, I., Stewart, I.J., and Punt, A.E. 2022. An ensemble approach to understand predation mortality for groundfish in the Gulf of Alaska. Fisheries Research **251**. doi:10.1016/j.fishres.2022.106303.

Bosley, K.M., Schueller, A.M., Goethel, D.R., Hanselman, D.H., Fenske, K.H., Berger, A.M., Deroba, J.J., and Langseth, B.J. 2021. Finding the perfect mismatch: Evaluating misspecification of population structure within spatially explicit integrated population models. Fish and Fisheries. doi:10.1111/faf.12616.

Clark, W.G. 2003. A model for the world: 80 years of model development and application at the international Pacific halibut commission. Natural Resource Modeling **16**(4): 491-503.

Clark, W.G. 2004a. Statistical distribution of IPHC age readings. IPHC Report of Assessment and Research Activities 2003. p. 99-110.

Clark, W.G. 2004b. Nonparametric estimates of age misclassification from paired readings. Canadian Journal of Fisheries and Aquatic Sciences **61**: 1881-1889.

Clark, W.G., and Hare, S.R. 2002. Effects of climate and stock size on recruitment and growth of Pacific halibut. North American Journal of Fisheries Management **22**: 852-862.

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

Clark, W.G., Hare, S.R., Parma, A.M., Sullivan, P.J., and Trumble, R.J. 1999. Decadal changes in growth and recruitment of Pacific halibut (*Hippoglossus stenolepis*). Canadian Journal of Fisheries and Aquatic Sciences **56**: 242-252.

Cox, S.P., Ianelli, J., and Mangel, M. 2014. IPHC Scientific Review Board Meeting. June 2014., 3 p.

Cox, S.P., Ianelli, J., and Mangel, M. 2017. 4.7 Reports of the IPHC Scientific Review Board (SRB). IPHC Report of Assessment and Research Activities 2016. p. 443-452.

Du, J. 2014. Uncertainty and ensemble forecast. National Weather service. 42 p.

Forsberg, J.E., and Stewart, I.J. 2015. Re-ageing of archived otoliths from the 1920s to the 1990s. IPHC Report of Assessment and Research Activities 2014. p. 405-428.

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

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences **68**: 1124-1138.

Francis, R.I.C.C. 2016. Revisiting data weighting in fisheries stock assessment models. Fish. Res. doi:10.1016/j.fishres.2016.06.006.

Goethel, D.R., and Berger, A.M. 2017. Accounting for spatial complexities in the calculation of biological reference points: effects of misdiagnosing population structure for stock status indicators. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1878-1894. doi:10.1139/cjfas-2016-0290.

Hamel, O.S. 2014. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES Journal of Marine Science **72**(1): 62-69. doi:10.1093/icesjms/fsu131.

Hamel, O.S., and Cope, J.M. 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research **256**. doi:10.1016/j.fishres.2022.106477.

Hamel, O.S., Ianelli, J.N., Maunder, M.N., and Punt, A.E. 2023. Natural mortality: Theory, estimation and application in fishery stock assessment models. Fisheries Research **261**. doi:10.1016/j.fishres.2023.106638.

Hamill, T.M., Brennan, M.J., Brown, B., DeMaria, M., Rappaport, E.N., and Toth, Z. 2012. NOAA's Future Ensemble-Based Hurricane Forecast Products. Bulletin of the American Meteorological Society **93**: 209-220.

Hicks, A., and Stewart, I. 2025. Updates to the IPHC MSE and a review of coastwide management procedures. IPHC-2025-MSAB021-07. 18 p.

Hulson, P.-J.F., and Williams, B.C. 2024. Inclusion of ageing error and growth variability using a bootstrap estimation of age composition and conditional age-at-length input sample size for fisheries stock assessment models. Fisheries Research **270**. doi:10.1016/j.fishres.2023.106894.

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. 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science **72**(1): 99-110. doi:10.1093/icesjms/fsu198.

Hutniczak, B., Tran, H., and Kong, T. 2025. Fisheries data overview. IPHC-2025-AM101-08 Rev_2. 15 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.

Ianelli, J., Holsman, K.K., Punt, A.E., and Aydin, K. 2016. Multi-model inference for incorporating trophic and climate uncertainty into stock assessments. Deep Sea Research Part II: Topical Studies in Oceanography **134**: 379-389. doi:10.1016/j.dsr2.2015.04.002.

Ianelli, J., Fissel, B., Stienessen, S., Honkalehto, T., Siddon, E., and Allen Akselrud, C. 2021. Assessment of the walleye pollock stock in the Eastern Bering Sea. Chapter 1. NPFMC Bering Sea and Aleutian Islands SAFE. 171 p.

IPHC. 2022. Report of the 21st session of the IPHC Scientific review board (SRB021). IPHC-2022-SRB021-R.

IPHC. 2025. Report of the 101st session of the IPHC Annual Meeting (AM101). Vancouver, BC, Canada, 27-31 January 2025. IPHC-2025-AM101-R. 52 p.

Karp, M.A., Blackhart, K., Lynch, P.D., Deroba, J., Hanselman, D., Gertseva, V., Teo, S., Townsend, H., Williams, E., and Yau, A. 2018. Proceedings of the 13th National Stock Assessment Workshop: Model Stability, Model Complexity, & Ensemble Modelling. U.S. Dept. of Commerce, NOAA. NOAA Tech. Memo. NMFS-F/SPO-189. 49 p.

Kong, T., Tran, H., and Prem, C. 2022. Fisheries data overview (2021). IPHC-2022-AM098-06 Rev_1. 15 p.

Li, Y., Bence, J.R., and Brenden, T.O. 2014. An evaluation of alternative assessment approaches for intermixing fish populations: a case study with Great Lakes lake whitefish. ICES Journal of Marine Science **72**(1): 70-81. doi:10.1093/icesjms/fsu057.

Litzow, M.A., Hunsicker, M.E., Bond, N.A., Burke, B.J., Cunningham, C.J., Gosselin, J.L., Norton, E.L., Ward, E.J., and Zador, S.G. 2020. The changing physical and ecological meanings of North Pacific Ocean climate indices. Proceedings of the National Academy of Sciences of the United States of America **117**(14): 7665-7671. doi:10.1073/pnas.1921266117.

Magnusson, A., Punt, A.E., and Hilborn, R. 2012. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fisheries. doi:10.1111/j.1467-2979.2012.00473.x.

Mangel, M., MacCall, A.D., Brodziak, J., Dick, E.J., Forrest, R., Pourzand, R., and Ralston, S. 2013. A perspective on steepness, reference points, and stock assessment. Canadian Journal of Fisheries and Aquatic Sciences **70**(6): 930-940.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.R., and Francis, R.C. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society **78**(6): 1069-1079.

Martell, S., and Stewart, I. 2014. Towards defining good practices for modeling time-varying selectivity. Fisheries Research **158**: 84-95. doi:10.1016/j.fishres.2013.11.001.

Maunder, M.N., Crone, P.R., Valero, J.L., and Semmens, B.X. 2015. Growth: theory, estimation, and application in fishery stock assessment models. CAPAM Workshop Series Report 2. 57 p.

McGilliard, C.R., and Palsson, W. 2021. Assessment of the rex sole stock in the Gulf of Alaska. NPFMC Gulf of Alaska SAFE. Chapter 6. 83 p.

McGilliard, C.R., Punt, A.E., Methot, R.D., Hilborn, R., and Jacobson, L. 2014. Accounting for marine reserves using spatial stock assessments. Canadian Journal of Fisheries and Aquatic Sciences: 1-19. doi:10.1139/cjfas-2013-0364.

Methot Jr, R.D. 2024. Stock Synthesis User Manual Version 3.30.23.1. NOAA Fisheries. Seattle, WA. December 5, 2024. 272 p.

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

Methot Jr, R.D., Wetzel, C.R., Taylor, I.G., Doering, K.L., and Johnson, K.F. 2021b. Stock synthesis user manual version 3.30.17. NOAA Fisheries, Seattle, WA. 238 p.

Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability in estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences **68**: 1744-1760.

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

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

Methot, R.D., Wetzel, C.R., and Taylor, I.G. 2019. Stock Synthesis User Manual Version 3.30.13. NOAA Fisheries. Seattle, WA. 213 p.

Methot, R.D., Wetzel, C.R., Taylor, I.G., Doering, K.L., Perl, E.F., and Johnson, K.F. 2024. Stock Synthesis user manual version 3.30.22.1. NOAA Fisheries, Seattle, Washington. January 31, 2024. 256 p.

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. doi:10.1006/jmsc.1999.0481.

Myers, R.A., Bowen, K.G., and Barrowman, N.J. 1999. Maximum reproductive rate of fish at low population sizes. Canadian Journal of Fisheries and Aquatic Sciences **56**: 2404-2419.

Piner, K.R., and Wischnioski, S.G. 2004. Pacific halibut chronology of bomb radiocarbon in otoliths from 1944 to 1981 and a validation of ageing methods. Journal of Fish Biology **64**: 1060-1071.

Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in Australia. Fisheries Research **168**: 85-99. doi:10.1016/j.fishres.2015.04.002.

Stewart, I., and Hicks, A. 2019a. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2018. IPHC-2019-AM095-09. 26 p.

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

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

Stewart, I., and Webster, R. 2022. Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. IPHC-2022-SA-02. 56 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 Hicks, A. 2024. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2023. IPHC-2024-SA-01. 37 p.

Stewart, I., and Hicks, A. 2025. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2024. IPHC-2025-SA-01. 40 p.

Stewart, I., and Webster, R. 2025. Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. IPHC-2025-SA-02. 57 p.

Stewart, I., Hicks, A., and Hutniczak, B. 2021. Evaluation of directed commercial fishery size limits in 2020. IPHC-2021-AM097-09. 28 p.

Stewart, I.J. 2014. Overview of data sources for the Pacific halibut stock assessment and related analyses. IPHC Report of Assessment and Research Activities 2013. p. 95-168.

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.

Stewart, I.J., and Martell, S.J.D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fisheries Research **158**: 40-49. doi:10.1016/j.fishres.2013.09.012.

Stewart, I.J., and Martell, S.J.D. 2015. Reconciling stock assessment paradigms to better inform fisheries management. ICES Journal of Marine Science **72**(8): 2187-2196. doi:10.1093/icesjms/fsv061.

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.

Stewart, I.J., and Monnahan, C.C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fisheries Research **192**: 126-134. doi:10.1016/j.fishres.2016.06.018.

Stewart, I.J., and Hicks, A.C. 2018. Interannual stability from ensemble modelling. Canadian Journal of Fisheries and Aquatic Sciences **75**: 2109-2113. doi:10.1139/cjfas-2018-0238.

Stewart, I.J., Wallace, J.R., and McGilliard, C.R. 2009. Status of the U.S. yelloweye rockfish resource in 2009. Pacific Fishery Management Council, Portland, OR. 236 p.

Stewart, I.J., Leaman, B.M., Martell, S., and Webster, R.A. 2013a. Assessment of the Pacific halibut stock at the end of 2012. IPHC Report of Assessment and Research Activities 2012. p. 93-186.

Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., and Kupschus, S. 2013b. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fisheries Research **142**(0): 37-46. doi:http://dx.doi.org/10.1016/j.fishres.2012.07.003.

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

Taylor, I.G., Doering, K.L., Johnson, K.F., Wetzel, C.R., and Stewart, I.J. 2021. Beyond visualizing catch-at-age models: Lessons learned from the r4ss package about software to support stock assessments. Fisheries Research **239**. doi:10.1016/j.fishres.2021.105924.

Thompson, G.G., and Lauth, R.R. 2012. Chapter 2: Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands area. *In* NPFMC Bering Sea and Aleutian Islands SAFE. p. 245-544.

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

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2016. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. doi:10.1016/j.fishres.2016.06.005.

Thygesen, U.H., Albertsen, C.M., Berg, C.W., Kristensen, K., and Nielsen, A. 2017. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics **24**(2): 317-339. doi:10.1007/s10651-017-0372-4.

Trijoulet, V., Albertsen, C.M., Kristensen, K., Legault, C.M., Miller, T.J., and Nielsen, A. 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. Fisheries Research **257**. doi:10.1016/j.fishres.2022.106487.

Ualesi, K., Jack, T., Rillera, R., and Coll, K. 2025. IPHC Fishery-Independent Setline Survey (FISS) design and implementation in 2024 IPHC-2025-AM101-09. 9 p.

Waterhouse, L., Sampson, D.B., Maunder, M., and Semmens, B.X. 2014. Using areas-as-fleets selectivity to model spatial fishing: Asymptotic curves are unlikely under equilibrium conditions. Fisheries Research **158**: 15-25. doi:10.1016/j.fishres.2014.01.009.

Webster, R. 2018. Space-time modelling of fishery independent setline survey data. IPHC-2018-AM094-07. 27 p.

Webster, R., and Stewart, I. 2022. Revision of the IPHC length-weight relationship. IPHC-2022-AM098-INF07. 29 p.

Webster, R.A. 2025. Space-time modelling of survey data. IPHC-2025-AM101-10. 9 p.

Webster, R.A., Clark, W.G., Leaman, B.M., and Forsberg, J.E. 2013. Pacific halibut on the move: a renewed understanding of adult migration from a coastwide tagging study. Canadian Journal of Fisheries and Aquatic Sciences **70**(4): 642-653. doi:10.1139/cjfas-2012-0371.

Webster, R.A., Soderlund, E., Dykstra, C.L., and Stewart, I.J. 2020. Monitoring change in a dynamic environment: spatio-temporal modelling of calibrated data from different types of fisheries surveys of Pacific halibut. Canadian Journal of Fisheries and Aquatic Sciences **77**: 1421-1432.

Appendices

Appendix A: Supplementary material

In addition to this document, supplementary material is available electronically, including:

1) Stock synthesis input files for each of the assessment models included in the proposed ensemble: data file, weight-at-age file, control file with model configuration, starter and forecast files with additional settings. Each of these files has been extensively annotated to aid in locating the various sections, as well as identifying which options and features were implemented or are irrelevant for the configuration.

2) Output from each of the stock assessment models: a sub-directory of all plotting and diagnostic output from each model created by the r4ss package (the entire set can be loaded at once via opening the "_SS_output.html" file), and the raw report (text) file from each model. The report file has not been annotated and contains some information not relevant to the Pacific halibut model configurations; content and formats can be determined from the stock synthesis user manual (Methot Jr et al. 2021a) and technical documentation (Methot and Wetzel 2013a).

3) Copies of the primary software documentation including the general modelling approach implemented in stock synthesis (Methot and Wetzel 2013b), the technical documentation (Methot and Wetzel 2013a) and the current user manual (Methot et al. 2024). From these documents, detailed model equations, data configurations, and control settings can be evaluated for the specific features implemented in the models for Pacific halibut.

4) The overview of data sources (Stewart and Webster 2025) and the stock assessment results (Stewart and Hicks 2025) from the 2024 stock assessment.

5) The documentation from the development of the most recent (2022) full stock assessment (Stewart and Hicks 2022).

6) Recent background papers describing the bootstrapping method employed for fishery and FISS age compositions (Hulson and Williams 2024; Stewart and Hamel 2014), the history of the halibut stock assessment (Stewart and Martell 2014), an evaluation of data weighting and process-error considerations (Stewart and Monnahan 2017), the general rationale for the ensemble approach (Stewart and Martell 2015), and the stability properties of ensemble assessments (Stewart and Hicks 2018).

7) A full record of the historical stock assessment documentation from 1978 to the present can be found on the IPHC's web site (https://www.iphc.int/management/science-and-research/stock-assessment). Individual Scientific Review Board reports and presentations (2013-2024) are available through the IPHC's meetings webpage (https://www.iphc.int/iphc-meetings).