INTERNATIONAL PACIFIC HALIBUT COMMISSION

2019 Pacific halibut (*Hippoglossus stenolepis*) stock assessment: Development

PREPARED BY: IPHC SECRETARIAT (I. STEWART, A. HICKS; 23 MAY 2019)

Table of Contents

Summary

This document reports preliminary analyses in development of the 2019 Pacific halibut (*Hippoglossus stenolepis*) stock assessment. It follows the previous full stock assessment conducted in 2015 (Stewart and Martell 2016; Stewart et al. 2016), and subsequent updates to that assessment in 2016 (Stewart and Hicks 2017), 2017 (Stewart and Hicks 2018a), and 2018 (Stewart and Hicks 2019). Following the review of this document in June 2019 (external peer review and SRB014), requested revisions will be considered and presented for final review in October 2019 (SRB015). Updated data sources, including the results of the 2019 Fishery-Independent Setline Survey (FISS), logbook and biological data from the 2019 commercial fishery, and (potentially) sex-ratio information from the 2018 commercial landings-at-age will be included for the final 2019 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 2014b). 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 2014b). One solution to the endless search for a better stock assessment model is to recognize that all models are simple approximations to reality, and that the uncertainty in our analyses can be better captured through the explicit use of multiple models: the ensemble approach. The ensemble approach utilizes multiple models in the estimation of management quantities and therefore adds explicit accounting for structural uncertainty about these quantities (Stewart and Hicks 2018b, Stewart and Martell 2015). This reduces potential for abrupt changes in management quantities as improvements and additional data are added to individual models, and provides a more realistic perception of uncertainty than any single model, and therefore a stronger basis for 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 2018 stock assessment data, models and results (Stewart and Hicks 2019; Stewart and Webster 2019), this analysis is sequentially updated to ‘bridge’ the changes toward

a preliminary assessment for 2019. This bridging analysis included a series of steps for which intermediate results and comparisons are provided. These steps included:

1) updating to the newest stock synthesis software (version 3.30.13; Methot et al. 2019),
2) adding newly available sex-ratio information from the 2017 commercial fishery landings,
3) extending the temporal length of the two short models to include the beginning of the available Fishery Independent Setline Survey (FISS) time series (1993),
4) updating the entire modelled FISS time series to include whale depredation criteria implemented in the survey in 2018, and
5) re-tuning the process and observation error components of these models to achieve internal consistency within each.

As documented in all recent analyses since 2013, a primary source of uncertainty has been the sex ratio of the commercial landings. The newly available data from 2017 allowed for a two-fold effect on this source of uncertainty: first, estimates of relative selectivity of males in the commercial fishery were decoupled from survey observations for the first time, second, this allowed improved overall fit to the various data sources and changes in the internally consistent levels of process and observation error within each model. In aggregate, the results of this preliminary assessment are consistent with those from recent assessments, but suggest a slightly higher absolute level of female spawning biomass, as well as a higher level of fishing intensity. Spawning biomass trends remain similar to recent analyses with large declines estimated from the late 1990s through around 2010, a brief period of stability and then gradual declines estimated since 2016. The uncertainty in stock dynamics also remains similar and high relative to that frequently reported for single-model or simple stock assessment analyses. This uncertainty will continue to be captured via the annual decision table, reporting the trade-offs between yield and various stock and fishery risks (i.e., Stewart and Hicks 2019).

Sensitivity and retrospective analyses were performed on all models contributing to the ensemble. Individual models showed differing sensitivity to specific important sources including the estimation of the steepness parameter, alternative values of female natural mortality (in the short models utilizing a fixed value of 0.15), and data weighting. Retrospective analyses suggested that these models are sensitive to new information, particularly the sex ratio information from 2017.

Given the challenges and uncertainties of the Pacific halibut population dynamics and stock assessment it is unlikely that some new future assessment model will provide substantially more precise and stable results. In light of the uncertainty and variability within which the Pacific halibut management occurs, the current effort to create and refine a robust management procedure through the IPHC's Management Strategy Evaluation (MSE) process (Hicks and Stewart 2019) may provide a much better prospect for future management success and stability than annual decisions based on annual 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 2018 stock assessment (Stewart and Webster 2019). Where specific improvements to existing data sources have been included in this assessment (i.e., sex-ratios from the 2017 commercial landings and the revised modelled survey time-series) changes are described below.

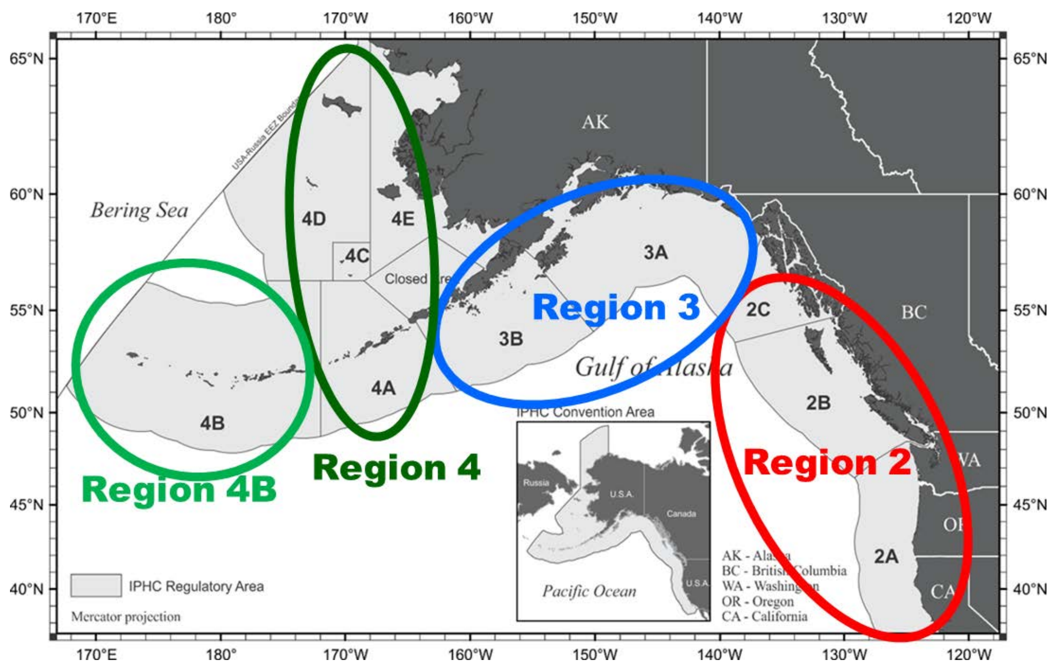

FIGURE 1. IPHC Regulatory Areas, Biological Regions, and the Pacific halibut geographical range within the territorial waters of Canada and the United States of America.

Overview of existing data

The time-series' of Pacific halibut data (Stewart and Webster 2019) provides a rich historical record including mortality estimates, abundance indices (CPUE) and age-composition data that extend back to the late 1800s and early 1900s (FIGURE 2). The IPHC's Fishery Independent Setline Survey (FISS; Erikson et al. 2019) 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, on average several years prior to entry into the retained catch. In aggregate, 42% of the FISS catch comprises smaller fish below the IPHC's 32 inch (82 cm) minimum size limit. The FISS also provides identification of the sex of each fish sampled. Commercial data is sampled at the point of landing (Erikson 2019), so it does not contain biological or catch-rate information on younger, smaller fish below the IPHC's 32 inch (82 cm) minimum size limit (Stewart and Hicks 2018b). Annual mortality estimates are provided to the IPHC from a variety of sources (Erikson 2019) including the directed halibut fisheries (commercial, recreational and subsistence) as well as incidental mortality associated with discards in directed fisheries and bycatch mortality in fisheries 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.

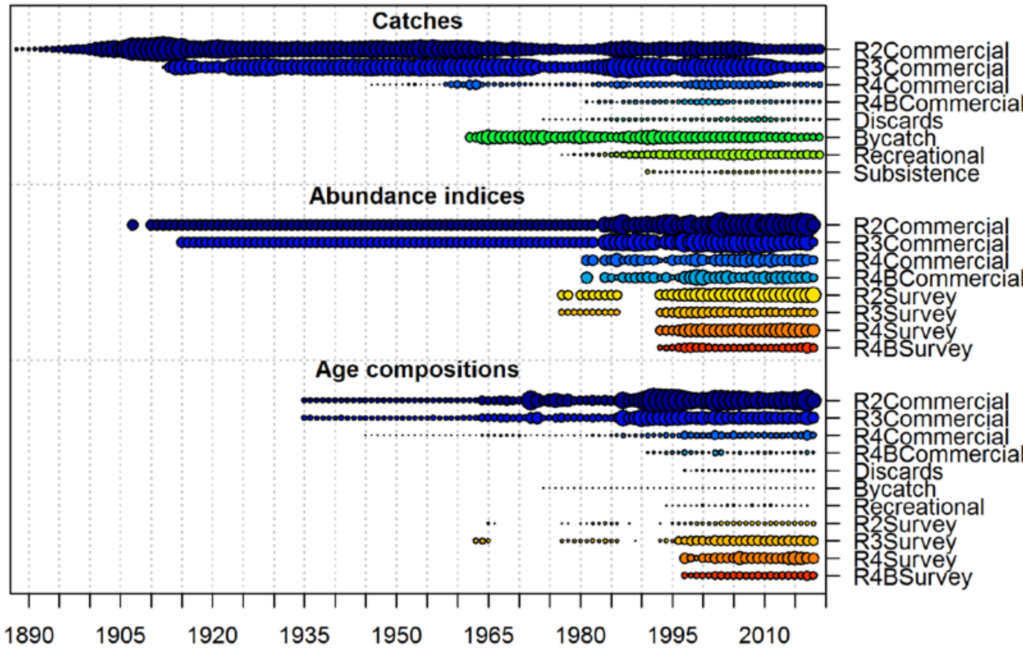

FIGURE 2. Data used in the stock assessment. Circle size is proportional to the magnitude of mortality (catches), inversely proportional to the variance (abundance indices) or proportional to the sample size (age-composition data).

Mortality

The industrial Pacific halibut fishery developed first off the west coast of the United States and Canada and sequentially moved to the north, only reaching full exploitation across all spatial areas in the last several decades (FIGURE 3). Mortality from bycatch in non-Pacific halibut directed fisheries 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.

Index data

The IPHC's FISS (Erikson et al. 2019) comprises the primary index of recent abundance and the primary source of biological data for use in the stock assessment. Index values from 1993-present (TABLE 1) are reported and used in this assessment in numbers of halibut captured per unit effort (NPUE). The time-series is based on the output of the IPHC's space-time model (Webster 2019) which estimates the degree of spatial and temporal correlation among survey stations in order to predict trends in biomass and abundance across the entire range of Pacific halibut within the IPHC Convention Area. For the recent time period (1993-2018) this index provides relatively precise trend information by IPHC Regulatory Area. Estimates from the space-time model are weighted by the relative spatial bottom area in each IPHC Regulatory Area, when combined up to Biological Regions and coastwide indices. 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 (Stewart and Martell 2016). 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.

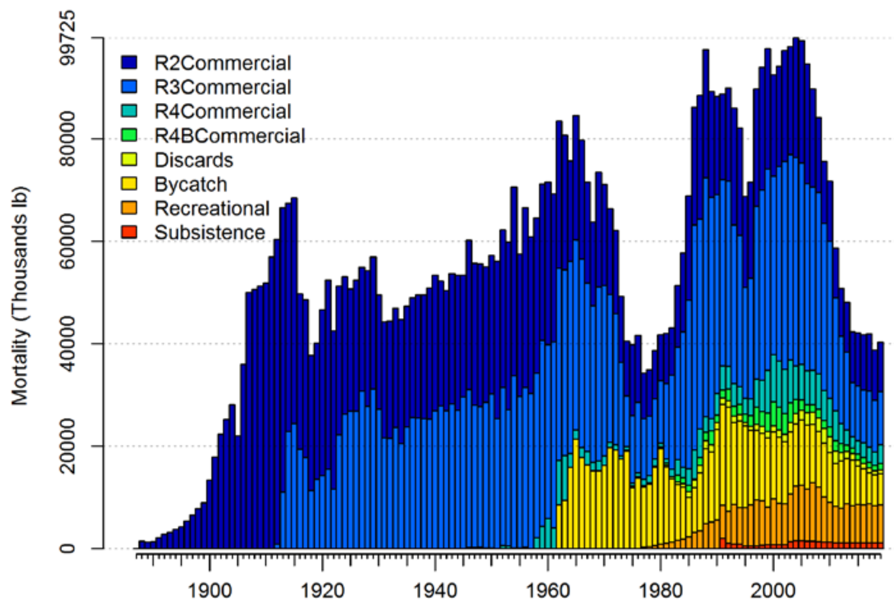

FIGURE 3. Time-series of mortality estimates used in the stock assessment. Commercial series is partitioned by Biological Regions, as in the Areas-As-Fleets models.

Commercial fishery landed Catch-Per-Unit-Effort (CPUE, generally referred to as Weight-Per-Unit-Effort or WPUE) is reported through mandatory logbooks (voluntary for vessels under 26 feet, 7.9 m, in length) collected by IPHC port samplers, and returned directly to the IPHC by mail. Commercial CPUE is available as far back as the early 1900s (Stewart and Webster 2019) 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) 1993-2018, raw average observed NPUE 1977-1986 and estimated log(SE); assumed values in italics.

	Region 2		Region 3		Region 4		Region 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)
1977	0.60	*0.124*	2.00	*0.246*	--	--	--	--	1.47	*0.322*
1978	0.80	*0.124*	1.30	*0.246*	--	--	--	--	1.11	*0.322*
1979	--	--	1.90	*0.246*	--	--	--	--	--	--
1980	1.20	*0.124*	2.50	*0.246*	--	--	--	--	2.01	*0.322*
1981	0.80	*0.124*	3.80	*0.246*	--	--	--	--	2.67	*0.322*
1982	1.85	*0.124*	3.80	*0.246*	--	--	--	--	2.88	*0.322*
1983	2.31	*0.124*	3.40	*0.246*	--	--	--	--	2.88	*0.322*
1984	6.75	*0.124*	11.60	*0.246*	--	--	--	--	9.31	*0.322*
1985	5.66	*0.124*	11.90	*0.246*	--	--	--	--	8.95	*0.322*
1986	4.55	*0.124*	7.80	*0.246*	--	--	--	--	6.26	*0.322*
1993	6.36	0.106	25.33	0.146	2.04	0.125	9.99	0.338	7.73	0.101
1994	7.66	0.112	25.14	0.124	2.12	0.116	10.22	0.268	7.96	0.083
1995	9.18	0.087	26.87	0.116	2.10	0.114	10.55	0.240	8.55	0.077
1996	8.14	0.075	27.44	0.096	2.36	0.096	10.76	0.187	8.66	0.064
1997	7.55	0.069	29.71	0.099	2.55	0.064	11.01	0.113	9.14	0.066
1998	6.35	0.071	25.21	0.091	2.70	0.062	11.13	0.112	8.15	0.059
1999	5.25	0.062	24.59	0.093	2.34	0.066	9.52	0.128	7.56	0.062
2000	5.79	0.062	26.66	0.093	2.49	0.061	8.70	0.141	8.11	0.063
2001	6.70	0.063	23.63	0.102	2.31	0.063	6.81	0.169	7.45	0.066
2002	6.71	0.059	26.25	0.106	2.19	0.063	4.95	0.202	7.81	0.072
2003	5.73	0.062	25.85	0.101	2.05	0.065	4.10	0.212	7.45	0.071
2004	5.26	0.060	29.17	0.107	2.02	0.066	3.81	0.207	8.00	0.079
2005	5.79	0.059	25.02	0.127	2.05	0.068	3.68	0.212	7.27	0.088
2006	5.70	0.056	23.82	0.126	2.09	0.059	4.29	0.194	7.08	0.085
2007	6.34	0.058	25.64	0.125	2.06	0.065	5.44	0.190	7.58	0.085
2008	6.32	0.058	23.31	0.124	2.41	0.067	5.25	0.179	7.31	0.080
2009	6.37	0.052	21.85	0.129	2.41	0.064	4.44	0.198	6.99	0.081
2010	6.35	0.053	22.10	0.129	2.28	0.060	4.18	0.209	6.94	0.083
2011	6.33	0.052	22.88	0.137	2.17	0.061	4.19	0.197	7.04	0.090
2012	7.42	0.051	23.75	0.143	2.21	0.053	3.77	0.201	7.38	0.092
2013	7.14	0.050	18.29	0.147	2.00	0.052	5.09	0.153	6.19	0.088
2014	7.38	0.049	21.64	0.147	2.07	0.050	4.54	0.153	6.91	0.093
2015	8.18	0.047	21.64	0.141	2.06	0.054	4.56	0.163	7.03	0.087
2016	8.26	0.048	21.54	0.137	1.87	0.057	5.02	0.137	6.93	0.086
2017	5.94	0.047	15.66	0.153	1.71	0.063	4.00	0.094	5.26	0.091
2018	5.04	0.045	14.65	0.160	1.59	0.066	4.03	0.146	4.85	0.097

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

	Region 2		Region 3		Region 4		Region 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	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 estimated log(SE); assumed values in italics.

	Region 2		Region 3		Region 4		Region 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	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	60.88	*0.100*	61.95	*0.100*	--	--	--	--	61.00	*0.100*
1975	61.85	*0.100*	66.76	*0.100*	--	--	--	--	61.00	*0.100*
1976	44.37	*0.100*	61.91	*0.100*	--	--	--	--	55.00	*0.100*
1977	64.14	*0.100*	65.57	*0.100*	--	--	--	--	63.00	*0.100*
1978	54.05	*0.100*	68.47	*0.100*	--	--	--	--	71.00	*0.100*
1979	55.84	*0.100*	67.33	*0.100*	--	--	--	--	75.00	*0.100*
1980	59.56	*0.100*	116.09	*0.100*	--	--	--	--	94.00	*0.100*
1981	73.95	*0.100*	148.86	*0.100*	137.27	*0.100*	99.00	0.078	111.00	*0.100*
1982	71.95	*0.100*	181.34	*0.100*	97.82	*0.100*	--	--	127.00	*0.100*
1984	152.14	0.045	491.33	0.046	350.30	*0.100*	161.00	0.103	291.00	*0.100*
1985	161.87	0.051	535.06	0.039	441.49	0.103	234.00	0.160	351.00	0.034
1986	137.49	0.035	506.00	0.042	325.84	0.059	238.00	0.372	315.00	0.041
1987	135.71	0.027	490.38	0.036	353.58	0.162	220.00	0.111	316.00	0.038
1988	168.60	0.054	560.55	0.042	405.68	0.105	224.00	0.122	363.00	0.036
1989	155.08	0.042	507.69	0.031	379.25	0.080	268.00	0.094	353.00	0.025
1990	194.77	0.043	403.54	0.036	362.91	0.097	209.00	0.103	315.00	0.029
1991	170.73	0.039	375.02	0.041	365.84	0.157	329.00	0.085	314.00	0.038

TABLE 4. Commercial fishery Weight-Per-Unit-Effort (WPUE) 1992-2018 and estimated log(SE).

	Region 2		Region 3		Region 4		Region 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)
1992	167.74	0.040	413.39	0.048	324.01	0.117	280.00	0.095	315.00	0.035
1993	200.10	0.031	439.11	0.096	400.28	0.447	218.00	0.220	369.00	0.100
1994	175.72	0.027	362.77	0.049	343.14	0.333	197.00	0.101	302.00	0.069
1995	190.75	0.025	439.48	0.043	330.22	0.100	189.00	0.336	326.00	0.037
1996	208.83	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	432.94	0.103	275.00	0.064	400.00	0.025
1998	221.23	0.029	512.59	0.036	433.49	0.084	287.00	0.058	403.00	0.025
1999	249.46	0.079	475.49	0.024	406.86	0.058	310.00	0.045	390.00	0.023
2000	229.96	0.036	494.83	0.026	415.81	0.082	320.00	0.048	399.00	0.023
2001	202.80	0.039	454.52	0.029	365.44	0.212	270.00	0.076	358.00	0.042
2002	214.84	0.032	466.46	0.025	303.90	0.080	245.00	0.081	356.00	0.020
2003	208.97	0.018	439.27	0.024	254.79	0.071	196.00	0.068	325.00	0.018
2004	192.93	0.028	425.79	0.026	242.57	0.070	202.00	0.061	315.00	0.019
2005	178.99	0.024	387.69	0.023	219.59	0.063	238.00	0.093	293.00	0.017
2006	180.18	0.024	360.70	0.022	174.18	0.066	218.00	0.111	267.00	0.019
2007	158.09	0.023	344.27	0.026	150.17	0.057	230.00	0.108	249.00	0.020
2008	138.82	0.020	318.17	0.024	162.55	0.071	193.00	0.069	229.00	0.017
2009	152.91	0.020	277.22	0.020	175.25	0.054	189.00	0.097	220.00	0.018
2010	185.13	0.037	242.32	0.024	141.52	0.081	142.00	0.063	202.00	0.020
2011	179.87	0.020	226.65	0.025	141.21	0.057	165.00	0.103	196.00	0.015
2012	193.90	0.020	213.46	0.032	136.03	0.081	149.00	0.066	193.00	0.021
2013	192.72	0.026	189.98	0.033	117.39	0.075	127.00	0.064	178.00	0.017
2014	210.33	0.026	182.93	0.039	108.29	0.098	146.00	0.070	183.00	0.022
2015	217.26	0.024	224.46	0.045	132.77	0.066	149.00	0.076	202.00	0.025
2016	212.58	0.019	216.22	0.044	126.67	0.067	123.00	0.083	196.00	0.020
2017	213.73	0.020	219.60	0.037	116.34	0.087	120.00	0.082	202.00	0.020
2018	204.55	*0.055*	191.36	*0.134*	104.87	*0.135*	133.00	*0.148*	180.00	*0.061*

Age data

Otoliths are sampled randomly from all stations FISS catches at variable rates across IPHC Regulatory areas, with a target of 1500 per Area. 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. 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 survey age data (1963-2018).

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	--	147	--	--	161
1995	103	120	--	--	223
1996	200	424	--	--	624
1997	212	429	221	74	936
1998	228	507	100	42	877
1999	332	556	61	82	1031
2000	242	553	153	83	1031
2001	334	522	148	83	1087
2002	313	558	154	82	1107
2003	323	518	153	82	1076
2004	330	527	148	71	1076
2005	342	509	152	83	1086
2006	321	529	243	84	1177
2007	330	540	181	74	1125
2008	339	552	184	76	1151
2009	336	559	179	84	1158
2010	336	533	182	78	1129
2011	365	554	172	79	1170
2012	361	524	174	72	1131
2013	368	537	170	80	1155
2014	386	567	247	77	1277
2015	365	540	248	82	1235
2016	352	549	230	78	1209
2017	374	527	175	124	1200
2018	467	538	168	77	1250

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 (1983-2018).

Year	Region 2	Region 3	Region 4	Region 4B	Coastwide
1983	133	106	23	--	268
1984	170	90	9	--	282
1985	171	99	14	--	286
1986	158	152	34	--	345
1987	531	498	76	--	1117
1988	278	258	19	--	571
1989	318	371	39	--	752
1990	491	560	50	--	1104
1991	718	496	62	12	1288
1992	1027	478	61	20	1586
1993	959	471	65	11	1506
1994	896	474	89	31	1490
1995	887	468	72	37	1464
1996	859	437	76	27	1399
1997	676	429	183	58	1346
1998	515	277	127	47	966
1999	454	303	118	24	899
2000	512	358	119	27	1016
2001	505	233	117	13	868
2002	561	284	163	53	1061
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	325	152	100	15	592
2018	319	164	100	16	599

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

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 schedule
3) fecundity information
4) weight-at-age
5) length-weight relationship
6) ageing error (bias and imprecision)
7) data based ‘priors’ on bycatch, discard, and recreational selectivity

There have been no significant changes to the treatment of these sources of information since the 2015 stock assessment (Stewart and Martell 2016), and they are updated (where appropriate) and described each year in the annual overview of data sources (Stewart and Webster 2019). For convenience, the treatment of each is briefly summarized in TABLE 8.

Sex-ratio of the commercial landings

A major source of uncertainty in the IPHC’s historical datasets is the sex-ratio of the commercial landings. Because Pacific halibut are legally required to be dressed at sea, port samplers are unable to easily determine the sex of fish at the time of landing. The sensitivity of the stock assessment to the relative selectivity of male and female Pacific halibut has been highlighted since the 2013 analysis (Stewart and Martell 2014a). Through consultation with the Scientific Review Board (SRB), several pilot studies were conducted to explore having fishermen identify the sex and voluntarily mark individual fish at sea (McCarthy 2015). The IPHC ultimately opted to use a genetic test that could be conducted in a cost efficient manner using tissue samples (Drinan et al. 2018). Beginning in 2017, fin clips were collected from all Pacific halibut sampled for length, weight, and age from the commercial fishery landings (Erikson and Kong 2018).

These data are available for this preliminary 2019 stock assessment, and were 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. Although not yet published, the data suggest a very high fraction of the commercial landings are female Pacific halibut (82% coastwide), with Biological Regions ranging from 65% female in Region 4B to 92% female in Region 4 (FIGURE 4). The differences among Biological Regions are most pronounced for ages-13 and greater (FIGURE 5). The effects of these new data on the stock assessment results are discussed as part of the bridging analysis described below.

TABLE 8. Summary of other information sources contributing directly to stock assessment input files (Stewart and Webster 2019).

Input	Summary	Key assumptions
Pacific Decadal Oscillation index	Monthly values (http://jisao.washington.edu/pdo/) averaged and compiled into a binary index for each year based on assignment to 'positive' and 'negative' phases	Only used as a binary indicator rather than annually varying values.
Maturity	Trimmed logistic from Clark and Hare (2006); 50% female maturity at 11.6 years old.	Based on visual assessments, treated as age-based and time-invariant.
Fecundity	Assumed to be proportional to body weight.	Temporal variability only via changes in weight-at-age.
Weight-at-age	Reconstructed from survey and fishery information by Biological Region.	Temporal 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.	Relationship has been shown to differ over space and time (Webster and Erikson 2017)and so may not provide an accurate translation from numbers to weight in some circumstances.
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.

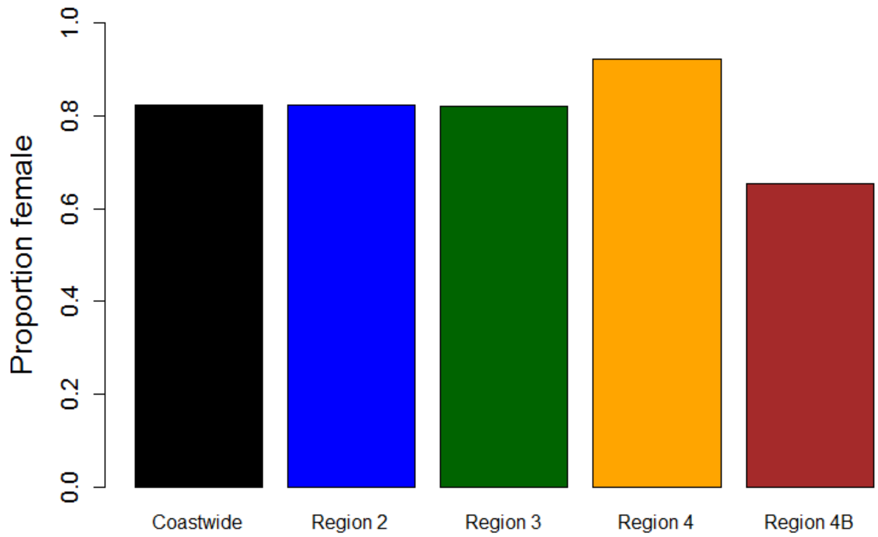

FIGURE 4. Estimates of the proportion female of the commercial landings (numbers of fish) by Biological Region.

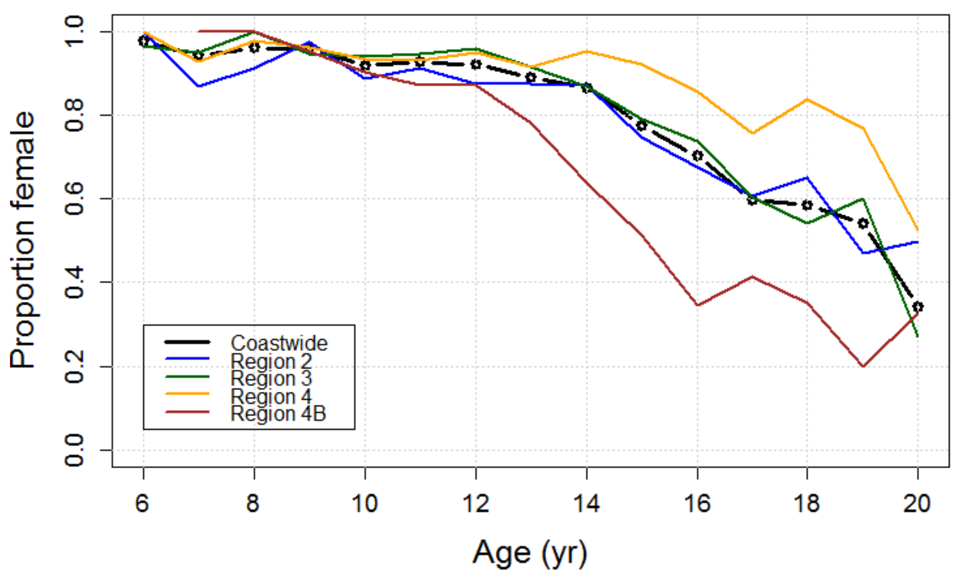

FIGURE 5. Estimates of the proportion of the commercial landings (numbers of fish) by Biological Region and age; data are aggregated at below age-six and above age-20 due to small sample sizes.

Revised Fishery-Independent Setline Survey (FISS) time-series

In 2017, the IPHC Secretariat reviewed historical criteria for determining when a FISS station had experienced whale depredation. Concerns that low levels of depredation and/or cryptic indications of whale activity on the gear might lead to unidentified depredation and therefore negatively biased catch rates led to a revision of the criteria for the 2018 FISS sampling season (Erikson et al. 2019). In order to retroactively apply these criteria to the historical time-series of FISS sampling (Soderlund et al. 2012), specifically including 1993-2017 (the years that are currently included in the space-time model), original field logs and other information had to be retrieved from the IPHC archives and inspected record-by-record. This effort was completed in February, 2019 and provided for this preliminary stock assessment analysis.

The annual station-by-station results, including type of whale interaction and station assignment are publicly available via the IPHCs interactive website (https://www.iphc.int/data/fiss-performance). Briefly, there were only a few geographical areas where enough stations were retroactively assigned as 'ineffective' to make an appreciable change to the modelled time-series. These were largely located in IPHC Regulatory Area 4A, and did not effect the 2018 estimate, because the revised criteria had already been applied to the 2018 data. In IPHC Regulatory Area 4A the variance increased slightly, and the index between 2004 and 2017 increased slightly due to removal of negatively biased catch-rates associated with now identified whale depredation (FIGURE 6). At the scale of Biological Regions and coastwide there was little change in the time-series estimates, and only a very small increase in the variance (FIGURE 7). The effects of these data on the stock assessment results are discussed as part of the bridging analysis described below.

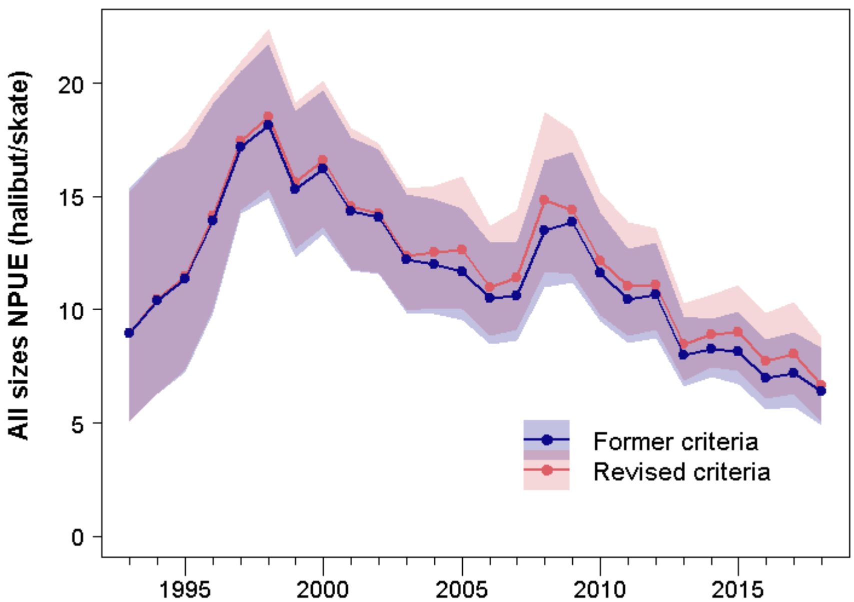

FIGURE 6. Comparison of modelled survey time series for Regulatory Area 4A with the old (former) and new (revised) whale depredation criteria applied to determine station effectiveness.

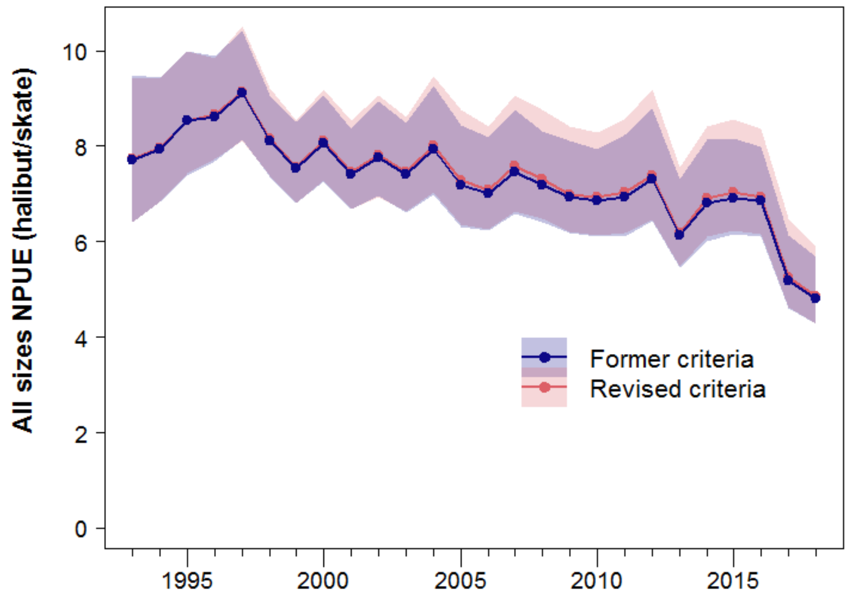

FIGURE 7. Comparison of modelled survey coastwide time series with the old (former) and new (revised) whale depredation criteria applied to determine station effectiveness.

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 2014b). 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 lack-of-fit metrics (Stewart and Martell 2014b). 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. Areas-As-Fleets (AAF) models was developed for the most recent full assessment analysis and review in 2015 (Stewart and Martell 2016). 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” 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 2018c), 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 2019). More detail on how the models are integrated can be found in the Ensemble section below.

Structural rationale

Consistent with the analyses from 2015-17, 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 (Methot and Wetzel 2013a; Methot et al. 2019), and the inputs and output formats are standardized, 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 in R (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 benefit of allowing direct fitting to observed length observations, as well as 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).

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 minimum size limits 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. Most 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 that are unknown, and almost certainly depend on the true underlying processes. 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 (Goethel and Berger 2017). Unfortunately, in the case of 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, commonly the quality of those data 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 recently developed geostatistical model since 1993 (Webster 2018). 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, variable and much less comprehensive survey coverage, lack of access to raw historical fishery data (ages, catch rates, etc.), and many others. The costs of using only a relatively short time-series include a lack of integration between harvest strategy calculations derived from full historical period, a lack of perspective on recent trends, the need for careful treatment of initial model conditions, 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 a 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 over a reasonable range of sex-ratios). 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. However, for Pacific halibut, highly dimorphic growth interacting with 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 skewed largely toward females. Historical modelling suggesting lower natural mortality for males and changing size-at-age all lead to the potential for a static assumption regarding sex-ratio to lead to a highly biased interpretation of stock status unless females and males are 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 the final risk-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 by r4ss (see Appendix A).

These models were configured to make use of relatively standard 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 estimation of fleet- and year-specific fishing mortality rate parameters, and should reasonably approximate the dynamics unless fishing mortality rates are extremely high. Catches were input in thousands of pounds (net weight; head-off and gutted, approximately 75% of round weight), so that the weight-at-age inputs were in pounds and the numbers-at-age tracked in thousands of individuals. Population dynamics contain ages 0-30, and female and male halibut are tracked separately in the dynamics.

The input data were partitioned via a fleet structure of: the directed fishery (by area in the AAF models), discards from the directed fishery, bycatch, recreational, subsistence, and survey (by area in the AAF models). TABLE 9 summarizes the data and key features of each model. Age data were partitioned by sex (the vectors for each year contain females, then males), 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 representing proportions less than 0.1%. This choice avoid 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.

TABLE 9. Comparison of structural assumptions among models.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Modelled period[1]	1992-2020	1888-2020	1992-2020	1888-2020
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	Fixed at 0.15	Estimated
Male M	Estimated	Estimated	Estimated	Estimated
Stock-recruit relationship	B-H	B-H	B-H	B-H
Initial conditions estimated	R_1, *N*-at-age: 1-19	R_0, *N*-at-age: 1-29	R_1, *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	0.55	0.75	0.5
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	Estimated	Estimated
Bycatch selectivity	Domed	Asymptotic	Domed	Domed
Recreational selectivity	Domed	Domed	Domed	Domed
Discard selectivity	Domed, by sex	Domed, by sex	Domed, by sex	Domed, by sex
Personal use selectivity	Mirrored to recreational	Mirrored to recreational	Mirrored to recreational	Mirrored to recreational

[1]Mortality estimates for 2020 were projected based on adopted IPHC limits.

All growth specifications in the control file 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 static vector of maturity-at-age (Stewart and Webster 2019).

For all estimated parameters (except temporal deviations), 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. TABLE 10 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 the two long time-series models. 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 (females from males), such that time-varying selectivity for one sex can be mapped into variability for both sexes without estimating a second set of 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).

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, except in a few cases where there was no improvement in model fit by allowing temporal variability. 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, 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 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 along with other changes for 2019.

TABLE 10. Comparison of estimated parameter counts among models.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Static				
Female *M*	--	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	1	4	4
Survey catchability	1	4	--	4
Fishery selectivity	5	5	20	18
Discard selectivity	6	7	5	6
Bycatch selectivity	4	2	4	3
Recreational selectivity	4	3	3	4
Survey selectivity	5	5	21	18
Total static	29	31	60	61
Time-varying				
Recruitment deviations[1]	51	165	51	165
Fishery catchability deviations	--	108	52	212
Fishery selectivity deviations	76	166	208	532
Survey selectivity deviations	75	84	182	236
Total deviations	202	523	493	1,145
Total	231	554	553	1,206

[1]Includes initial age structure and four forecast years (the latter only included here such that counts will match model output).

In all models, fit to the age data used a multinomial likelihood with initial input sample sizes representing the number of fishery trips or survey stations contributing to that observation, subsequently adjusted down via a multiplicative scalar for each fleet in the control file (more discussion below). Indices of abundance from both the setline survey and commercial fishery (by area in the AAF models) were fit using a log-normal likelihood and input log(*SE*)s. Survey indices were fit in numbers of fish to avoid converting numbers to weights in the data and then weights back to numbers in the model predictions (as recommended by the Scientific Review Board in 2014). Weight-per-unit-effort is the native scale for the fishery indices.

As developed for the 2015 assessment, discard mortality, bycatch and recreational selectivity are estimated, but the age composition data are downweighted to avoid imparting any significant information on recruitment strengths from these uncertain and likely non-representative data sets. 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 to be the case, and there has been little relationship between the magnitude of discards and the magnitude of commercial landings when this has been evaluated for previous reviews. 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 no appreciable effect on the model results but still allowed for the direct estimation of selectivity. This approach propagates uncertainty in the estimated selectivity, and lends itself to direct inclusion of observer data on discarded halibut when it becomes available.

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. 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). Recent work has shown that under some circumstances both components can be iteratively (or by other more statistically rigorous means) estimated simultaneously (Thorson 2018; 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 number of survey sets and fishery trips (and not the number of individual fish measured, which would be much larger), were considerably larger than commonly applied weighting for stock assessment models would suggest (TABLE 5-7). These values were iteratively reduced based on evaluation of three considerations: the relative magnitude of the standardized residuals, comparison of the input value for each fleet with the harmonic mean effective sample size which is an unbiased estimator for a set of independent multinomial samples (Stewart and Hamel 2014), and the scaling suggested by the Francis (2011) method (as implemented in the r4ss package). 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 those derived from the number of trips or stations represented in the data.

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^2_{devs}) 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):

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

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 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, and an improvement on the similar but simpler approach employed in the 2015 stock assessment.

As a final model-building step, models were regularized via adjusting parameterizations through removing and/or fixing 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 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 automated calculation for this is automated in the r4ss package and does not require external calculations. The output of that calculation 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, except where the central tendency of the stock-recruitment relationship was changed (e.g., sensitivity analyses estimating steepness).

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 2014b). Tuning diagnostics and results specific to each model are provided below.

In order to provide for direct transparent comparisons from this preliminary stock assessment through the final results for 2019, the initial step in this analysis was to extend the modelled time-series to 2020, using the projected mortality associated with the limits set by the IPHC for 2019 (IPHC 2019). Weight-at-age was assumed to remain constant from 2018 to 2019 (it will be updated when new data become available) and no other information was needed for this single year projection.

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.

The coastwide short model employed a Beverton-Holt stock recruitment relationship (a change from 2015, as described below) 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 (see sensitivity analyses). 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 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 10.

Age-based selectivity for female halibut in both the setline survey 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 no deviation was estimated for the terminal year (2019), because there were no data yet in the model. This means that the actual mortality in 2019 may have a different effect on the projections when updated from projections and removed via an informed selectivity schedule in the final assessment. 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. In the coastwide short model, with fixed female natural mortality and direct overlap between all years of fishery and survey age data, the male offset parameters for the fishery have been estimated in recent assessments. These parameters have been informed by the weak information on sex-ratio included the sex-aggregated age data. In aggregate, there were five estimated base parameters each for the survey and fishery and annual deviations on the ascending limb parameters (TABLE 10).

As in the 2015 assessment, the scale of male selectivity for both the survey and fishery were allowed to vary over time as a random walk. With only sex-aggregated commercial fishery age compositions (except in 2017; see below), it is not clear how strongly the temporal variability in the scale of male selectivity is informed (and potentially how correlated it would be with female natural mortality, which is fixed in this model). However, the addition of time-varying deviations on the scale parameters was found to improve the residual patterns in the fit to the fishery age-data, and did not show signs of erratic estimation over sensitivity and alternative model runs.

Coastwide long

Initial conditions for the coastwide long time-series model were represented simply as the equilibrium stock condition, as the model period began well before (1888) the first age data were available (1935), and therefore there was a substantial 'burn in' for recruitment variability. 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 2002a, 2006), the coastwide long model allowed for the possibility that recruitment variability is correlated with the regimes of the Pacific Decadal Oscillation (PDO). 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 was converted to a binary indicator (PDO_{regime}) where productive regimes (e.g., 1977-2006) were assigned a value of 1.0, and poor regimes 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:

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

Although the specific parameterization changed in the newest version of stock synthesis (Methot et al. 2019), it was possible to configure the control file to achieve an algebraically identical approach to that used since 2015. This parameterization allows for the β parameter to be estimated at a value of 0.0 if there is no correlation between the putative environmental index and underlying mean recruitment; 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, except that the scale of male selectivity was only estimable after adding the sex-ratio information from 2017 (see below) and was highly unstable when those data were removed (consistent with the 2015 assessment results). Therefore, no attempt was made at present to allow this parameter to vary over time. 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).

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 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, although this could be investigated in future modelling efforts.

The second difference between the short time-series models was in the treatment of the scale of male selectivity for the fishing fleets in each of the four areas. Similar to the coastwide long model, the three parameters defining the male offset to female selectivity for the commercial fishery in each area were only estimable after the addition of 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 Regions 2 and 3.

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.

Convergence criteria

Over the past four years, hundreds of alternative model runs for these four model configurations have been evaluated for evidence of lack of convergence. Tools employed have included monitoring of the maximum gradient component, alternative phasing and initial values (including the use of overdispersed starting points – ‘jitter analyses’) as well as likelihood profiles, and a limited amount of Bayesian integration (see section below).

For this preliminary 2019 assessment, all individual models all had a maximum gradient component < 0.003. A series of preliminary and intermediate runs did not indicate any signs that the estimates reported here represented a local minima, nor did the models have difficulty converging and producing a positive definite Hessian matrix under the broad range of alternative and sensitivity analyses (some presented in this document, but many used only for development).

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 (where there were a high fraction of males in 2017 and presumably throughout the time series) was estimated to be at the bound of 1.0, 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.

Changes from 2018

In the intervening period between the last full stock assessment analysis and review (conducted in 2015) and this preliminary analysis for 2019 a number of important data sources have been changed or added. These changes have been documented and their effects evaluated singly in each year (Stewart and Hicks 2018a, 2019; Stewart and Hicks 2017); however, the cumulative effects on data weighting, parameter estimability, and the tuning of process error in selectivity and recruitment variation has not been fully evaluated. Key changes to the data sources since the full assessment in 2015 include:

- A 44% increase in the number of years of FISS index observations from 1997-2014 to 1993-2018, including the addition of newly collected data and the extension of the time series to include 1993+ in 2017.
- FISS expansions in 2015-2018 supplementing historical gaps in sampling with an effect on both the time-series values and uncertainty.
- Addition of age data from non-standard FISS stations not previously included (2017).
- Design- to space-time model-based survey time series, with changes in the values and uncertainty (generally reduced).
- Use of measured commercial fishery individual fish weights instead of predictions from the length-weight relationship L-W predictions beginning with the 2015 data.

These changes, in tandem with the specific changes described below, result in changes to estimates for a number of model parameters, and the relative tuning of sample sizes and process error variances. These results are described sequentially below, via the ‘bridging’ analysis.

Software version update

Prior to 2019, this stock assessment has used stock synthesis version 3.24 (Methot 2015; Methot and Wetzel 2013b). For 2019, all of the features used in the Pacific halibut stock assessment models have been implemented in stock synthesis version 3.30.13 (Methot et al. 2019). Although some options have been reparameterized (e.g., the treatment of initial model conditions relative to the stock-recruitment curve), in all cases near perfect back-compatibility was retained. The estimated spawning biomass time series and uncertainty intervals for the coastwide and AAF short models were essentially unchanged after updating all of the input files and parameterizations (FIGURE 8). The two long time-series models differed slightly, mainly in the initial conditions, likely as a function of recoding those calculations in the newest version (FIGURE 9). The results from the updated software version were separated from the rest of the bridging analysis to more easily identify these minor differences; all subsequent comparisons were made using the version 3.30 results.

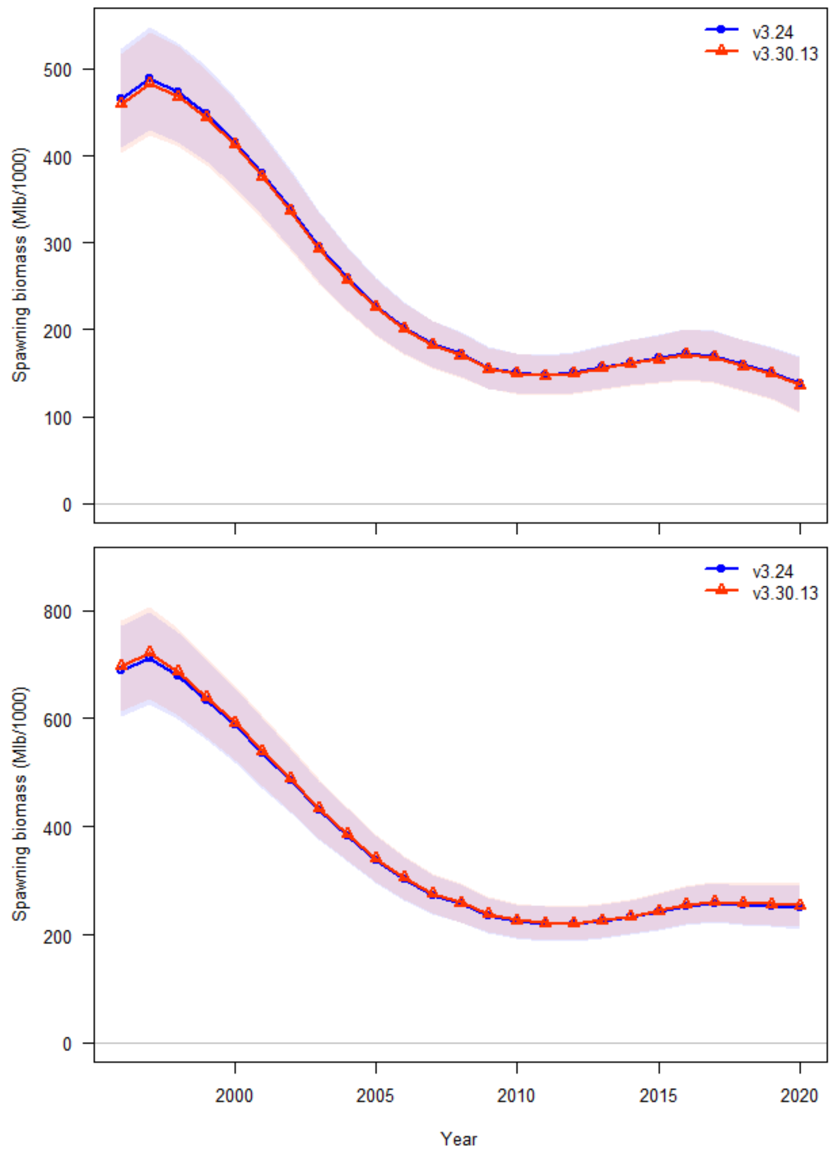

FIGURE 8. Comparison of estimated biomass time series for the coastwide (upper panel) and Areas-as-fleets (lower panel) short models before and after updating to the newest version of stock synthesis.

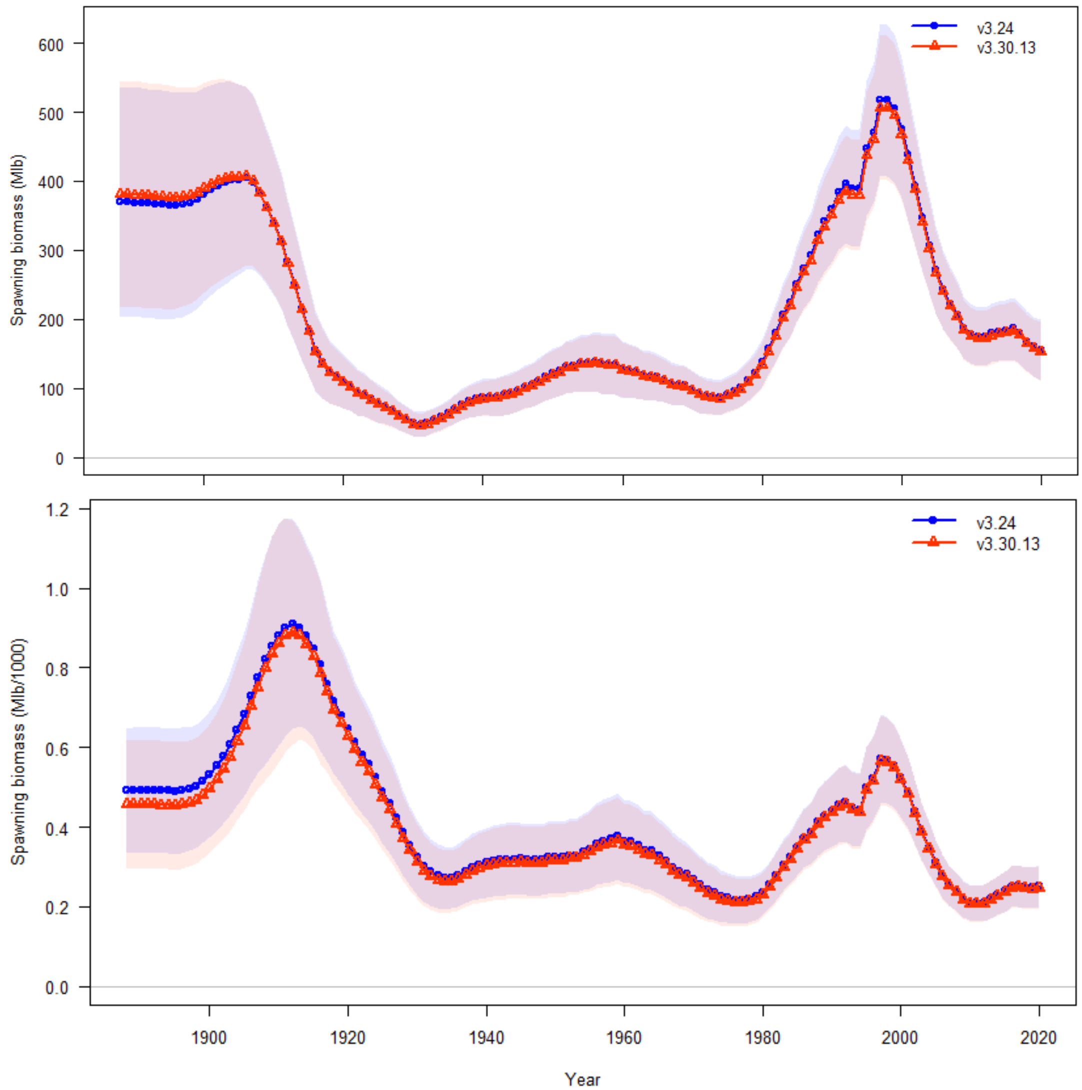

FIGURE 9. Comparison of estimated biomass time series for the coastwide (upper panel) and Areas-as-fleets (lower panel) long models before and after updating to the newest version of stock synthesis.

Updated data sources

There were four steps taken to update from the 2018 stock assessment (implemented in the newest version of stock synthesis) to the preliminary results for 2019:

1) Add the newly available sex-ratio data from the 2017 commercial fishery landings and estimate male selectivity scale parameters.
2) Extend the time series (for the two short models) from 1996 to 1992 and add a stock-recruitment function to these models.
3) Replace the modelled FISS time-series with the series corrected for whale depredation.
4) Regularize and tune each model to be reliable and internally consistent given all the changes that had been made.

The results of each of these steps is reported sequentially for each of the four stock assessment models.

Adding the sex-ratio data to the coastwide short model had no appreciable effect on the trend, but changed the scale slightly, estimating a somewhat larger spawning biomass throughout the modelled period (FIGURE 10). Extending the time-series to include the entire time series of available modelled FISS data and adding a stock-recruitment relationship also increased the spawning biomass estimates slightly and steepened the downward trend over the last several years. The new data also substantially increased the level of recruitment estimated for 1995 and 1994. The modelled FISS time-series including stricter criteria for whale depredation had no visible effect on the results of the short coastwide model. Regularizing and tuning the final configuration including all of the new data also had very little effect on the results.

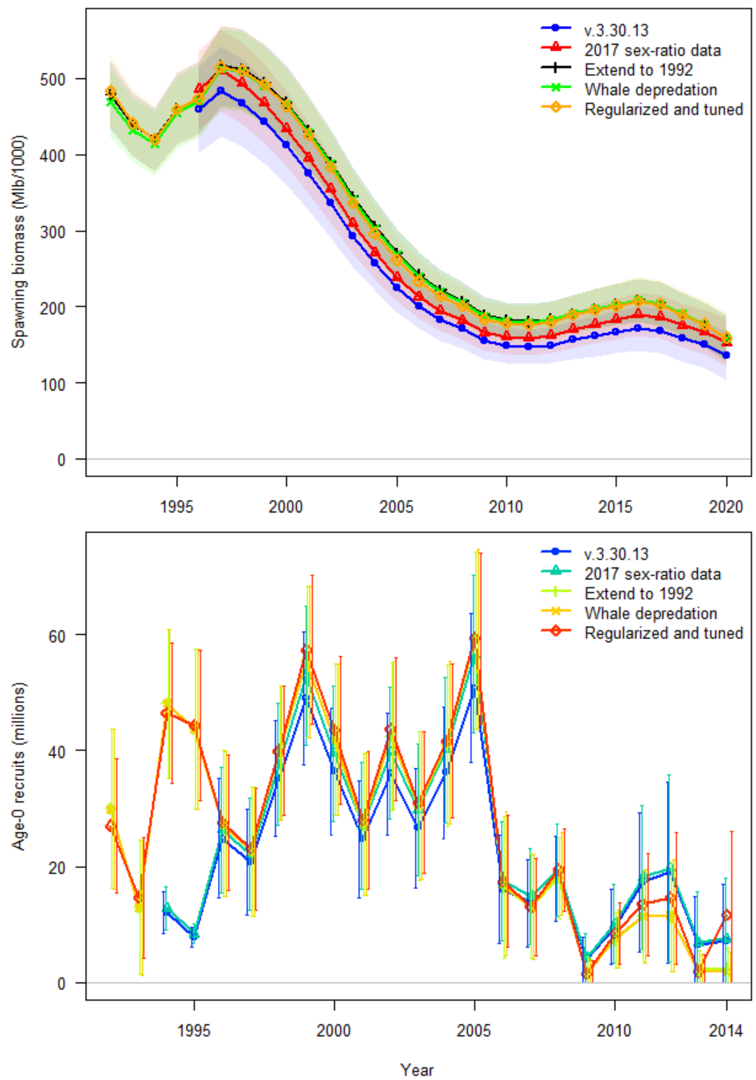

FIGURE 10. Comparison of estimated biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2018 to preliminary 2019 coastwide short models.

Adding the sex-ratio data to the coastwide long model and estimating the scale of the male selectivity (rather than assigning the value estimated for the survey as was done in previous assessments) had little effect on the recent spawning biomass trend (FIGURE 11). However, it did increase the scale of the estimated spawning biomass over most of the time-series, as it suggested fewer male halibut in the commercial landings than in the survey (and therefore previously assumed). The modelled FISS time-series including stricter criteria for whale depredation again had no visible effect on the results of the long coastwide model. Regularizing and tuning the final configuration including all of the new data also increased the scale of the spawning biomass at the end of the time series, and had small but variable effects on the rest of the results.

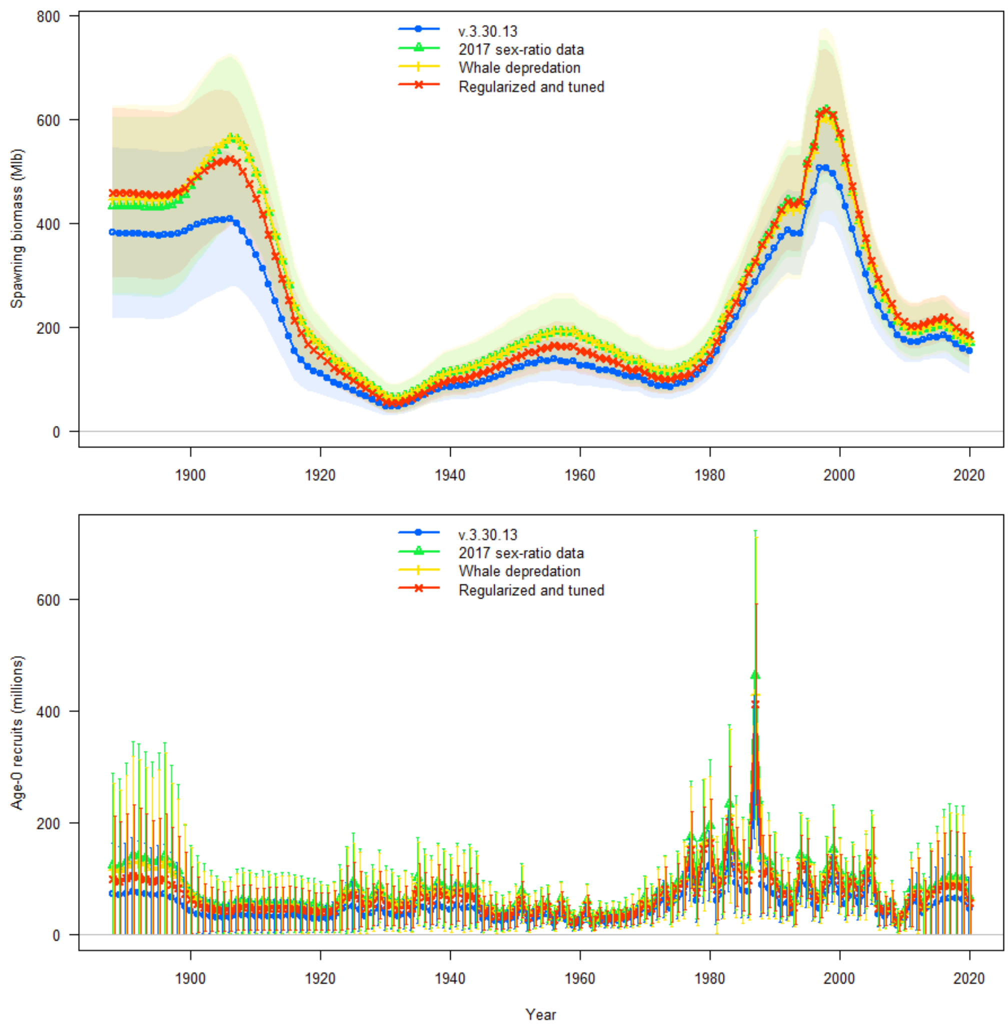

FIGURE 11. Comparison of estimated biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2018 to preliminary 2019 coastwide long models.

Adding the sex-ratio data to the AAF short model and estimating the scale of the male selectivity for each Region (rather than assigning the values estimated for the survey in each Region) again increased the scale of the estimated spawning biomass substantially, suggesting fewer male halibut in the commercial landings than in the survey (FIGURE 12). Extending the time-series increased the scale of the spawning biomass estimates at the end of the modelled period, and adjusted upward the 1994-1995 year-class strengths. As in the other models, the modelled FISS time-series including stricter criteria for whale depredation again had no visible effect on the results of the short AAF model. Regularizing and tuning the final configuration including all of the new data produced a noticeably flatter trend at the end of the modelled period.

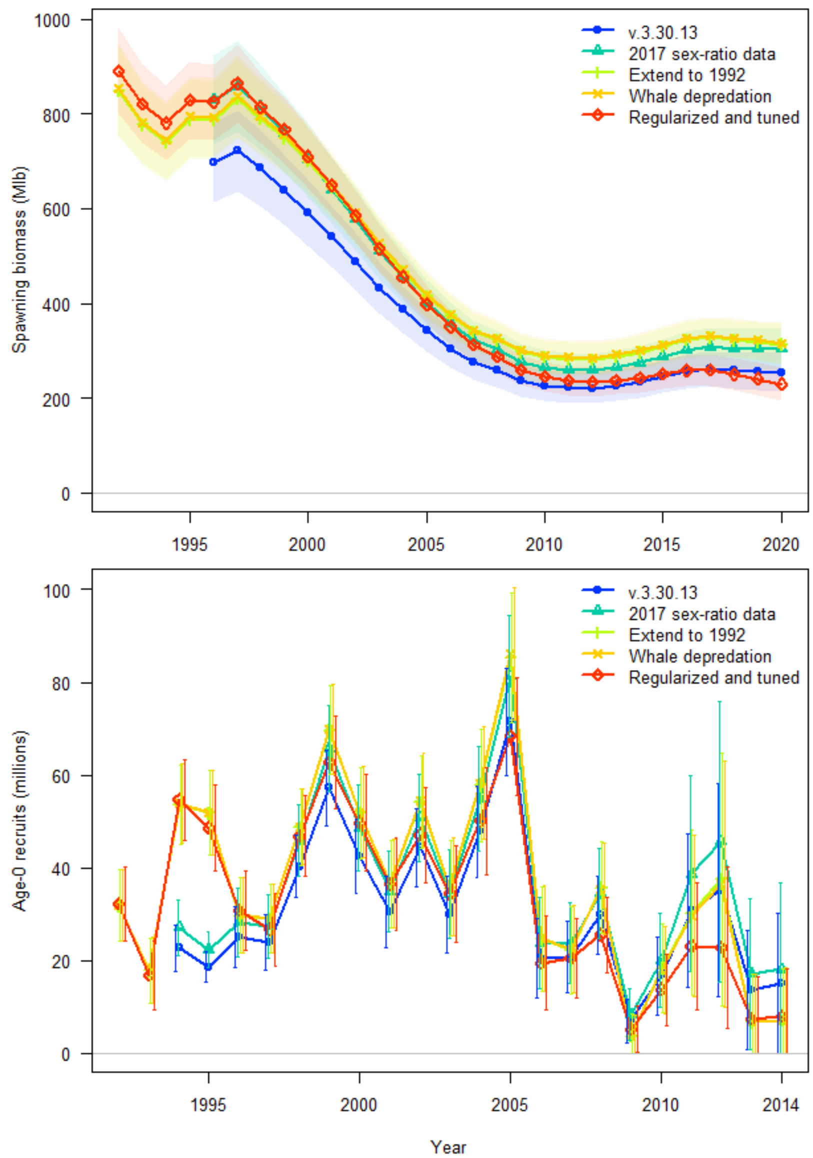

FIGURE 12. Comparison of estimated biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2018 to preliminary 2019 AAF short models.

Adding the sex-ratio data to the AAF long model and estimating the scale of the male selectivity for each Region (rather than assigning the values estimated for the survey in each Region) again increased the scale of the estimated spawning biomass substantially, suggesting fewer male halibut in the commercial landings than in the survey (FIGURE 13). A very large peak in the historical recruitment series, prior to the information content of the age data (beginning in 1935) appeared in this model where the 2018 assessment had estimated a short period of higher recruitment rather than a single large annual deviation. Again for the AAF long model, the modelled FISS time-series including stricter criteria for whale depredation had no visible effect on the results. Regularizing and tuning the final configuration including all of the new data revised the historical trends substantially, decreasing historical stock sizes before about 1960 (and eliminating the single large recruitment that appeared in the first bridging model) and increasing stock sizes from 1960through the mid-2000s. Despite these changes, the scale and trend at the very end of the time-series was similar to that from the 2018 stock assessment.

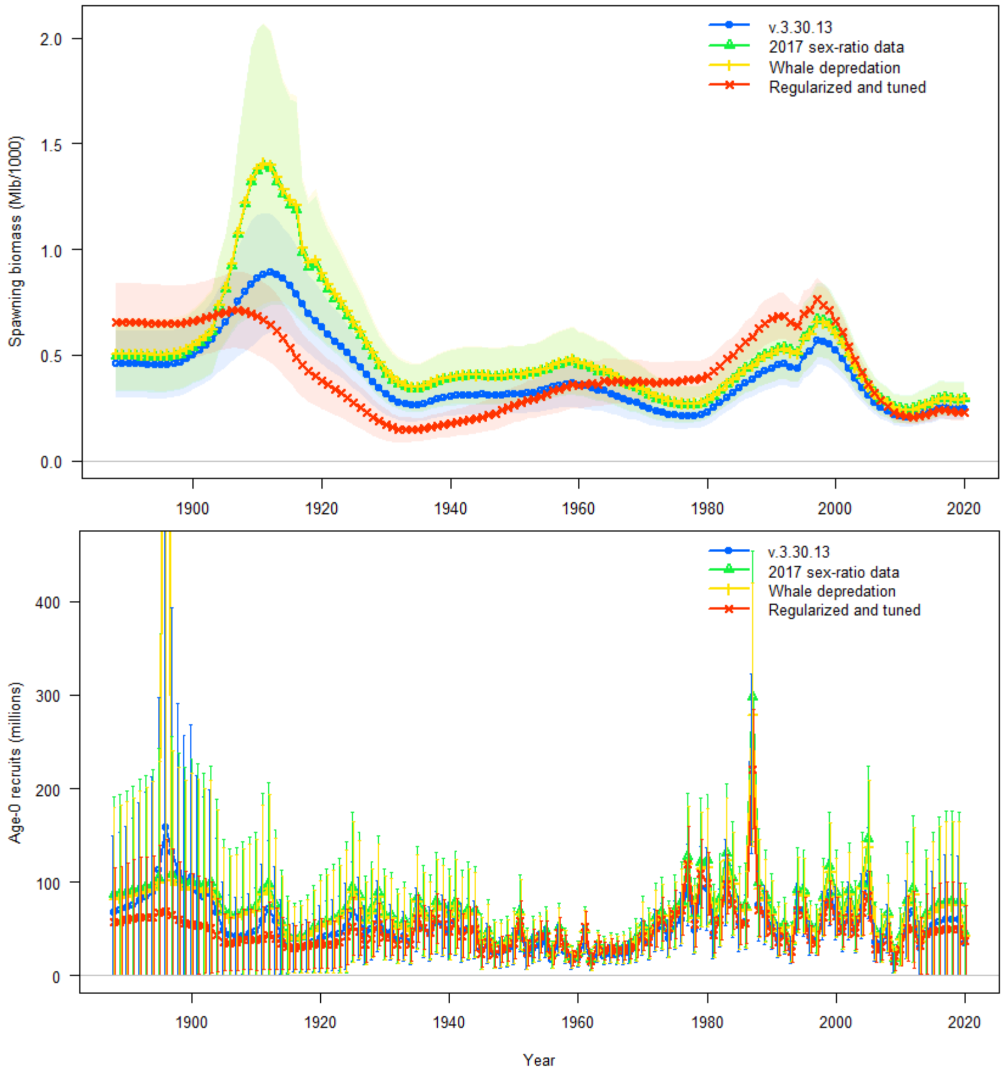

FIGURE 13. Comparison of estimated biomass (upper panel) and recruitment time series (lower panel) over sequential changes from the 2018 to preliminary 2019 AAF long models.

Overall, the inclusion of sex-ratio data resulted in higher spawning biomass for all models, and the updated whale depredation data made little difference. Extending the time-series back in the short models resulted in higher estimates of recruitment for 1994 and 1995. Regularizing and tuning the series had different effects on each model.

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 report the fit and residuals to every data component, but to summarize the basic performance of the model and specifically highlight areas of 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 pros and cons of each 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 14). In the 2018 assessment, a small amount of process error was allowed on fishery catchability. In this preliminary assessment, the iterative tuning of the annual deviations suggested it as no longer needed, and was therefore removed from the model. The predicted aggregate age distributions also matched the observed distributions quite well, for both the fishery and survey indicating that the selectivity approach was generally capturing differences in both the age-structure (and sex-ratio for the survey; FIGURE 15). The 2017 sex-ratio specific commercial data were not tuned separately from the remainder of the data, and it the model did not fit this information as closely as the rest of the series. Some lack of fit was also evident in the aggregate age composition data for the discard fleet; due to the downweighting of these data, several parameters were highly correlated and were fixed in the final model. Average input sample size by fleet (after adjustments) was substantially below the harmonic mean effective sample sizes for both the survey and fishery and the fishery data were weighted relatively less to achieve consistency with Francis weights (TABLE 11) which likely contributed to the very tight fit to the index time-series.

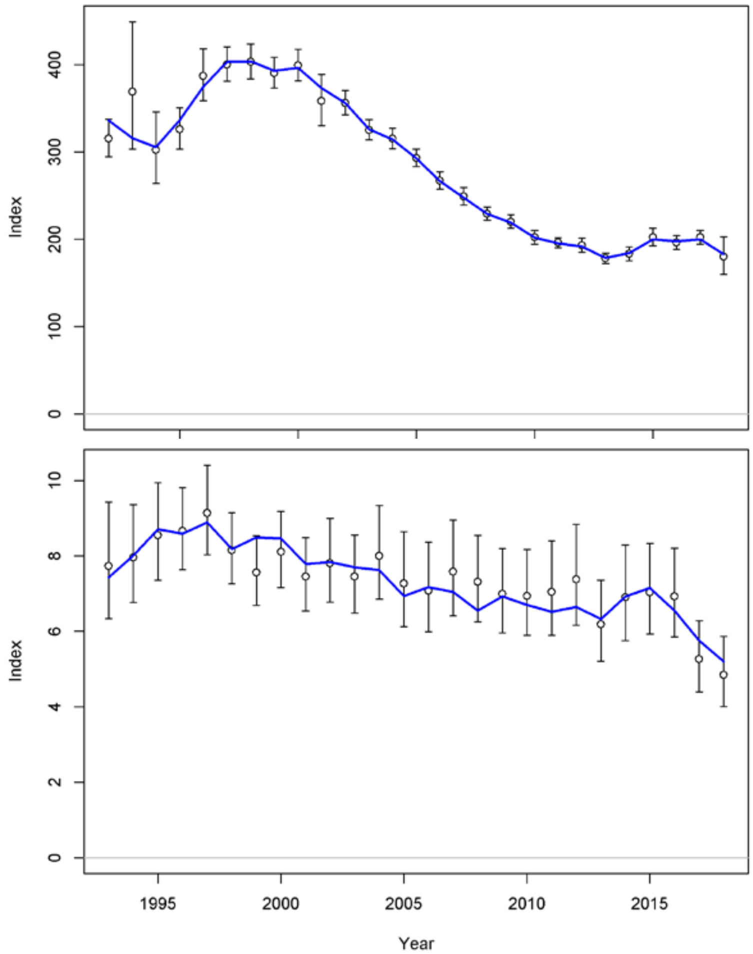

FIGURE 14. Fit to fishery (upper panel) and survey (lower panel) indices of abundance in the coastwide short model; note that the scale of the y-axes differ appreciably.

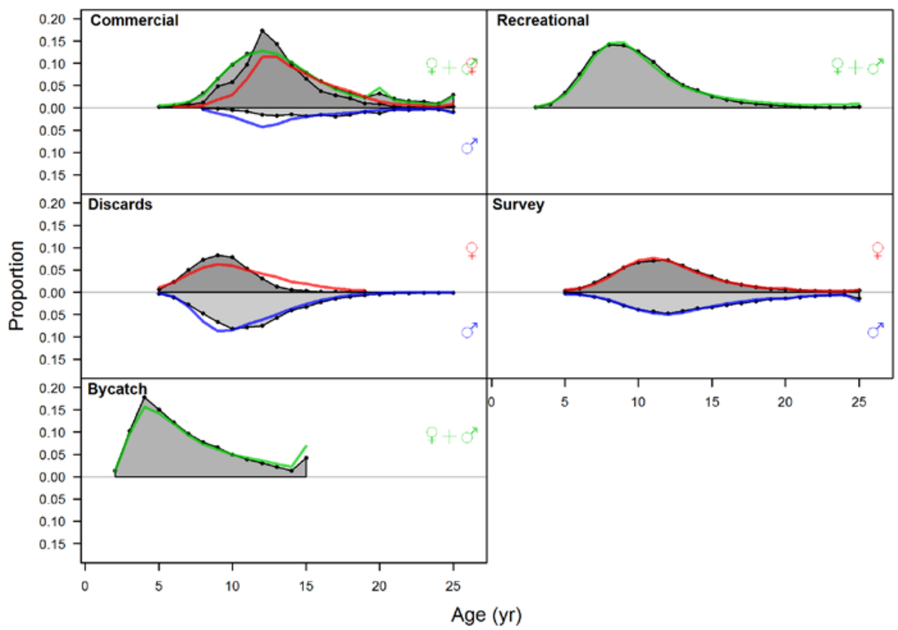

FIGURE 15. Aggregate fit to all age data by model fleet in the coastwide short model; sex-specific distributions for the commercial fishery represent only 2017.

Fit to the annual setline survey age compositions were good (FIGURE 16), although some patterning was visible in the standardized residuals (FIGURE 17). Specifically, there was a pattern of negative residuals in the plus group for male halibut; however, this was almost imperceptible in the fits themselves. The fits to the annual fishery data were also acceptable (FIGURE 18). 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 in the background materials.

TABLE 11. Post-iteration sample size diagnostics for age-composition data by model and fleet.

	Average iterated input	Harmonic mean effective	Francis weight samples	Maximum Pearson residual
Coastwide short				
Fishery	37	244	37	1.58
Discards[1]	9	126	79	0.89
Bycatch[1]	5	56	49	1.65
Sport[1]	5	109	35	0.93
Survey	372	724	372	2.48
Coastwide long				
Fishery	140	391	148	4.15
Discards[1]	6	234	118	0.58
Bycatch[1]	2.5	37	5	1.38
Sport[1]	2.5	118	23	0.72
Survey	125	196	125	3.81
AAF short				
Region 2 Fishery[2]	136	591	218	3.97
Region 3 Fishery[2]	127	570	229	2.20
Region 4 Fishery	40	64	40	3.80
Region 4B Fishery[2]	23	114	55	1.69
Discards[1]	6	216	134	0.73
Bycatch[1]	5	51	65	1.10
Sport[1]	5	117	27	0.70
Region 2 Survey	185	411	187	1.14
Region 3 Survey	240	575	235	1.93
Region 4 Survey	87	195	90	2.98
Region 4B Survey	40	188	40	1.34
AAF long				
Region 2 Fishery[2]	270	347	513	3.72
Region 3 Fishery[2]	167	347	334	3.76
Region 4 Fishery	30	61	30	5.28
Region 4B Fishery[2]	22	104	57	1.81
Discards[1]	6	222	95	3.82
Bycatch[1]	2.5	45	7	1.26
Sport[1]	5	132	24	0.68
Region 2 Survey	9	101	9	1.30
Region 3 Survey	43	154	43	1.85
Region 4 Survey	82	198	87	3.45
Region 4B Survey	40	192	42	1.56

[1]Inputs downweighted, and not iteratively reweighted – see text.

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

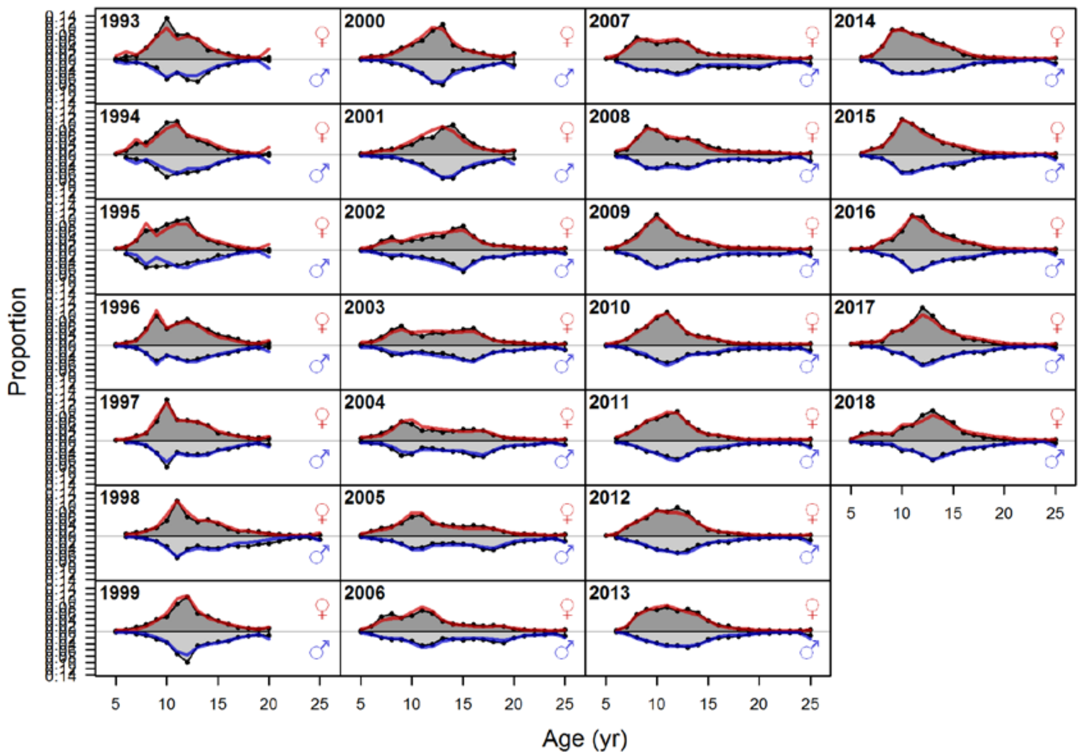

FIGURE 16. Fit to annual age data from the FISS survey in the coastwide short model.

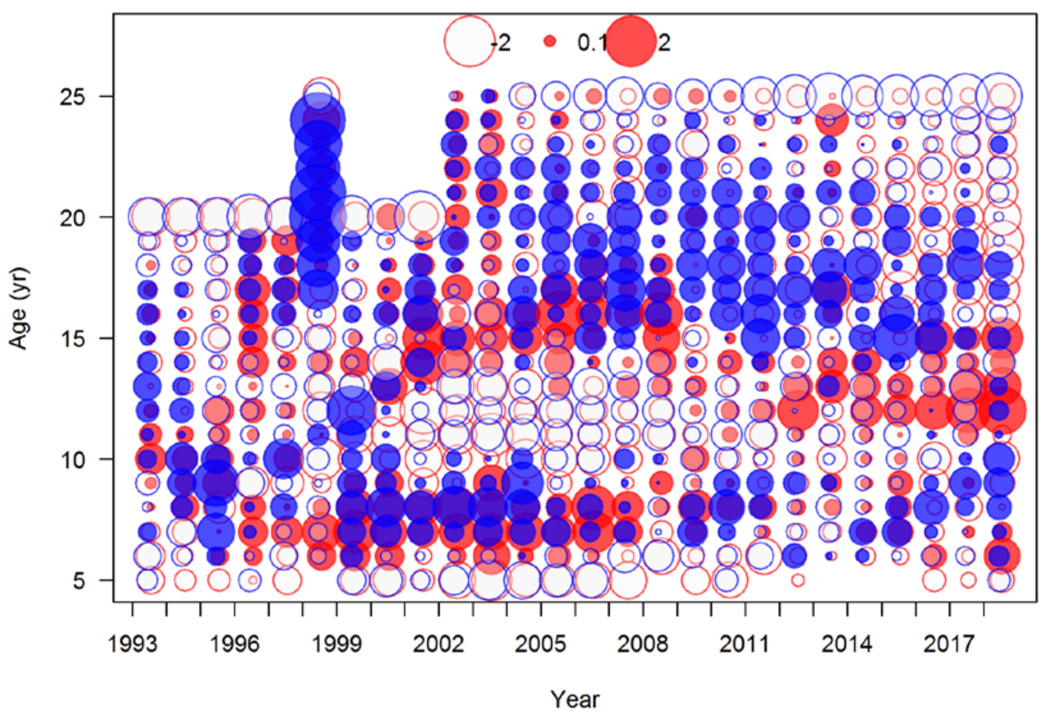

FIGURE 17. Pearson residuals for fit to annual age data from the FISS survey in the coastwide short model; red circles denote female residuals, and blue circles denote male residuals.

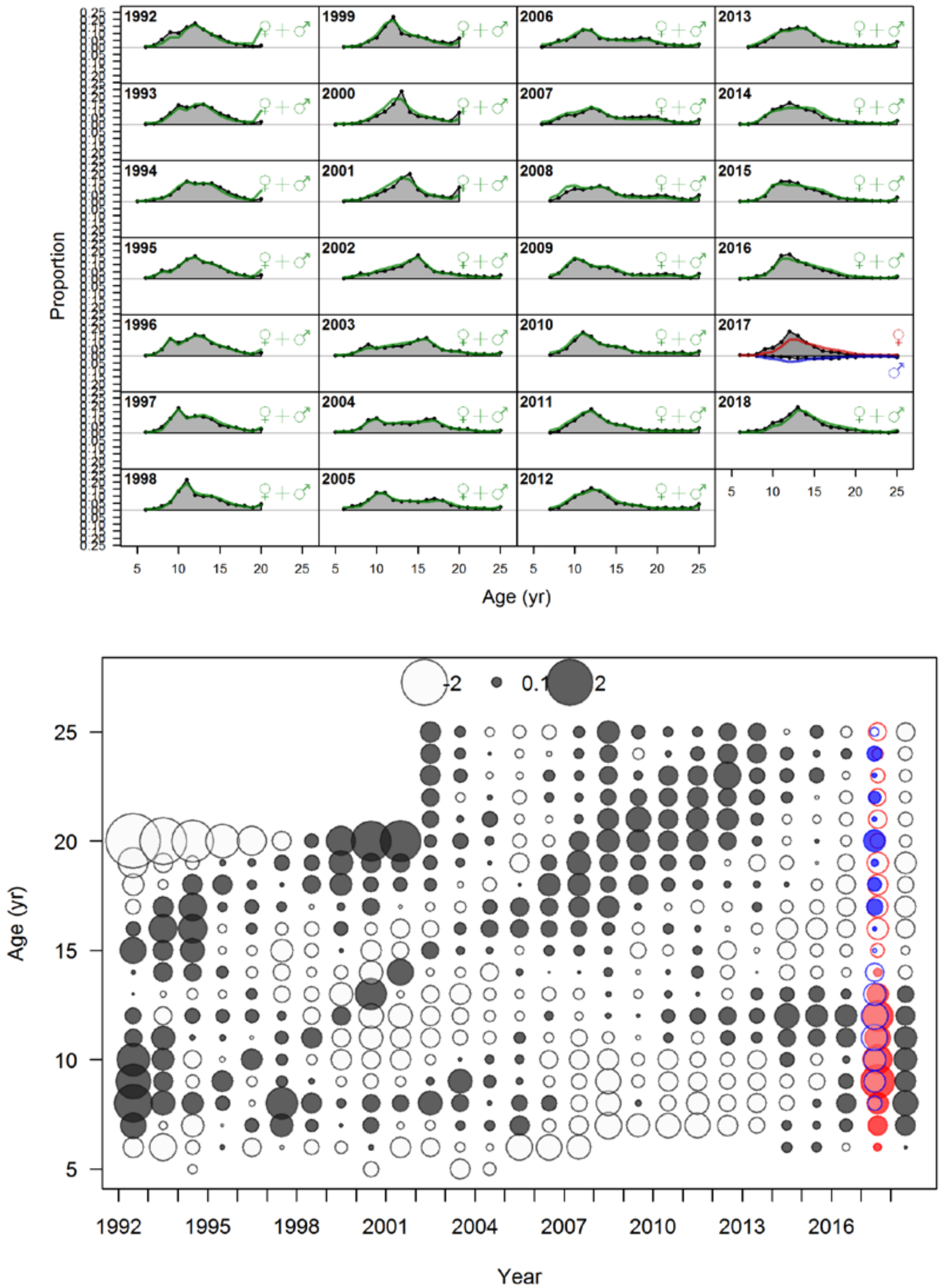

FIGURE 18. Fit (upper panel) and Pearson residuals (lower panel) for fit to annual age data from the commercial fishery landings in the coastwide short model; grey circles denote unsexed residuals, red circles denote female residuals, and blue circles denote male residuals.

Neither the survey nor the fishery selectivity was estimated to have a highly variable ascending limb over the short time-series (FIGURE 19). The estimated fishery selectivity showed a trend toward increasing selection of males in recent years (FIGURE 20), perhaps a function of the catch distribution shifting toward the Eastern side of the stock where fast-growing males are much more common, as well as the decline in the strong cohorts from the 1980s which produced an abundance of older females. The survey estimates did not show this trend, but selected a much larger relative fraction of females.

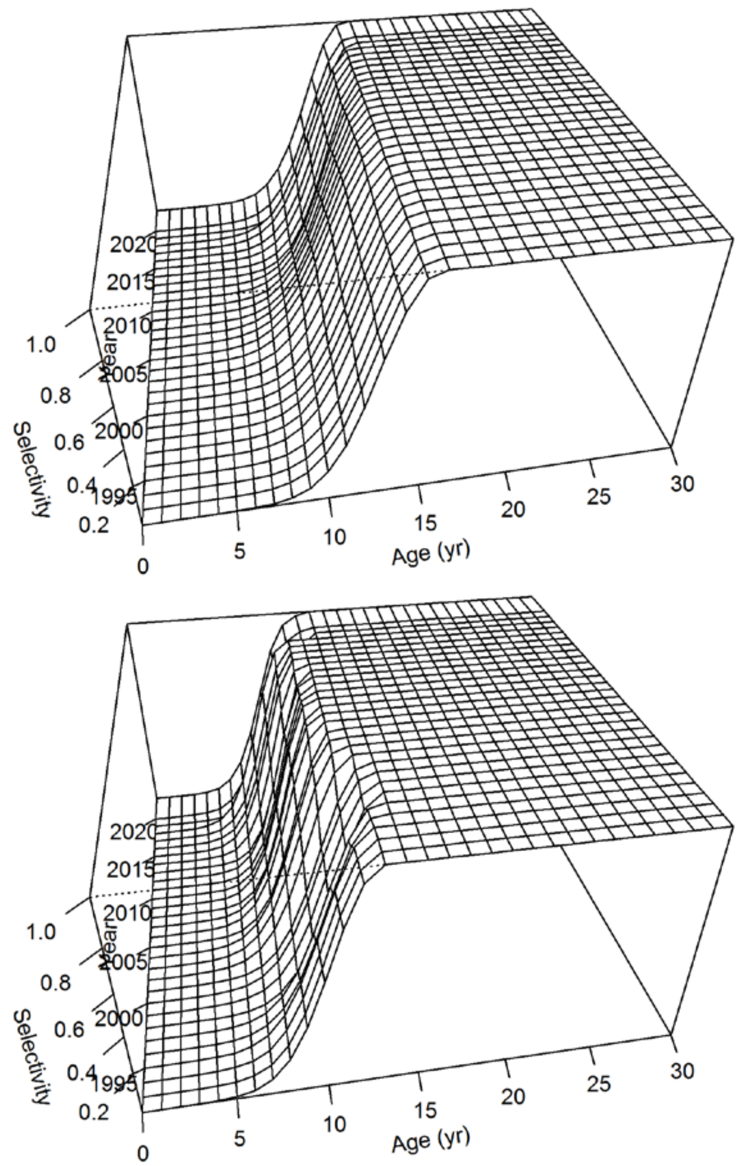

FIGURE 19. Estimated time-varying female selectivity curves for the commercial fishery landings (upper panel) and FISS survey (lower panel).

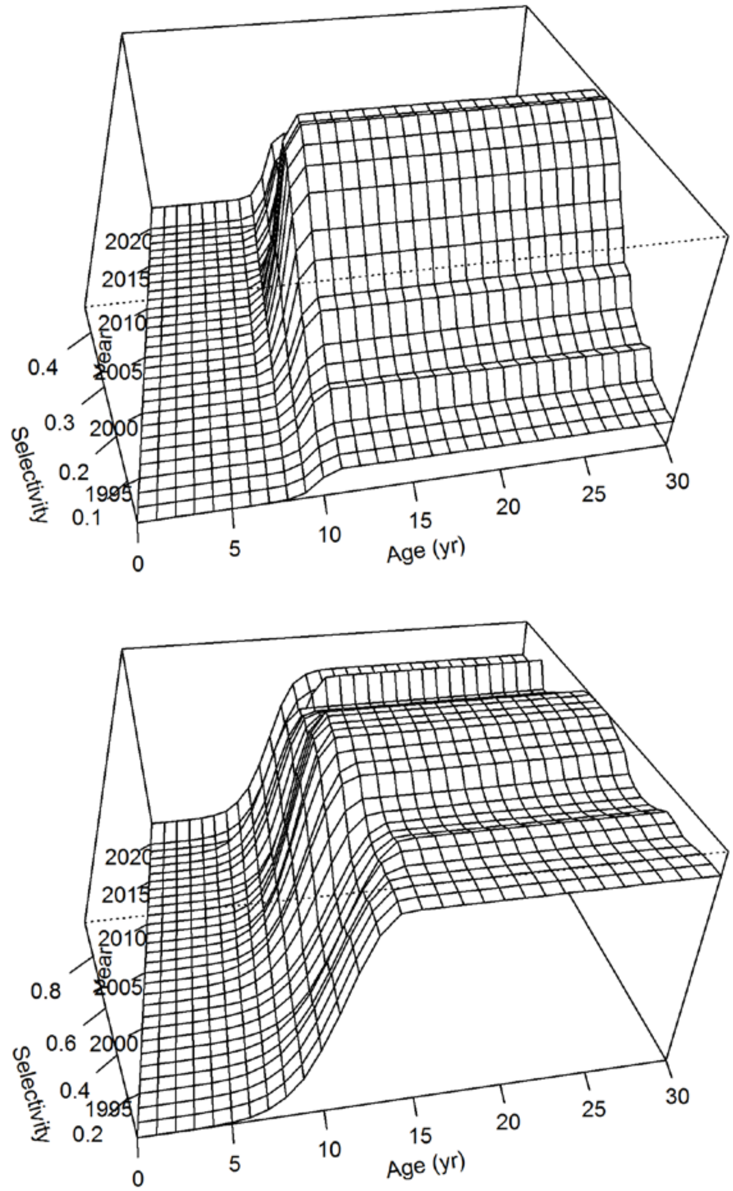

FIGURE 20. Estimated time-varying male selectivity curves for the commercial fishery landings (upper panel) and FISS survey (lower panel).

Estimated selectivity for the discard fleet selected fewer males than females (FIGURE 21). Estimated selectivity for the bycatch fleet showed a peak at age-4 and a domed relationship. Recreational selectivity was shifted to the left of the commercial fishery discards (and therefore the survey), reflecting the increased numbers of halibut age-7 and younger in the data from the Gulf of Alaska.

Male natural mortality was estimated to be slightly higher (0.155) than the fixed value assumed for females of 0.15 (TABLE 12), which differed from the slightly lower value estimated in the previous assessment (although still inside the 95% interval).

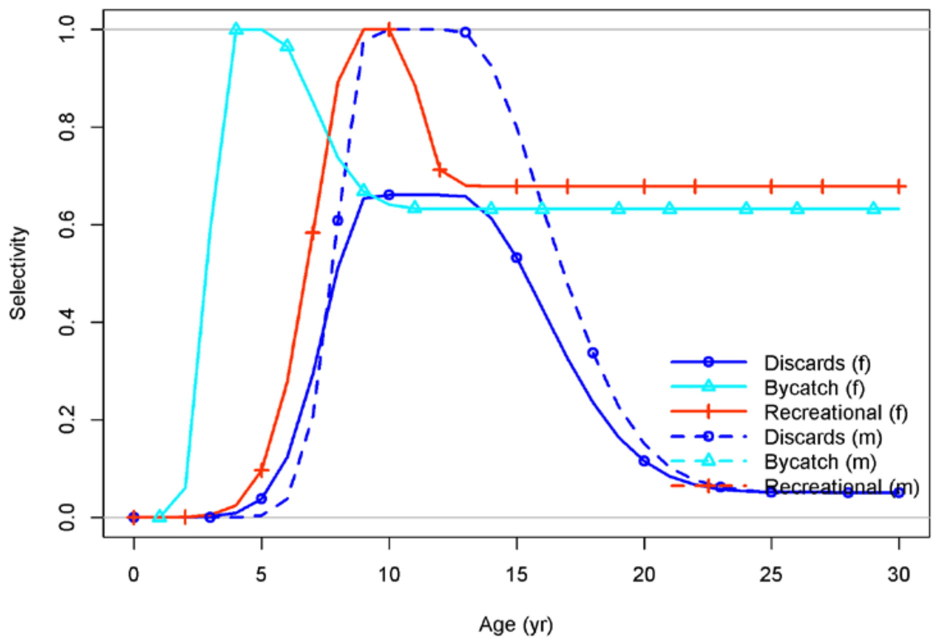

FIGURE 21. Estimated selectivity curves for discard, bycatch and recreational fleets in the coastwide short model.

Summary of pros and cons for the coastwide short model:

Pros:

- Lowest technical overhead (complexity) of the four models in the ensemble
- Fits the fishery and survey indices very well
- Fits the survey age data (males and females) relatively well
- Allows for changes in sex-ratio of the commercial landings over time
- Parameter estimates are derived from the most recent time period
- Internally consistent data weighting

Cons:

- Does not include uncertainty in female natural mortality
- 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
- Commercial age data is not heavily weighted and there are therefore residual patterns despite allowing for process error in selectivity

TABLE 12. Select parameter estimates (maximum likelihood value and approximate 95% confidence interval) by model and Region (where applicable).

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Biological				
Female *M*	0.150 (Fixed)	0.213 (0.188-0.238)	0.150 (Fixed)	0.173 (0.157-0.189)
Male *M*	0.155 (0.143-0.167)	0.199 (0.184-0.214)	0.140 (0.134-0.147)	0.155 (0.145-0.165)
Log(R_0)	10.63 (10.45-10.81)	11.06 (10.72-11.40)	10.68 (10.53-10.82)	10.66 (10.35-10.96)
Initial R_0 offset	-1.274 (-1.474--1.075)	NA	-0.659 (-0.833--0.485)	NA
Environmental Link (β)	NA	0.398 (0.167-0.629)	NA	0.293 (0.078-0.508)
Survey Log(q) Δ1984 (transition to circle hooks)	NA	0.943 (0.011-1.874)	NA	R2:1.209 (0.863-1.554) R3:2.100 (1.825-2.375)
Fishery Log(q) Δ1984	NA	0.654 (0.493-0.816)	NA	R2:0.573 (0.387-0.758) R3:0.934 (0.734-1.135) R4:0.784 (0.591-0.977) R4B:0.446 (0.263-0.629)
Scale of male survey selectivity (max value relative to females)	Time-varying	0.501 (0.354-0.648)	R2: 0.308 (0.196-0.419) R3: 0.604 (0.516-0.692) R4: 0.414 (0.340-0.488) R4B: 1.000 (Fixed at bound)	R2: 0.315 (0.222-0.408) R3: 0.494 (0.402-0.586) R4: 0.371 (0.310-0.432) R4B: 1.000 (Fixed at bound)
Scale of male fishery selectivity (max value relative to females)	Time-varying	0.362 (0.263-0.461)	R2: 0.113 (0.079-0.147) R3: 0.234 (0.171-0.298) R4: 0.086 (-0.002-0.174) R4B: 0.856 (0.455-1.000)	R2: 0.106 (0.074-0.139) R3: 0.220 (0.160-0.279) R4: 0.088 (0.033-0.143) R4B: 1.000 (Fixed at bound)

Coastwide long

Both the fishery and survey indices were fit well (FIGURE 22), with breaks in catchability to accommodate the change from “J” to circle hooks which were very conspicuous in both series (TABLE 12). In aggregate, the predicted age compositions matched the observed data well (FIGURE 23); however, there were notable differences among years within the time-series. Fits to the setline survey were quite poor in the early portion of the time series, improving where the data became more comprehensive in the mid-1990s, and quite good in the most recent years (FIGURE 24). Fishery data fit reasonably well for the entire time-series (FIGURE 25), with patterns in the residuals corresponding to relatively small differences with observed distributions. Harmonic mean effective sample sizes were much larger than adjusted inputs when Francis weights were close to 1.0 (TABLE 11).

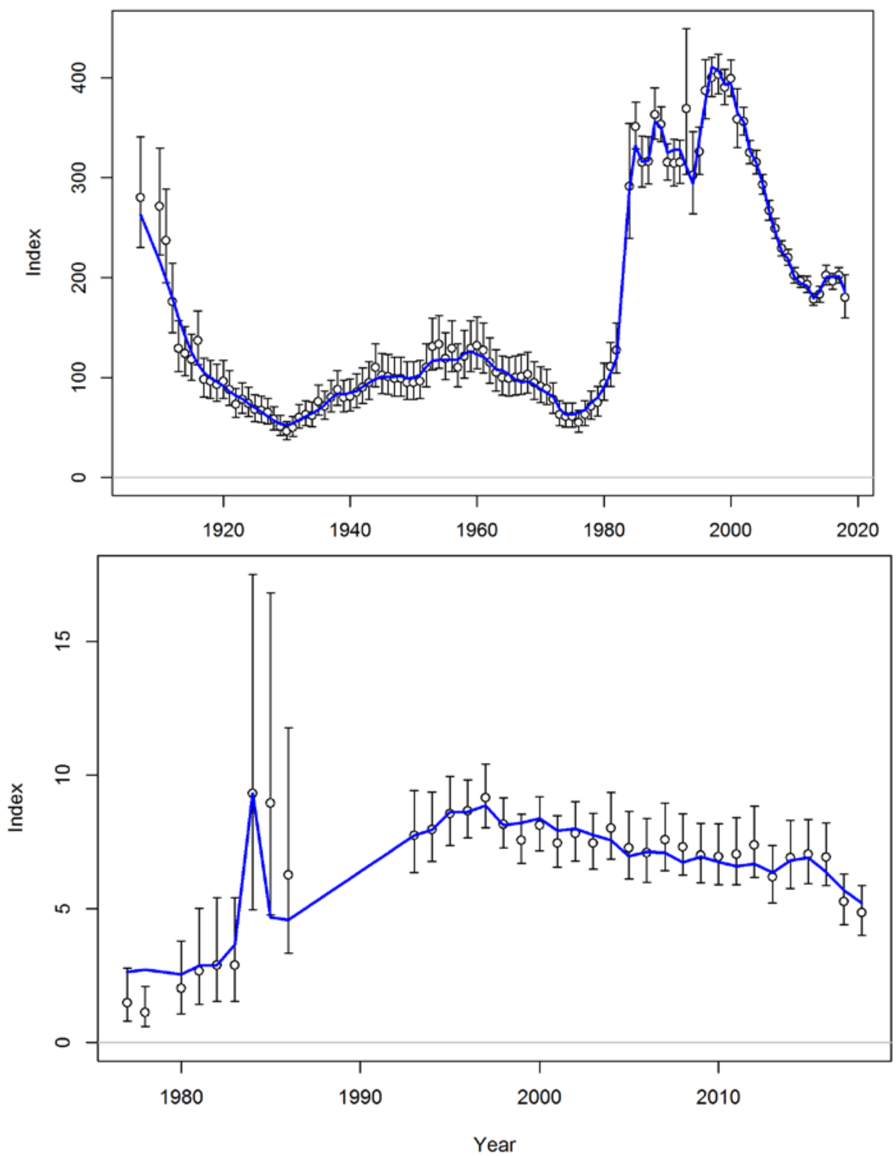

FIGURE 22. Fit to fishery (upper panel) and survey (lower panel) indices in the coastwide long model.

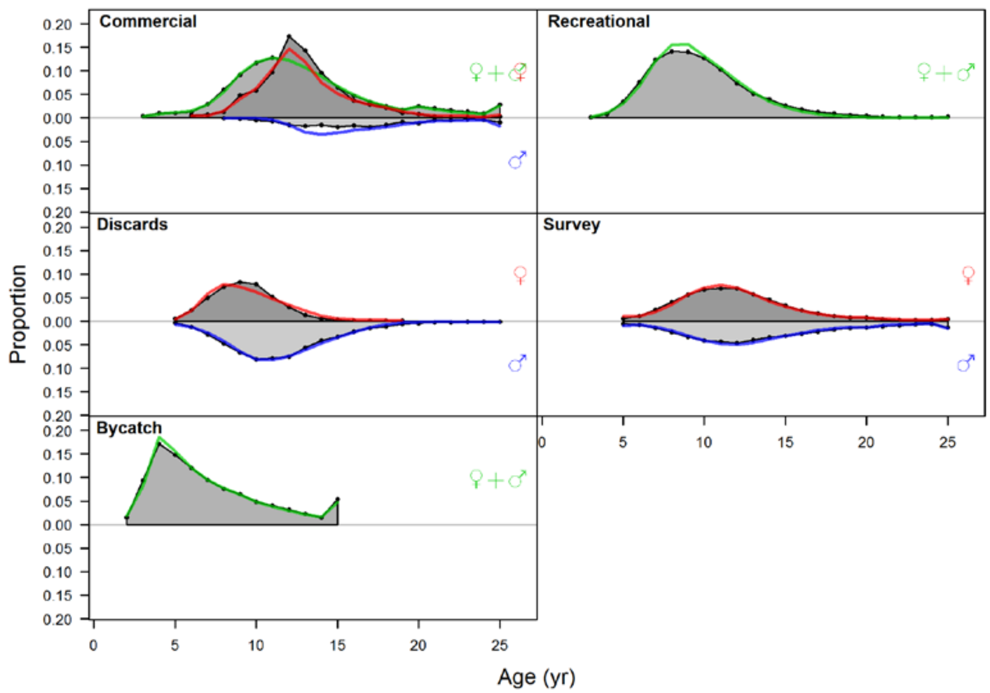

FIGURE 23. Aggregate fit to all age data by model fleet in the coastwide long model; sex-specific distributions for the commercial fishery represent only 2017.

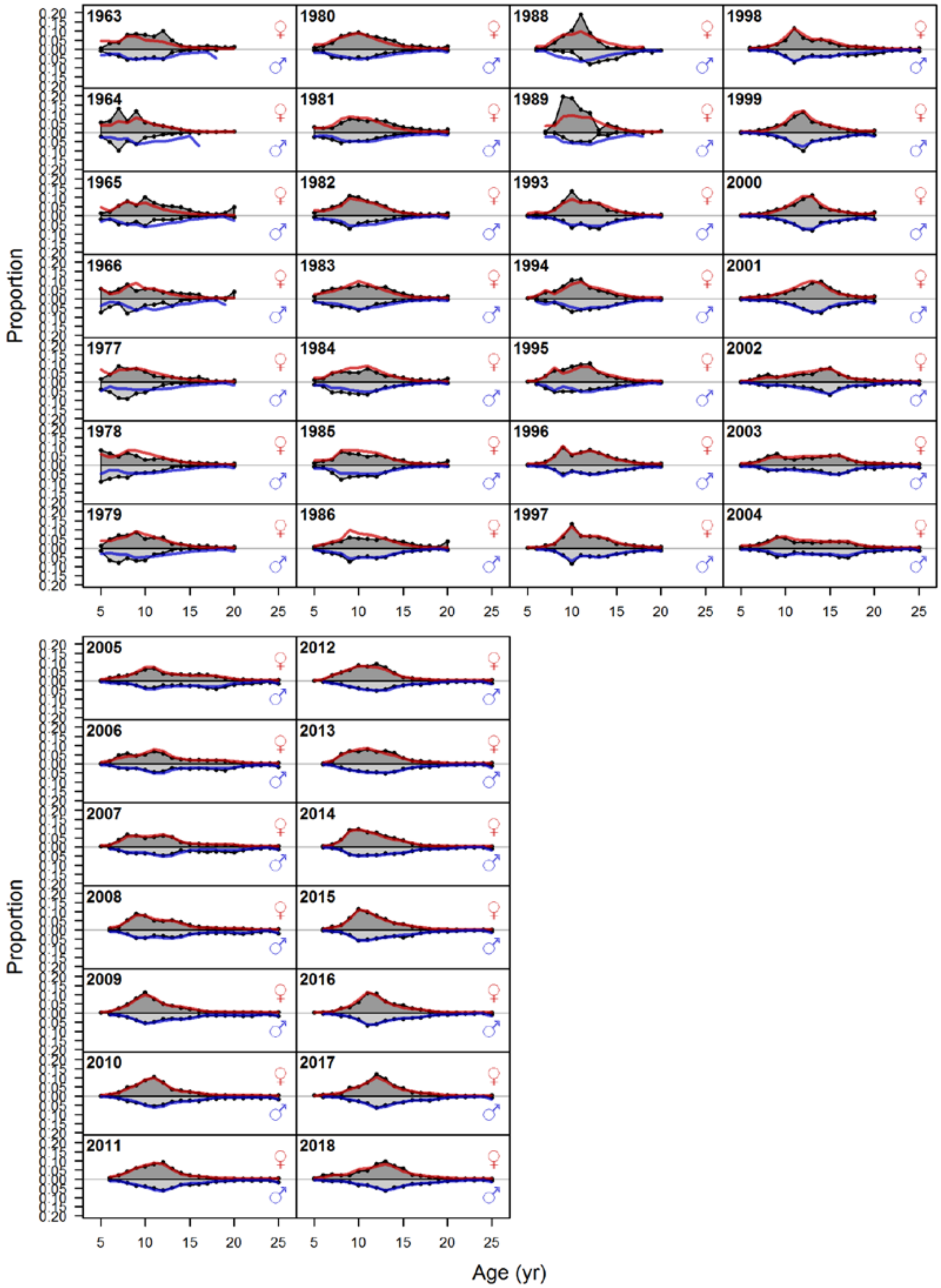

FIGURE 24. Fit to survey age data in the coastwide long model.

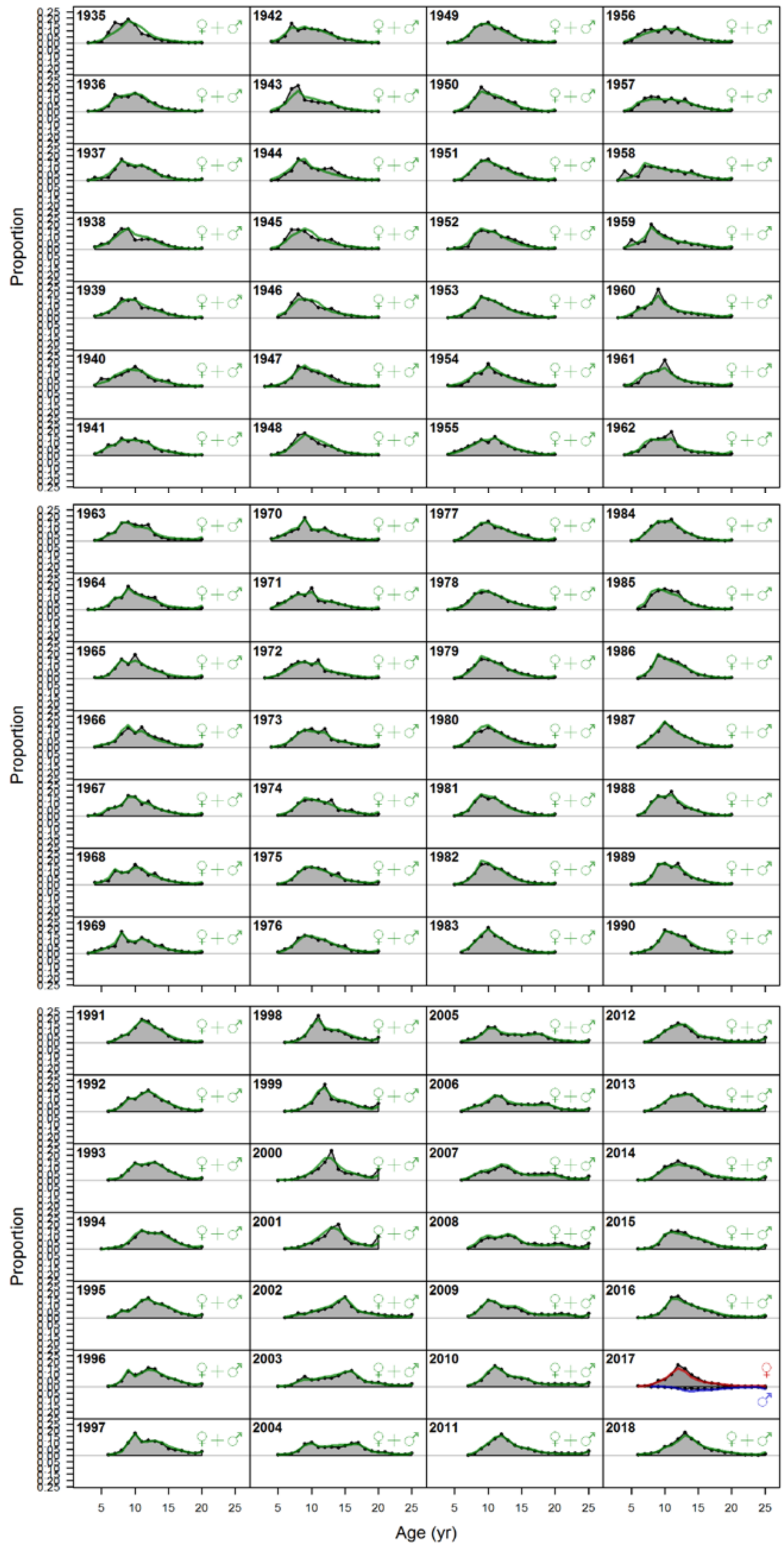

FIGURE 25. Fit to fishery age data in the coastwide long model.

Fishery selectivity generally showed a pattern toward selecting fewer younger fish in the latter half of the time series, but a similar trend to the setline survey in the most recent years (FIGURE 26). This may be consistent with changes in both the age-structure of the stock and the spatial distribution. Fishery catchability showed a very large (unconstrained) increase associated with the change from “J” to circle hooks (TABLE 12, FIGURE 27). Older halibut were more represented in the bycatch age data prior to 1992, and therefore the estimated selectivity was 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 (0.213) than for males although the 95% intervals overlap (0.199; TABLE 12, FIGURE 28). The environmental link parameter (β) was estimated to be positive (0.398), with no density below a value of 0.0 (TABLE 12, FIGURE 29). However, the time series of estimated recruitment deviates (FIGURE 30) suggested that some residual effect and/or mismatch in the relationship might still be present, as the poor PDO period from 1947-1977 and the positive phase from 1978-2006 generally correspond to negative and positive deviations, respectively (FIGURE 31).

Summary of pros and cons for the coastwide long model:

Pros:

- 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

Cons:

- 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, natural mortality) over long historical period (beyond environmental effects)

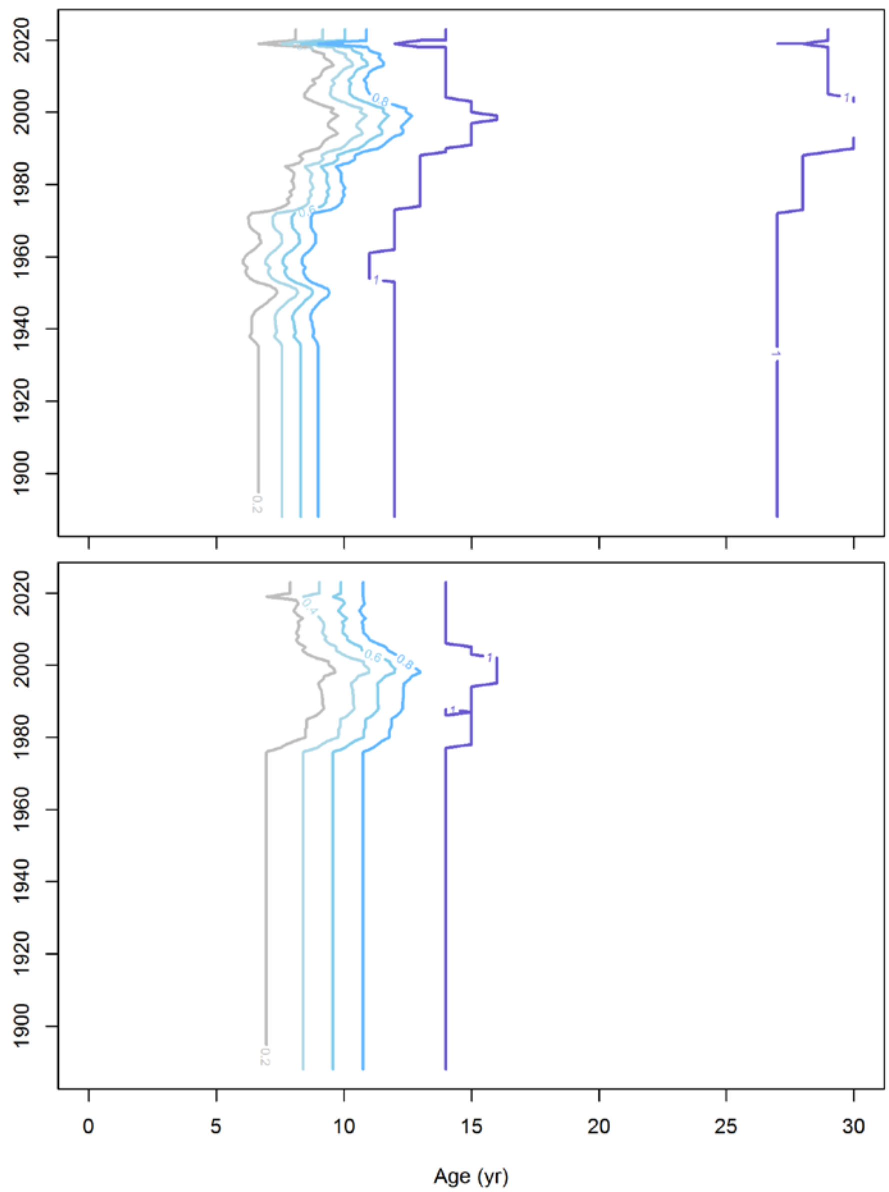

FIGURE 26. Estimated selectivity for females in the commercial fishery landings (upper panel) and survey (lower panel) in the coastwide long model; note that the apparent dip near the end of the time-series just corresponds to the fixed deviation in that year where there are not yet any data.

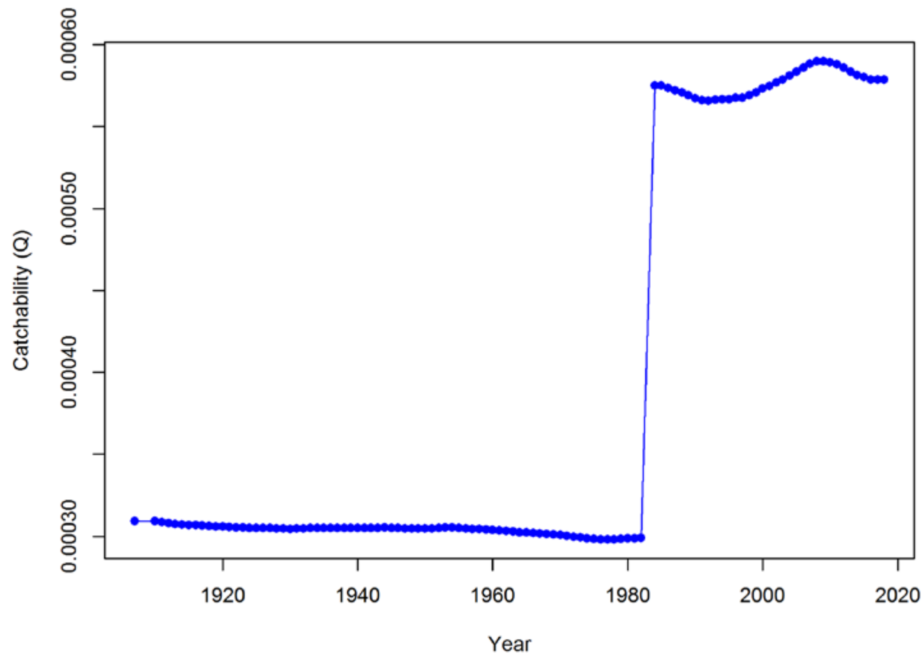

FIGURE 27. Time-varying fishery catchability in the coastwide long model.

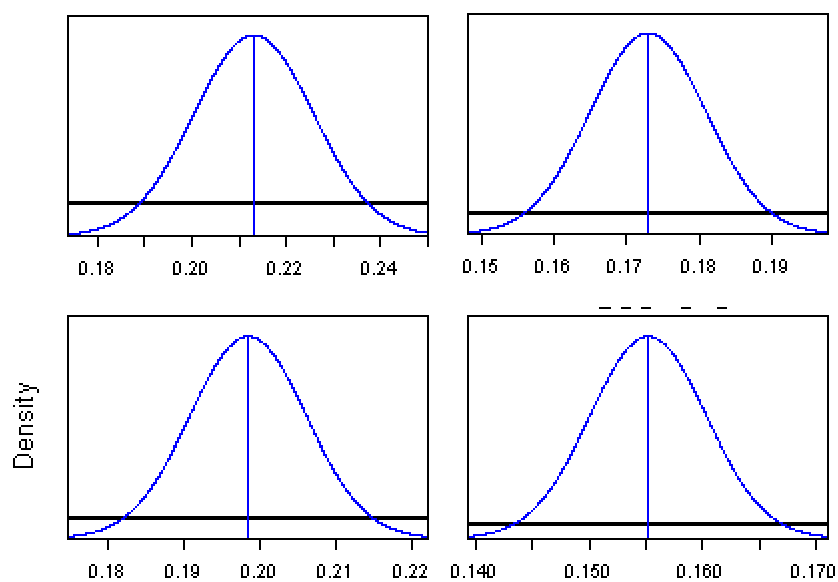

FIGURE 28. Estimated parameter distributions for female (upper panels) and male (lower panels) natural mortality from the coastwide long model (left panels) and the AAF long model (right panels); horizontal lines indicate uniform priors, vertical lines the maximum likelihood value.

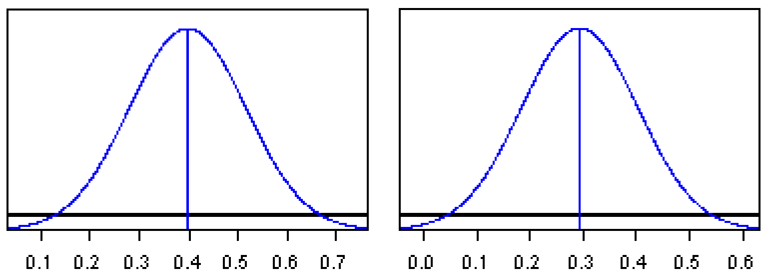

FIGURE 29. Estimated parameter distributions for the environmental regime parameters from the coastwide long model (left panel) and the AAF long model (right panel); horizontal lines indicate uniform priors, vertical lines the maximum likelihood value.

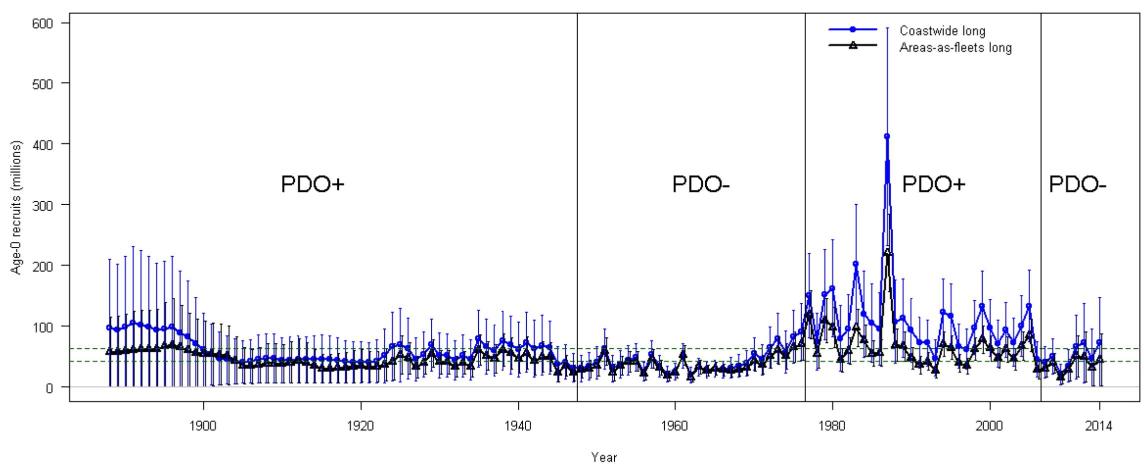

FIGURE 30. Estimated recruitments and assumed PDO regimes from the coastwide long and AAF long models (right panel); horizontal lines indicate equilibrium values.

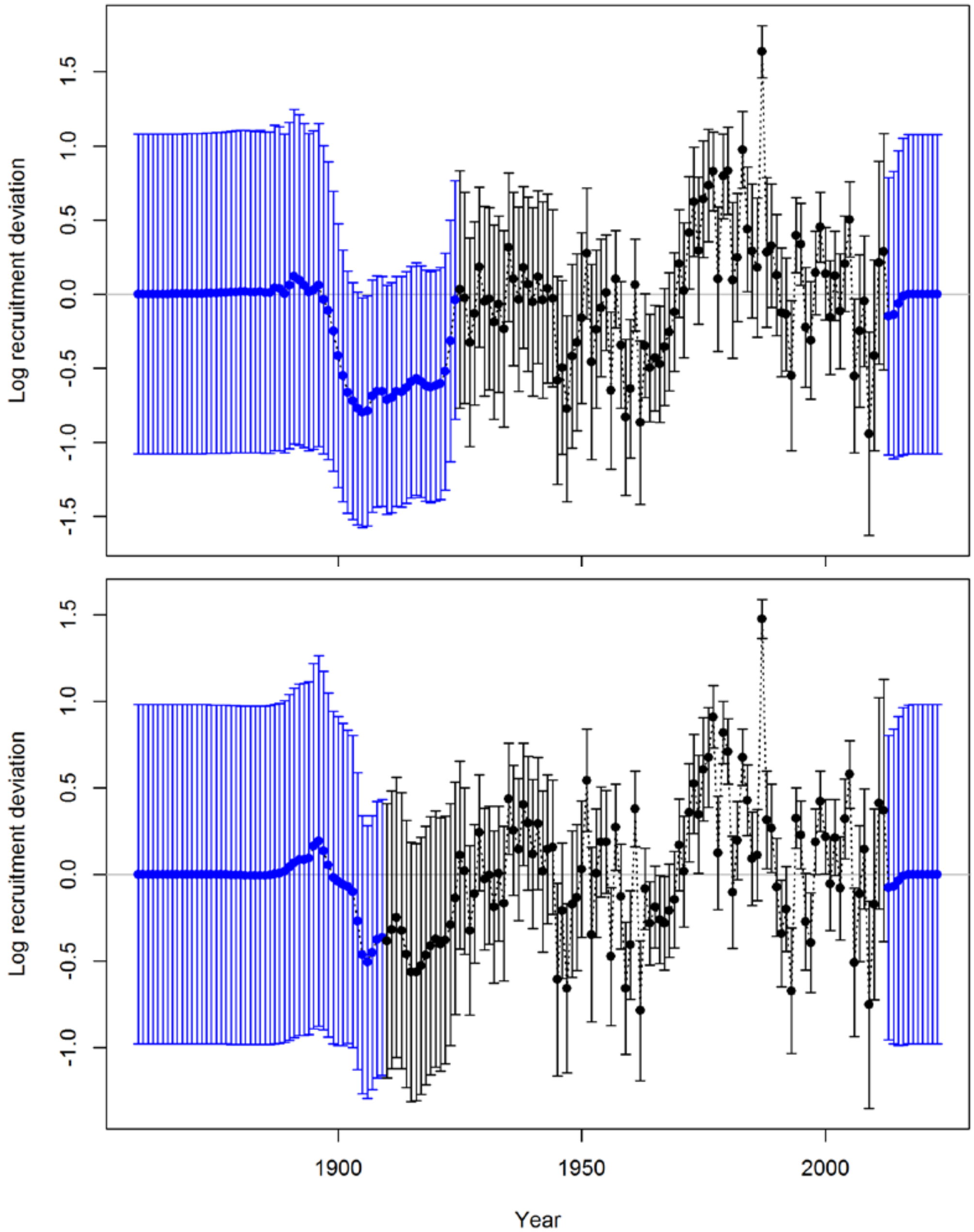

FIGURE 31. Estimated recruitment deviations coastwide long (upper panel) and AAF long (lower panel) models; horizontal lines indicate expected values based only on the stock-recruitment functions.

AAF short

The AAF short model fit the observed trends in all fishery and survey indices relatively well (FIGURE 32-33). These fits were somewhat better than those from the 2015 stock assessment, particularly for Regions 2 and 3 (Stewart and Martell 2016). Fit to the aggregate age data for each model fleet clearly illustrated the differences in age structure (FIGURE 34). The biggest differences between female and male halibut occurred in the Region 3 survey, and generally Regions 4 and 4B were predicted (and observed) to have the greatest fraction of older halibut, particularly males. The fit to the annual survey age data generally captured these patterns, with the worst fit in Region 2 (FIGURE 35); the Francis weight still suggested a relatively high weighting for the Region 2 survey despite these patterns (TABLE 11). Although showing a reasonably good aggregate fit, the fit to annual commercial fishery landings in Regions 4 and 4B (FIGURE 36) did not capture the strong peaks created by the 1987 year-class in the late 1990s and early 2000s; however of these fleets only the Region 4 data were downweighted from the number of samples collected based on the Francis weighting (TABLE 11). No model configurations evaluated during model development were able to fit the peak observations of this cohort observed in Regions 4 and 4B, which may be a reflection of the spatial nature of the dynamics not well approximated by an AAF approach.

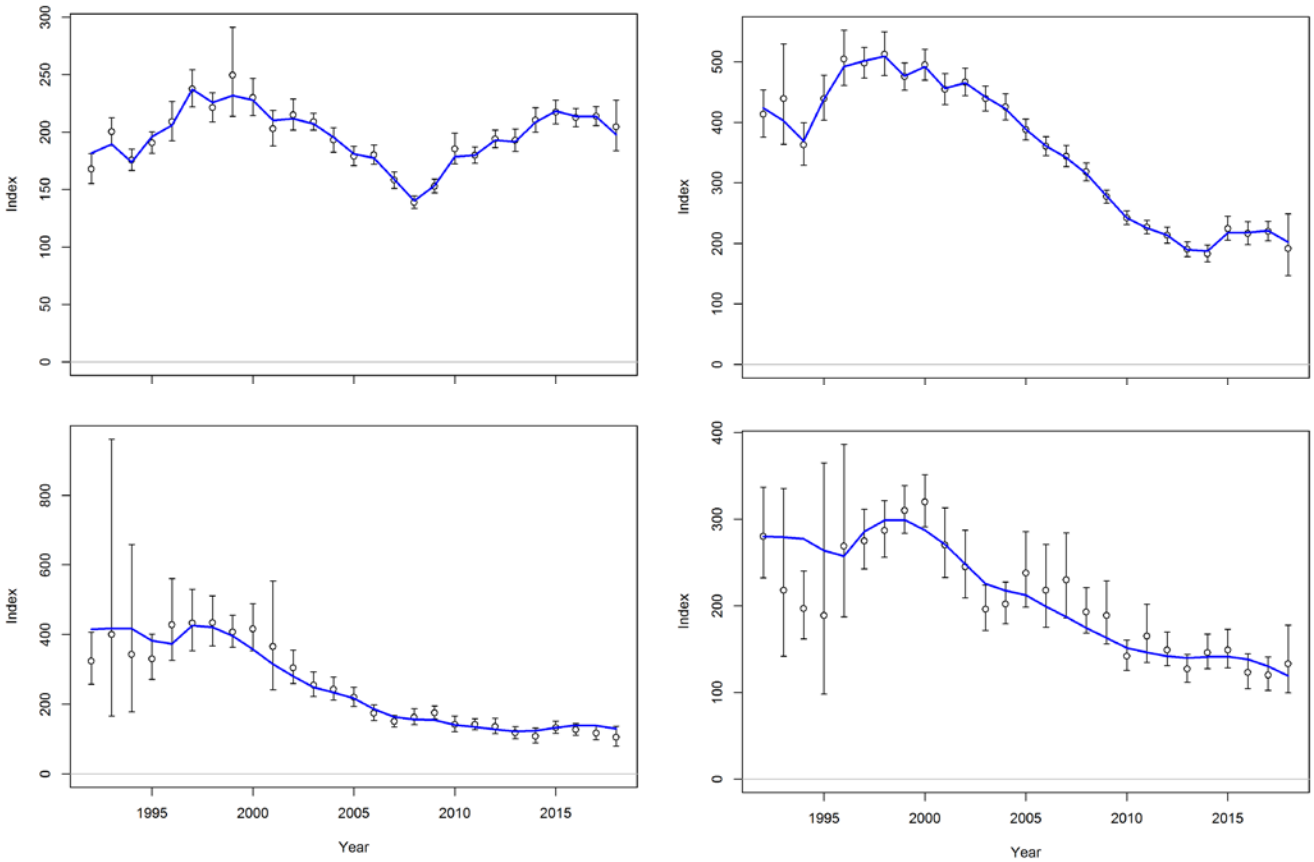

FIGURE 32. Fit to fishery trends in Regions 2, 3, 4, and 4B (top left to bottom right) in the AAF short model.

Male selectivity was estimated to be much less (0.31-0.6) relative to female selectivity for the survey in all Regions except 4B, where both were estimated to be fully selected and have a similar ascending limb (TABLE 12). Fishery selectivity was estimated to be shifted to the right of survey selectivity, and males were estimated to achieve a lower full selection relative to females in all Regions (0.086-0.856; TABLE 12). Bycatch, sport and discard selectivity estimates were similar to those from the coastwide short model.

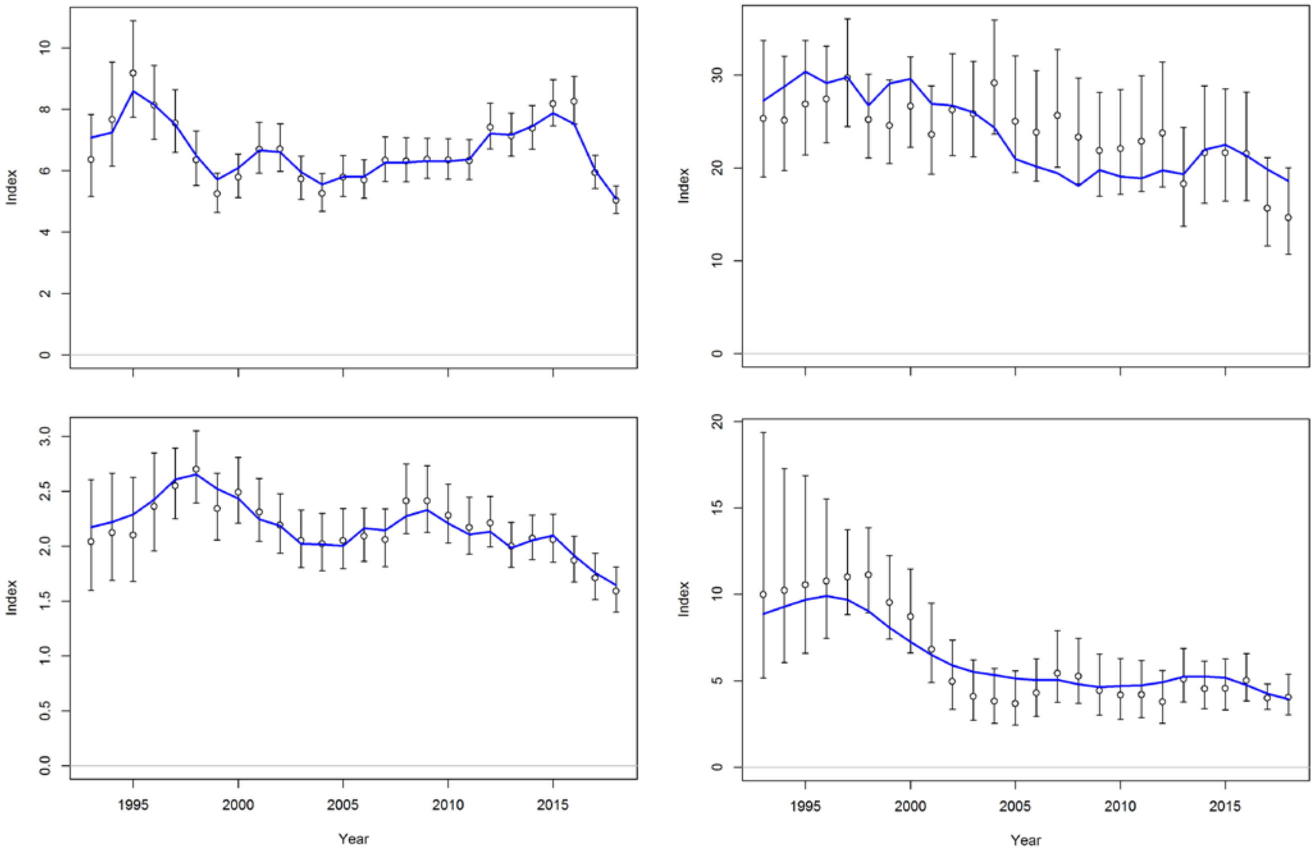

FIGURE 33. Fit to survey trends in Regions 2, 3, 4, and 4B (top left to bottom right) in the AAF short model.

Estimated fishery catchability showed differing temporal patterns and scale in Regions 2 and 3, with neither obviously showing a large amount of interannual variability (FIGURE 37). Although explored, tuning of process error deviations in catchability did not suggest retaining time-varying catchability for the Regions 4 and 4B fishery, and the fit to the indices remained consistent with the variance associated with the observations (FIGURE 32).

The estimate of male natural mortality in the AAF short model (0.14) was slightly lower than in the coastwide short model (TABLE 12) and the 95% intervals did not overlap that estimate. 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.

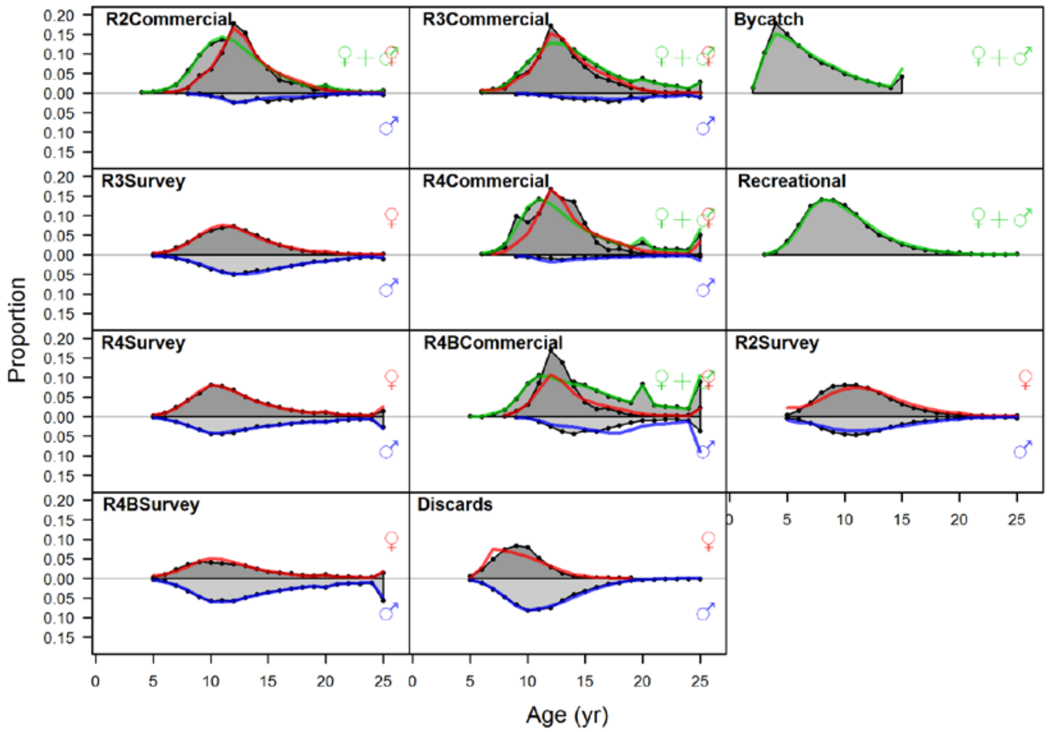

FIGURE 34. Aggregate fit to age data for each model fleet in the AAF short model; sex-specific distributions for the commercial fishery represent only 2017.

Summary of pros and cons for the AAF short model:

Pros:

- Parameter estimates are derived from the most recent time period
- Avoids aggregating data over Regions with differing trends and biological patterns
- Fits the Regional fishery and survey indices well
- Fits Region 2 and 3 fishery and Region 3 survey age data well
- Internally consistent data weighting

Cons:

- Does not includes uncertainty in female natural mortality
- Does not include environmental information to inform recruitment
- Modest technical overhead (complexity)
- Residual patterns in Region 4 and 4B fishery and survey age data
- Fits Region 2 survey age data poorly
- Does not include extensive historical data

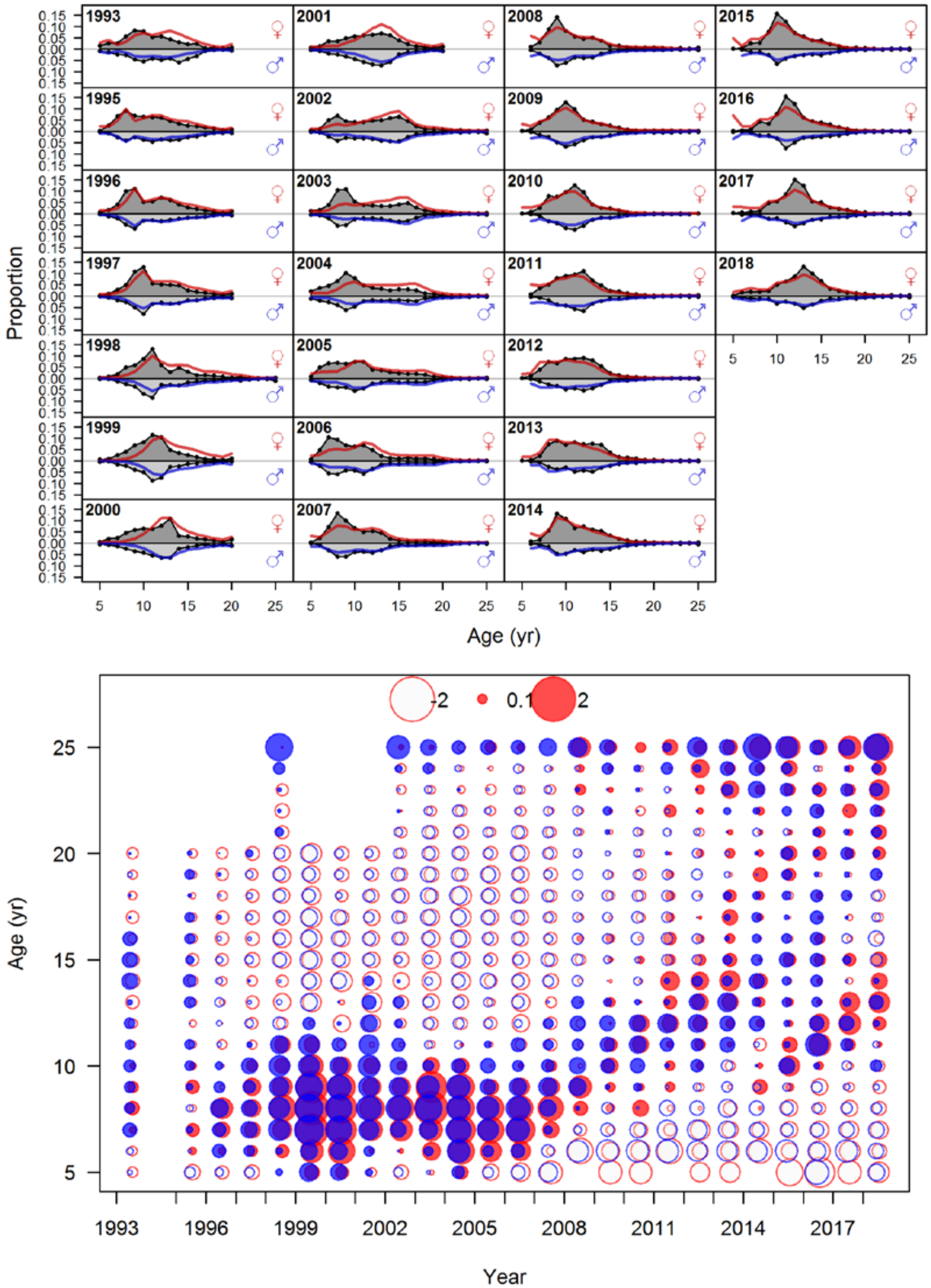

FIGURE 35. Fit to age data (upper panel) and Pearson residuals (lower panel) from the Region 2 survey in the AAF short model; red circles denote female residuals, and blue circles denote male residuals.

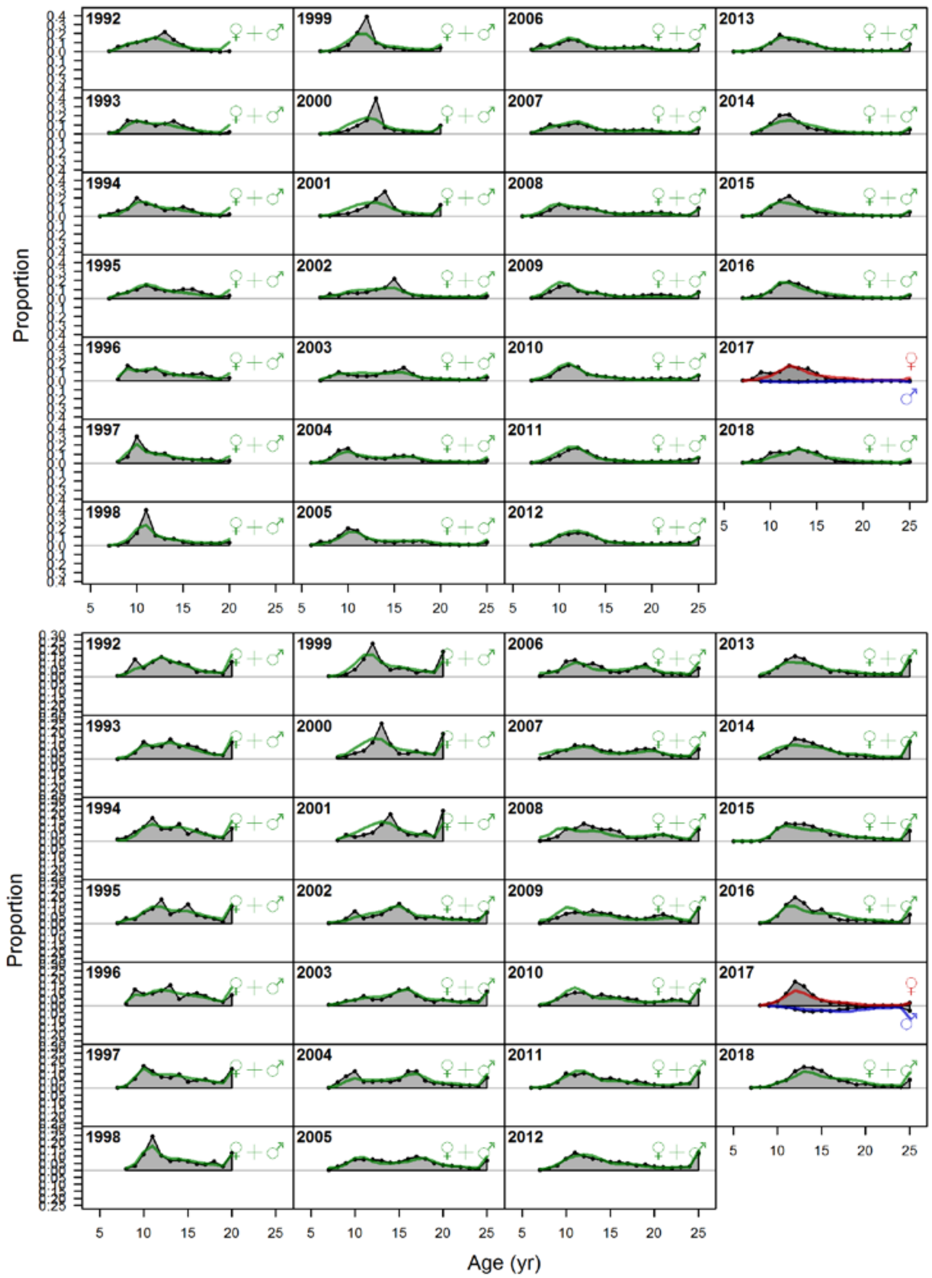

FIGURE 36. Fit to age data from the Region 4 (upper panel) and Region 4B (lower panel) commercial fishery landings in the AAF short model.

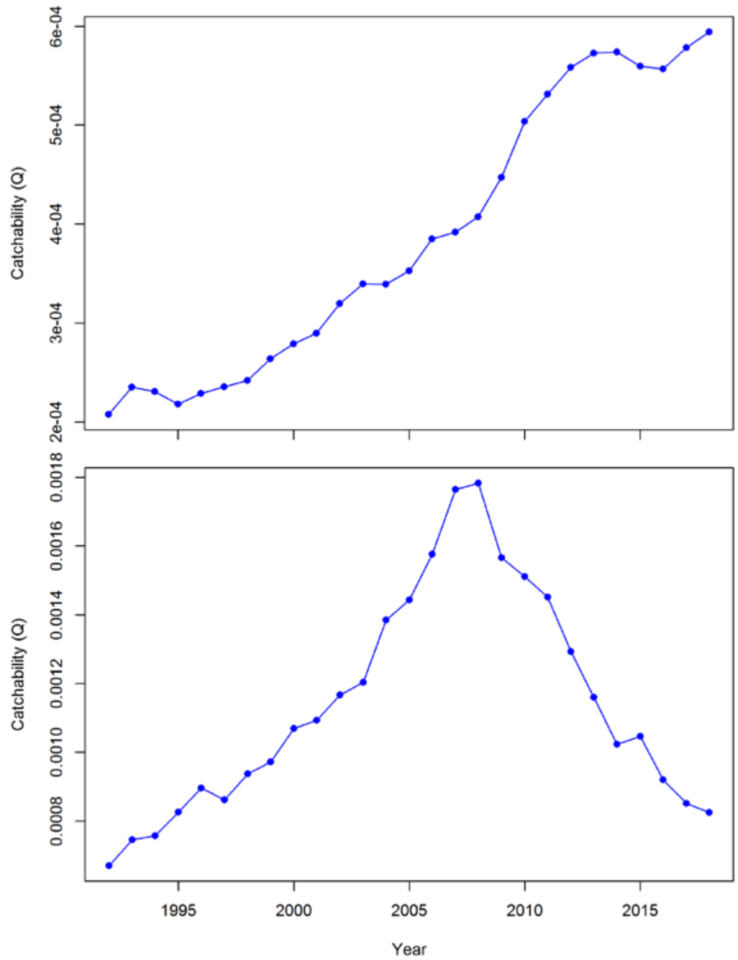

FIGURE 37. Estimated trends in the Region 2 (upper panel) and Region 3 (lower panel) commercial fishery catchbility in the AAF short model.

AAF long

Like the AAF short model, the AAF long model fit both the fishery and survey trends relatively well (FIGURE 38-39). Aggregate fits to the survey age composition data showed similar patterns to those observed in the AAF short model (FIGURE 40). Generally, the fit to the survey age data improved over the time series. The Region 2 survey age data was relatively downweighted in order to achieve consistency with the Francis weighting TABLE 11, and this resulted in the worst fit by fleet (FIGURE 41). Lack of fit to the Region 3 survey data occurred primarily in the early part of the time-series. Among the fishery fleets, only the Region 4 data were downweighted from the number of samples TABLE 11. Generally, as a function of the iterative weighting and the separation of commercial male selectivity (from the strong assumption in previous models that peak male selectivity was equal to that in the survey) the fits to the age data in this preliminary assessment were improved over previous analyses.

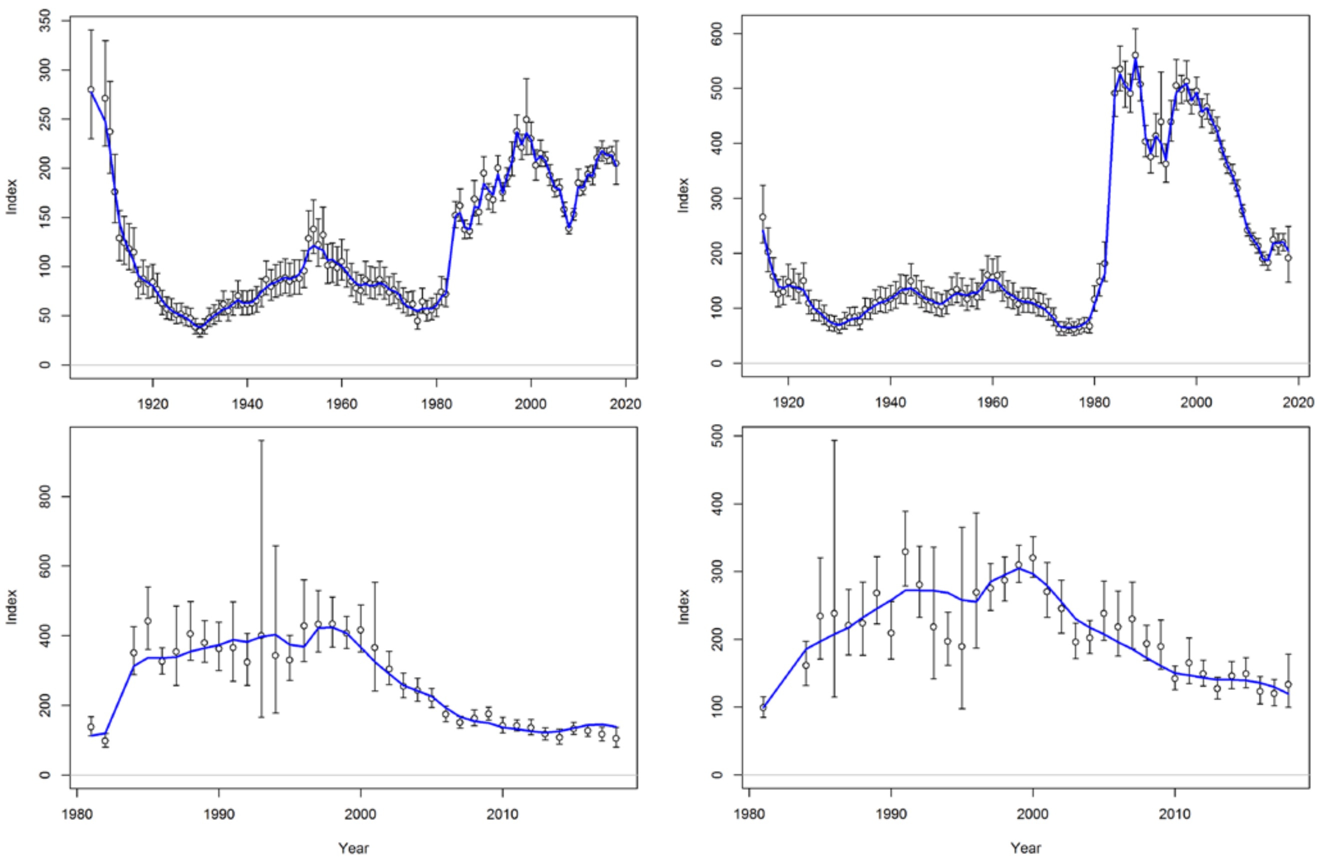

FIGURE 38. Fit to fishery trends in Regions 2, 3, 4, and 4B (top left to bottom right) in the AAF long model.

Similar to the AAF short model, peak male survey selectivity was estimated to be asymptotic only for Region 4B, ranging from 0.32-0.49 among the other Regions (TABLE 12). Peak male selectivity in the commercial fishery landings was estimated to be much less (0.09-0.22), except in Region 4B where it was also asymptotic. Fishery catchability was estimated to be strongly increasing in Region 2 and decreasing in Area 3 at the end of the time series (FIGURE 43). As in the AAF short model, tuning eliminated time-varying catchability for the Region 4 and 4B commercial fisheries. All fleets with data extending past the transition from J to circle hooks in 1984 showed a strong offset in the unconstrained deviation in catchability for that year (TABLE 12). Discard and recreational selectivity estimates were similar in the AAF long model to those estimated in the coastwide long model. Bycatch 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 (0.173 for females, and 0.155 for males) than in the coastwide long model (TABLE 12).

The environmental link coefficient was estimated to be slightly weaker (0.293) than in the coastwide long model, although the 95% interval did not contain zero (TABLE 12, FIGURE 29)

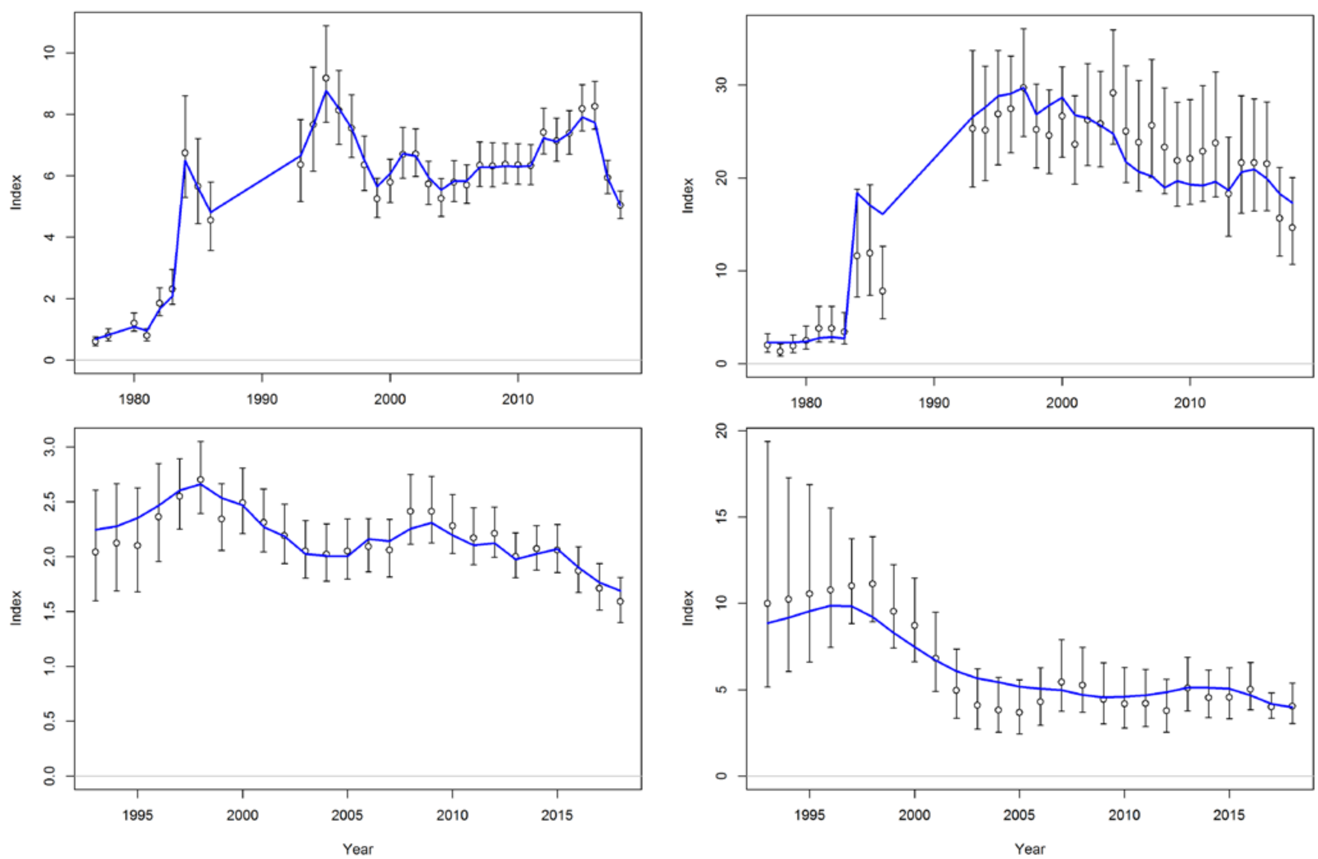

FIGURE 39. Fit to survey trends in Regions 2, 3, 4, and 4B (top left to bottom right) in the AAF long model.

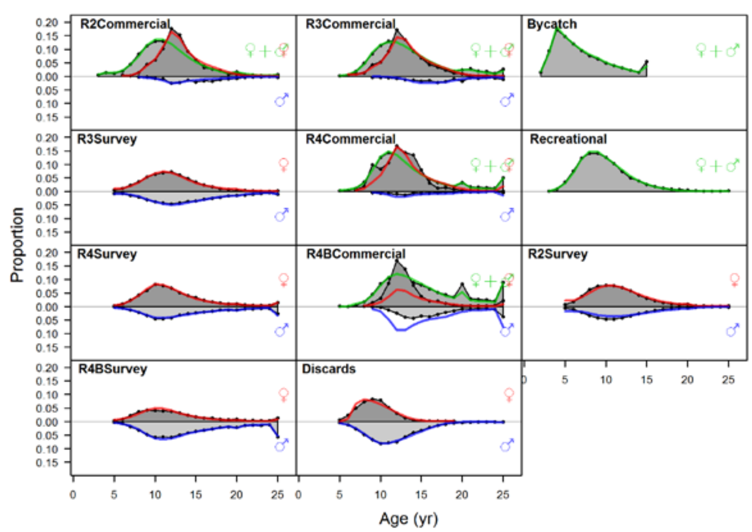

FIGURE 40. Aggregate fit to age data for each model fleet in the AAF long model; sex-specific distributions for the commercial fishery represent only 2017.

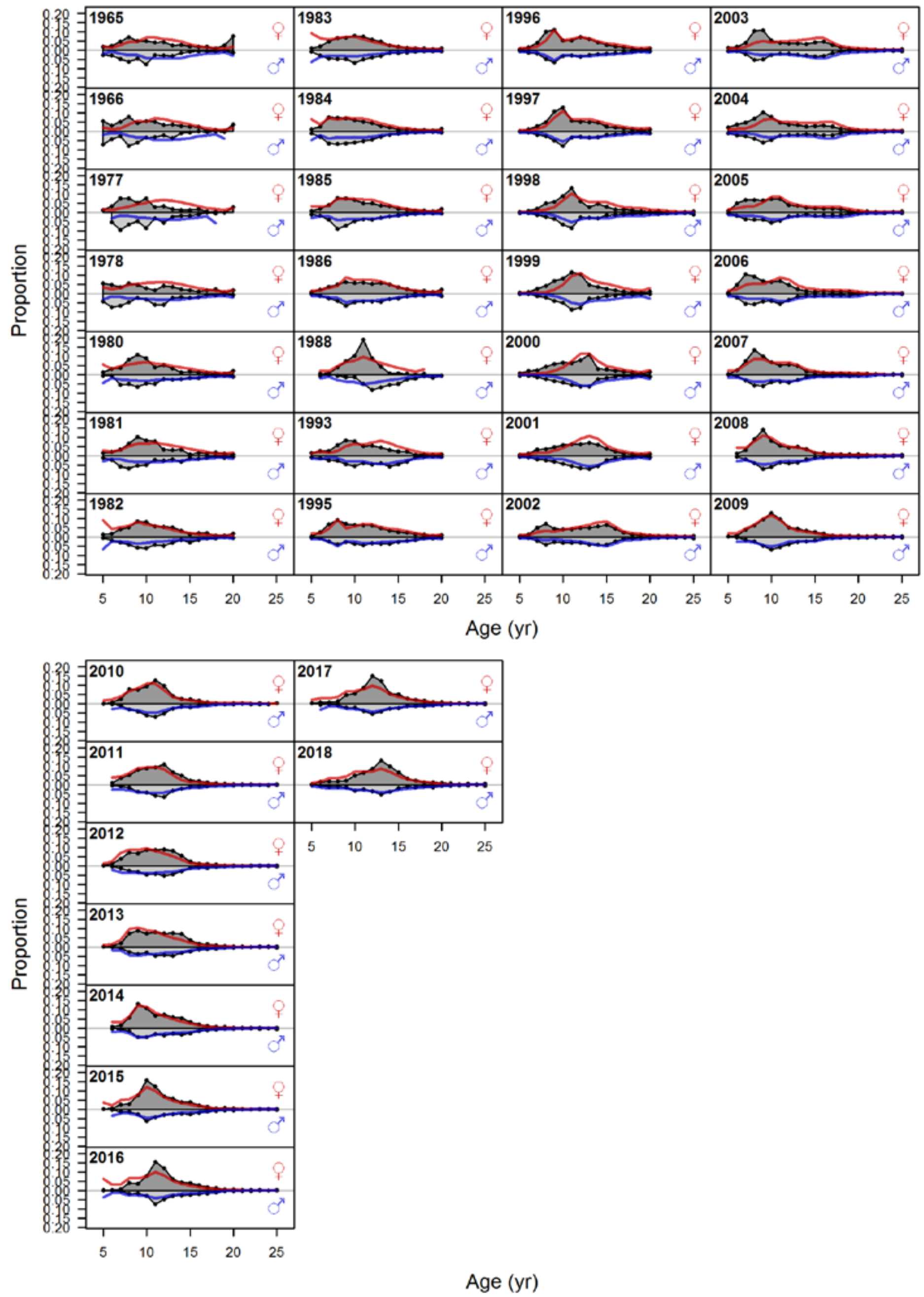

FIGURE 41. Fit to age data from the Region 2 survey in the AAF long model.

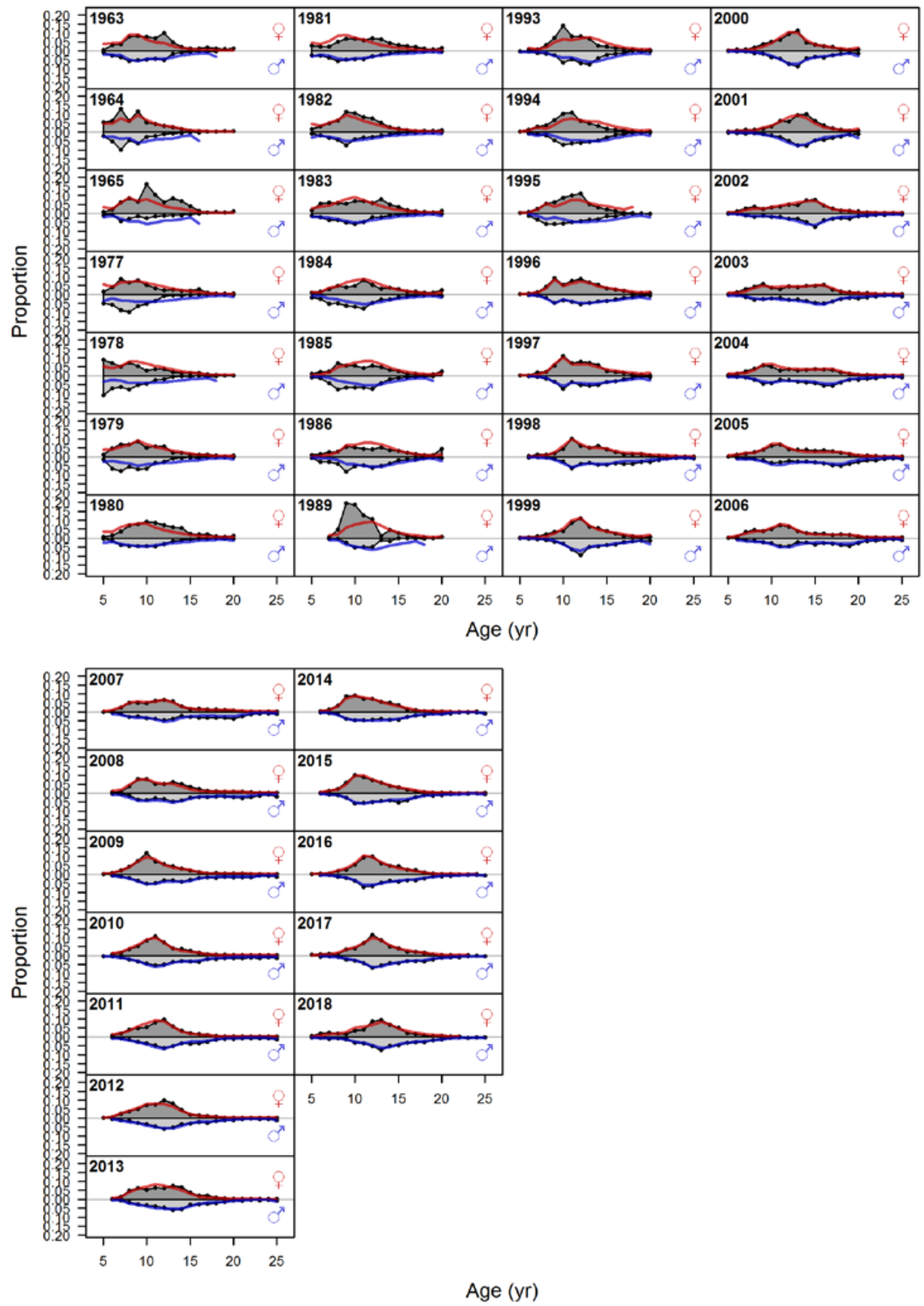

FIGURE 42. Fit to age data from the Region 3 survey in the AAF long model.

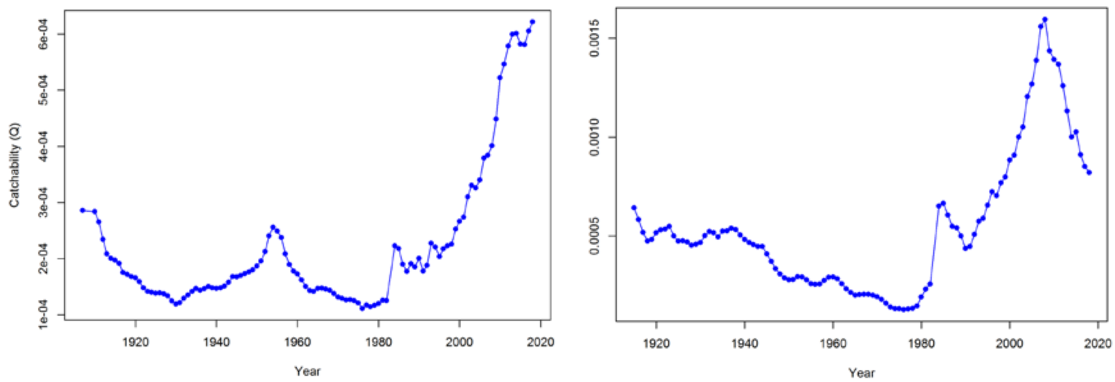

FIGURE 43. Estimated trends in the Region 2 (upper panel) and Region 3 (lower panel) commercial fishery catchability in the AAF long model.

Summary of pros and cons for the AAF long model:

Pros:

- Includes uncertainty in female natural mortality
- 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 survey age data well
- Internally consistent data weighting

Cons:

- Highest technical overhead (complexity) of the four models
- Relies heavily on only fishery trends over the historical period
- Implicitly assumes stationarity in some processes (e.g., the stock-recruitment function, natural mortality) over long historical period (beyond modelled environmental effects)
- Residual patterns in Region 4 fishery age data
- Fits Region 2 and 3 survey age data poorly

Sources of uncertainty

The four models evaluated here represent 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 is also embedded in the differences observed among the model results. These factors lead to differences in both scale and trend. In aggregate, the four models together reflected much more uncertainty than any single model, while still showing a similar basic trend over the recent time-series' of 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. These results were selected to try to highlight the features of each of the four models to which there appeared to be the strongest response in stock size and trend estimates, or to illustrate the effect of specific model features of specific interest.

The large differences in the scale of the spawning biomass in the historical period between the two long time series models represent the range of assumptions about the connectivity of the stock via spatial availability (FIGURE 44). Specifically, domed selectivity for Region 2 and Region 3 in the long AAF model implicitly assumes that older fish (located in northern and western areas) were historically less available and therefore less mobile. 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 of these 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.

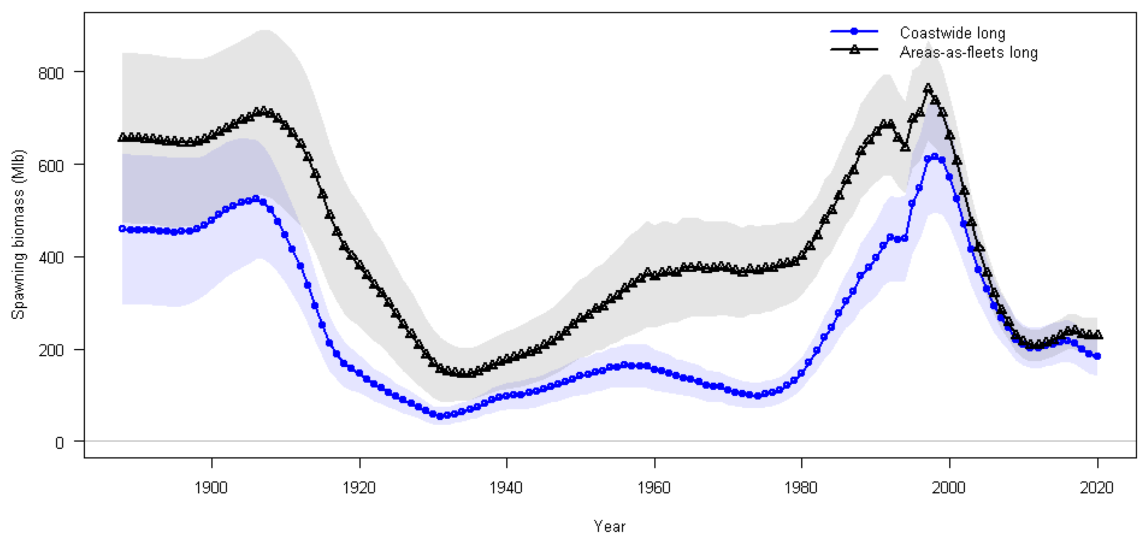

FIGURE 44. Comparison of the spawning biomass for the long coastwide and AAF models.

Steepness of the stock-recruitment relationship was fixed at a value of 0.75 for all four models. Exploratory model runs revealed that when estimated, steepness was either very imprecisely informed or maximum likelihood values occurred at the upper bound of 1.0. The effects of estimating steepness on the spawning biomass time series varied among the four models. The two short time-series models showed little difference in the estimated time series when steepness was estimated (FIGURE 45), likely due to the flexible initial conditions and the full information content of the entire series directly informing all recruitment deviations. The long AAF model also showed little difference when steepness was estimated (FIGURE 46), and was the only model where steepness did not go to a value of 1.0 (however the 95% interval did contain 1.0). In contrast, the coastwide long model showed an increase in the scale of both the spawning biomass and recruitment estimates across the entire time-series when steepness was estimated (FIGURE 47). This is likely due to an interaction between the very low relative stock sizes estimated during the historical period and the relatively small value of the σ constraining recruitment deviations (0.55; TABLE 9), and the higher estimated natural mortality in this model. This sensitivity was not investigated further to determine whether retuning the recruitment σ would result in a smaller difference in the overall results.

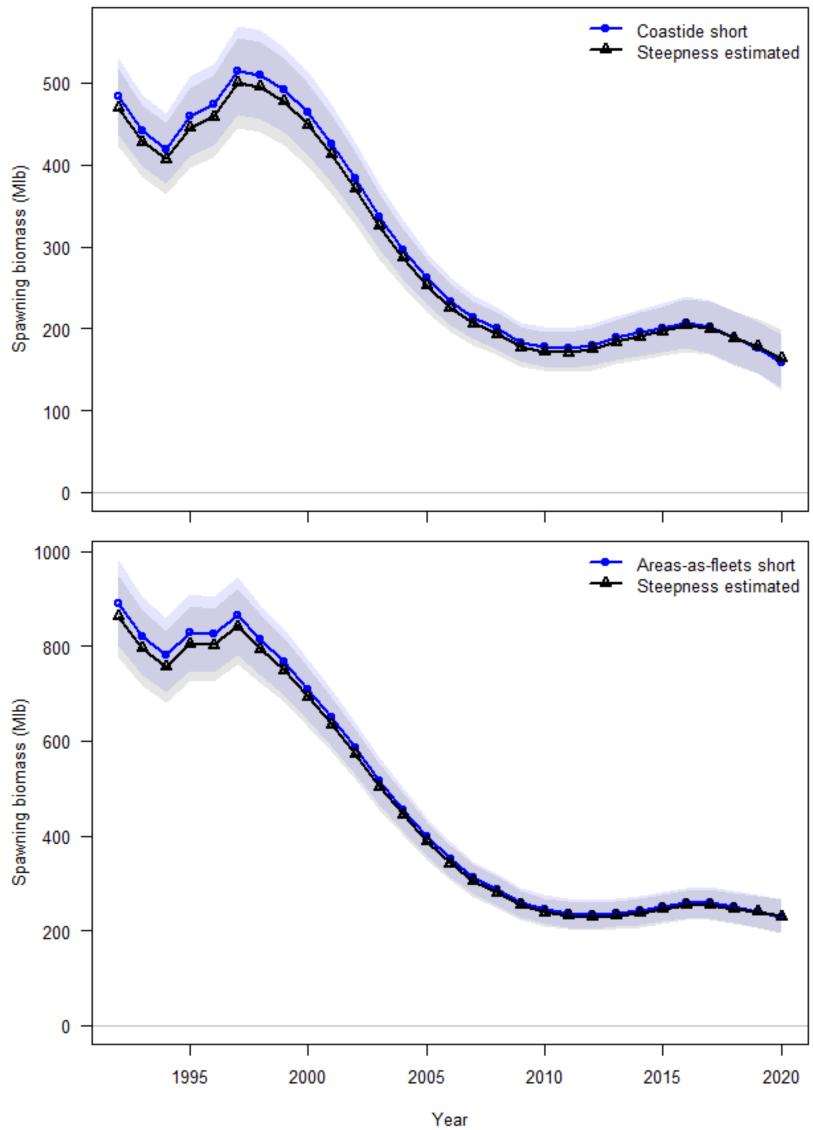

FIGURE 45. Effect of freely estimating steepness (*h*) in the coastwide short model (upper panel) and in the AAF short model (lower panel).

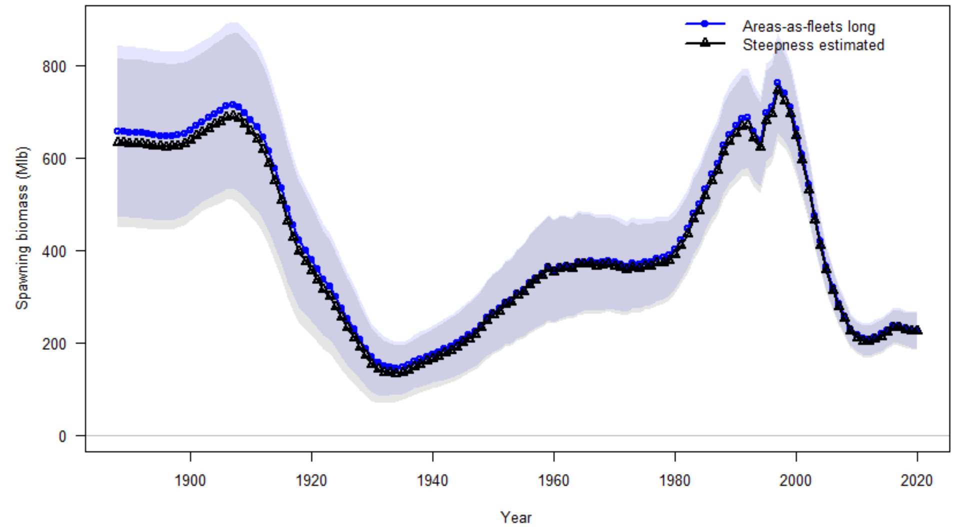

FIGURE 46. Effect of freely estimating steepness (*h*) in the AAF long model.

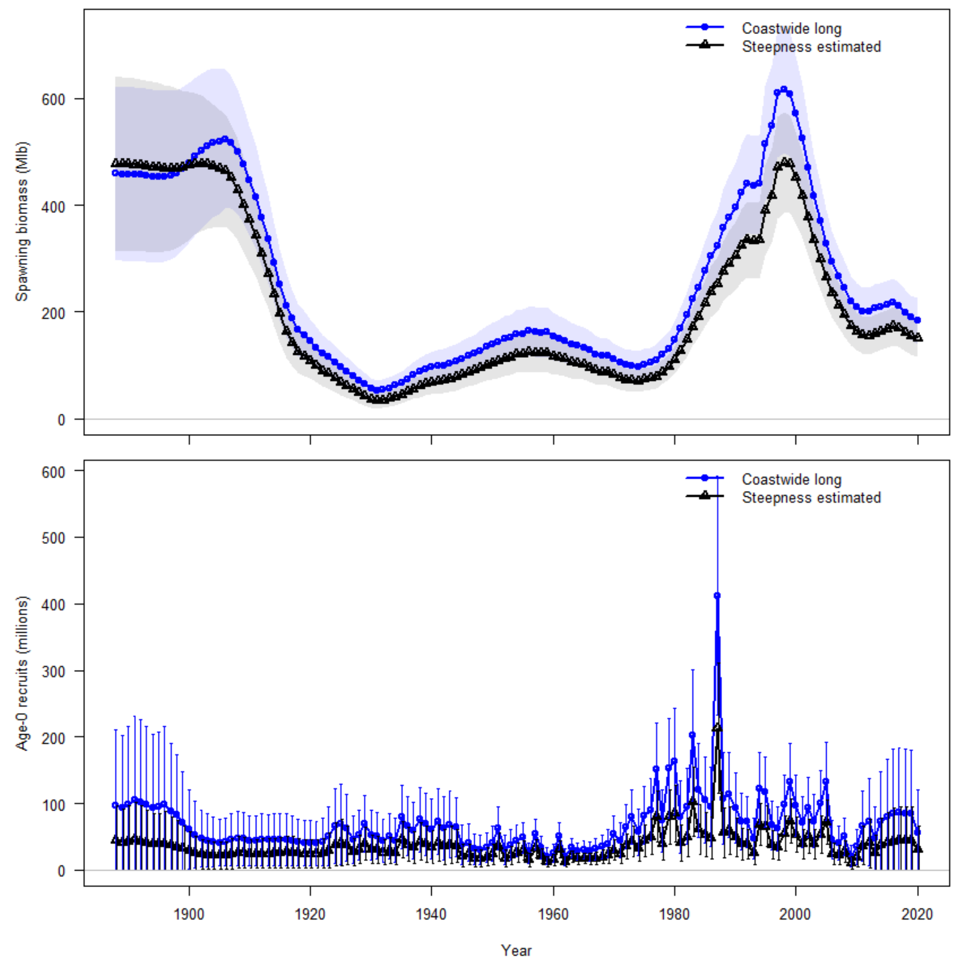

FIGURE 47. Effect of freely estimating steepness (*h*) on spawning biomass (upper panel) and recruitment estimates (lower panel) in the coastwide long model.

Female natural mortality (*M*) is fixed at 0.15 in the two short time-series models, representing a very strong assumption about the scale and productivity of the estimated population. In exploratory analyses, the values of female and male natural mortality were not jointly estimable with only the short time-series of data to inform them given the other estimated processes in these models. To evaluate the degree of uncertainty missing from these models, lower and higher values were constructed based approximately on the width of the intervals from the two long models where female natural mortality is freely estimated. Centered on the fixed value of 0.15, models with a lower value of 0.13 and a higher value of 0.17 were run. The results were consistent with previous sensitivity analyses: female natural mortality is a direct scalar on the scale of spawning biomass and recruitment in both the coastwide and AAF short models (FIGURE 48-49). Higher values of natural mortality corresponded to larger stock sizes and age-0 recruitment estimates; however, the trends in both series were nearly identical to those from the model assuming female natural mortality was 0.15.

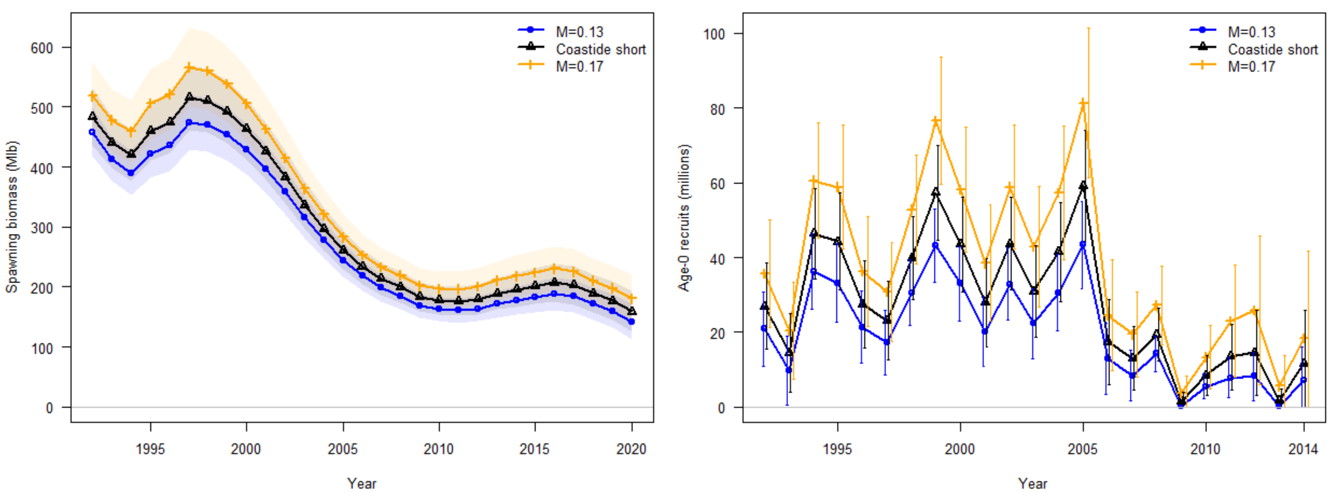

FIGURE 48. Effect of alternative fixed values of natural mortality relative to the base value (0.15) in the coastwide short model.

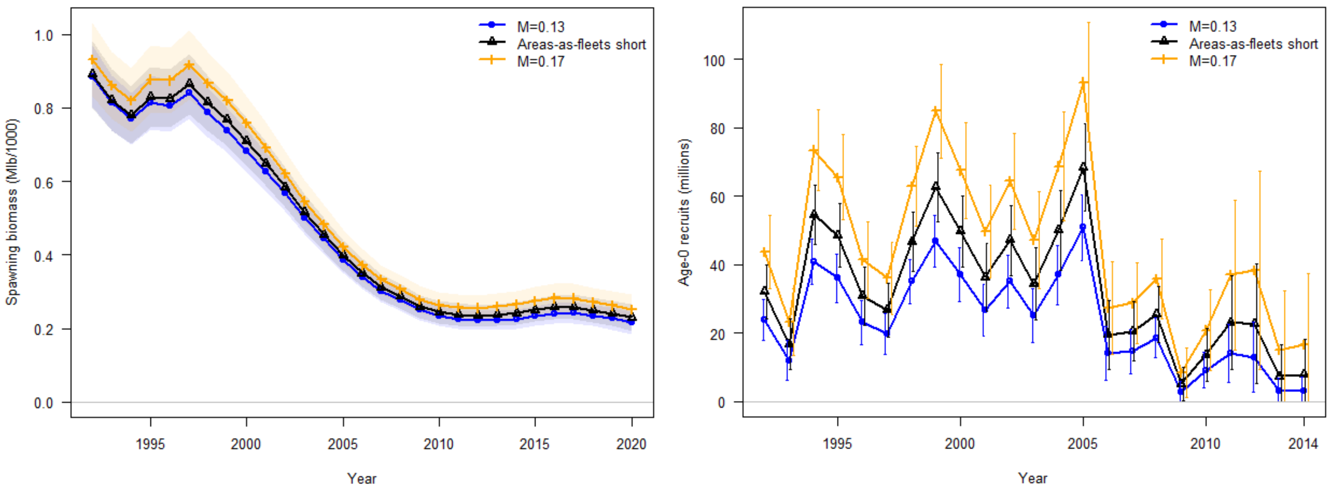

FIGURE 49. Effect of alternative fixed values of natural mortality relative to the base value (0.15) in the AAF short model.

For each of the models where one or more sources of age data were relatively weighted much less than the others a sensitivity was conducted to determine if this weighting consistently effected the biomass in one direction and how strongly. To conduct these sensitivity analyses, the lowest weighted age data (depending on the model) was increased to be roughly consistent in input sample size with other sources without making any other changes (i.e., retuning process and observation error) in that model. Increasing the weight of the commercial fishery age data in the coastwide short model led to a reduction in the scale of the estimated spawning biomass (FIGURE 50). Conversely, increasing the weight of the survey age data had a positive effect on the scale and trend of the spawning biomass time-series in both the AAF short (FIGURE 51) and long models (FIGURE 52). These results suggest that it may be worthwhile to explore reparameterizing the selectivity curves and process error (time-varying parameters) for fleets receiving lower weighting, in order to search for an approach that could fit these data better, but still retain internal consistency. It is unclear whether similar effects of the biomass time-series would be realized, but this sensitivity analysis underscores the importance of internally consistent data weighting, and the relative sensitivity of three of these models for Pacific halibut to the conflicting signals in the data. The degree to which these conflicting signals may be a result of unmodelled spatial processes is unknown, but there may not be a dramatic improvement using only nonspatially-explicit approach for these population dynamics.

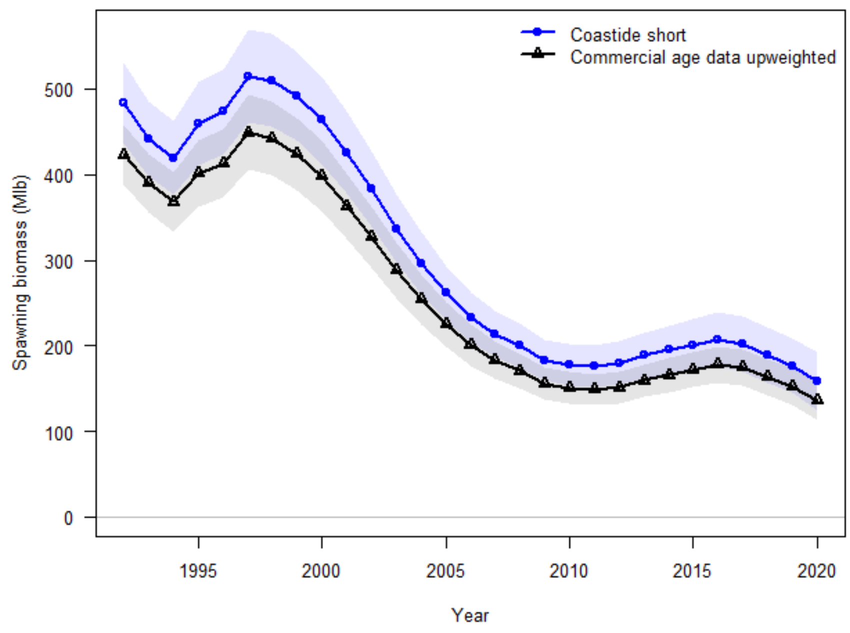

FIGURE 50. Effect of upweighting the commercial fishery age data in the coastwide short model.

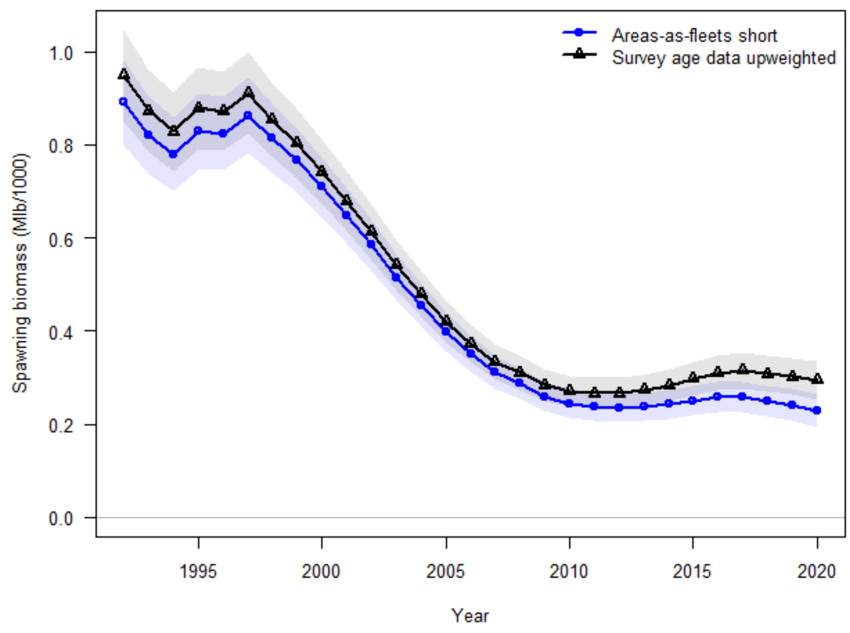

FIGURE 51. Effect of upweighting the survey age data in the AAF short model.

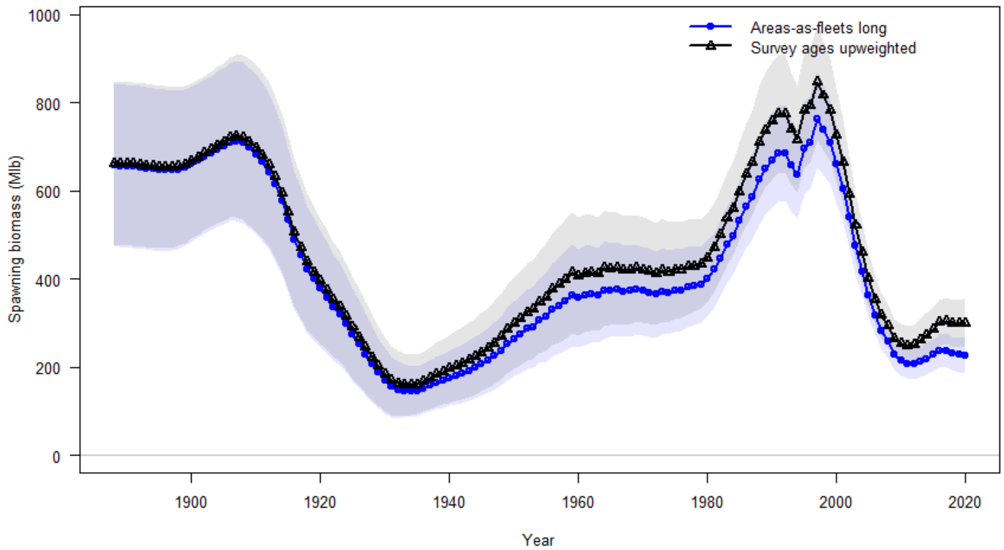

FIGURE 52. Effect of upweighting the survey age data in the AAF long model.

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 2014b; 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.

Retropective analyses were conducted for these preliminary 2019 models by sequentially removing the terminal six years from the model (a five-year retrospective, since the terminal year currently contains no information other than mortality projections). Both the coastwide and AAF short models showed variability in the scale of the spawning biomass estimates, with the only apparent trend (increasing) occurring in the AAF short model after the important sex-ratio data from the 2017 commercial fishery landings were removed in the third year of the retrospective (FIGURE 53-54). The coastwide long time series model was also sensitive to the retrospective removal of data, again particularly so after the sex-ratio data from 2017 had been removed and fits to the data and parameter estimates became unreliable (FIGURE 55). A slightly increasing trend was observed in the AAF long model, although retrospective estimates remained inside the 95% intervals from the base model (FIGURE 56).

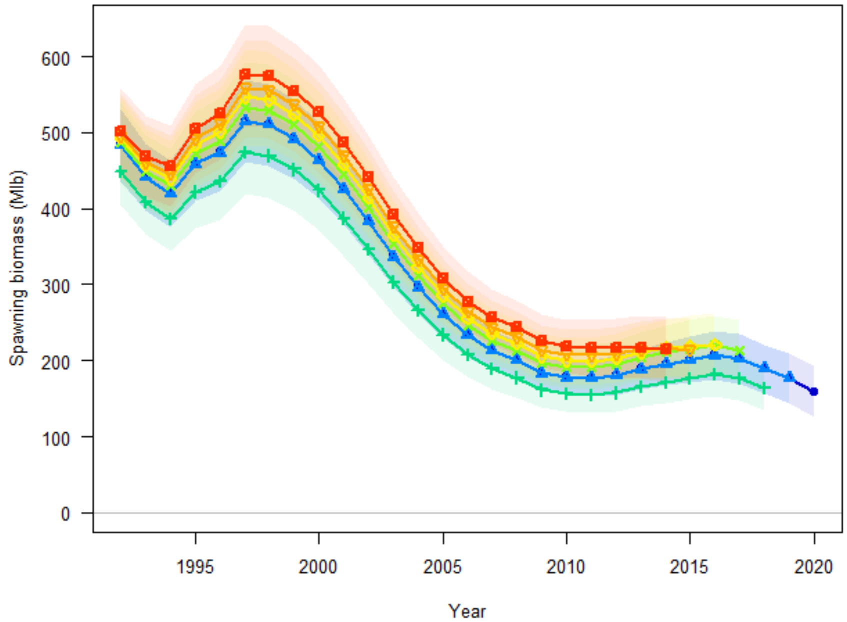

FIGURE 53. Six-year retrospective analysis of spawning biomass (1st year is a projection with no data) based on the coastwide short model.

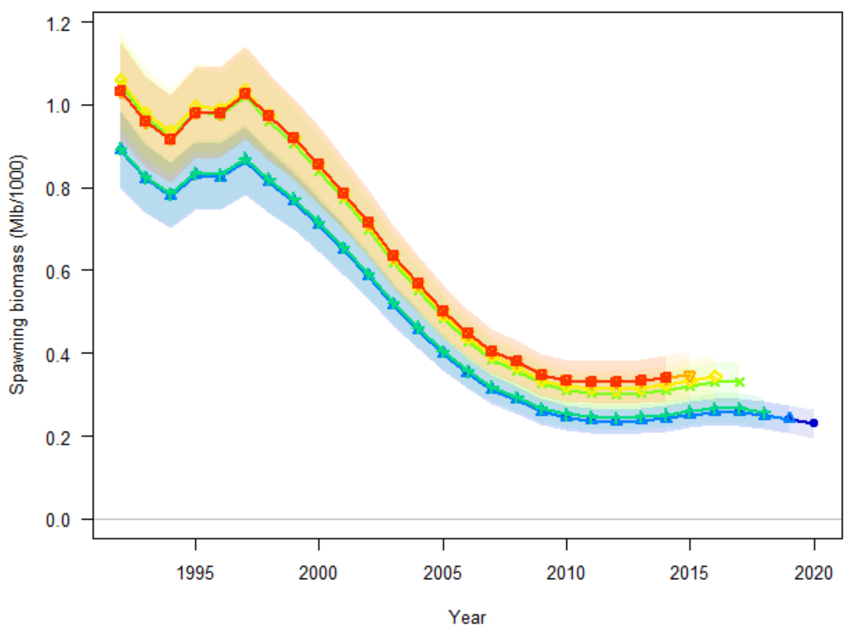

FIGURE 54. Six-year retrospective analysis of spawning biomass (1st year is a projection with no data) based on the AAF short model.

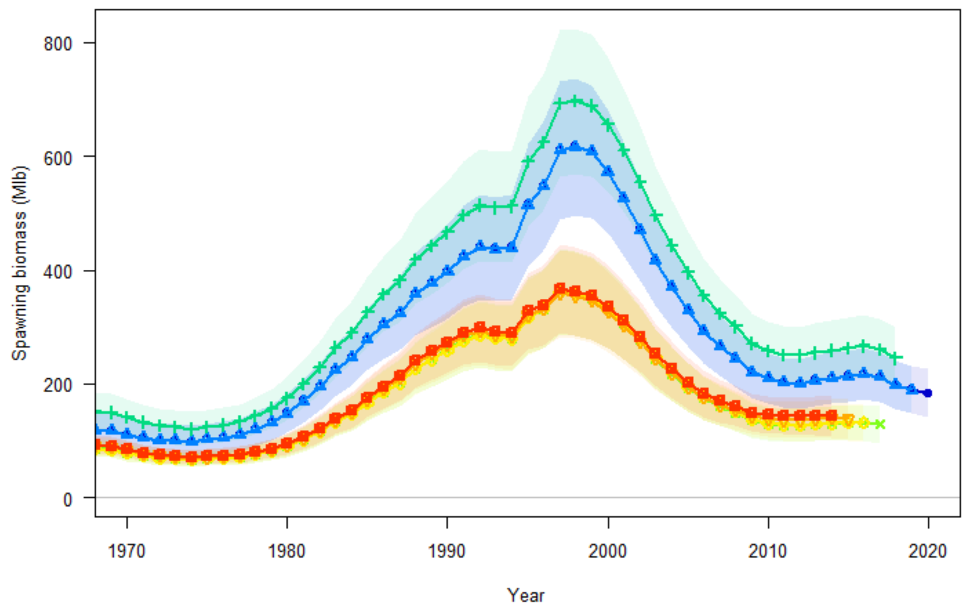

FIGURE 55. Six-year retrospective analysis of spawning biomass (1st year is a projection with no data) based on the coastwide long model.

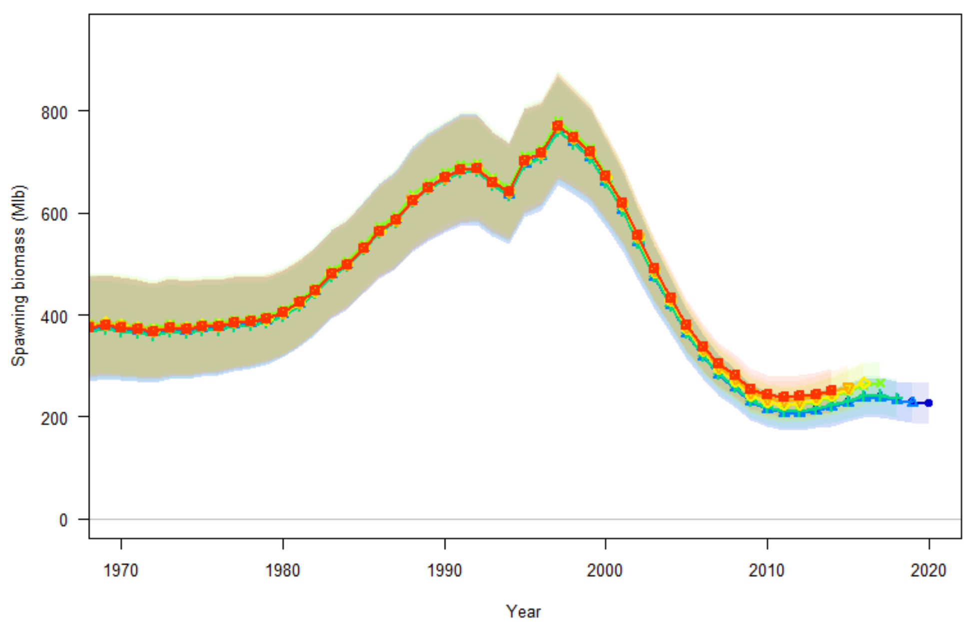

FIGURE 56. Six-year retrospective analysis of spawning biomass (1st year is a projection with no data) based on the AAF long model. Time-series is truncated in 1972 so that differences in the terminal years are more visible.

In order to better understand the interaction between the retrospective analysis and the important change in information content provided by the 2017 sex-ratio information (allowing for the estimation of the scale of male selectivity in the commercial fisheries), a second series of retrospective analyses were conducted for the short AAF, and two long time-series models. This set of retrospective analyses fixed the scale of male selectivity at the estimates from the base models, and then sequentially removed each year of data as above. The results indicated little difference in the retrospective patterns for the AAF short (FIGURE 57) and AAF long models (FIGURE 58). In contrast, the coastwide long model showed very little retrospective pattern when the scale of the male selectivity was fixed at the base estimate, illustrating the sensitivity to, and importance of this piece of information to the current stock assessment (FIGURE 59).

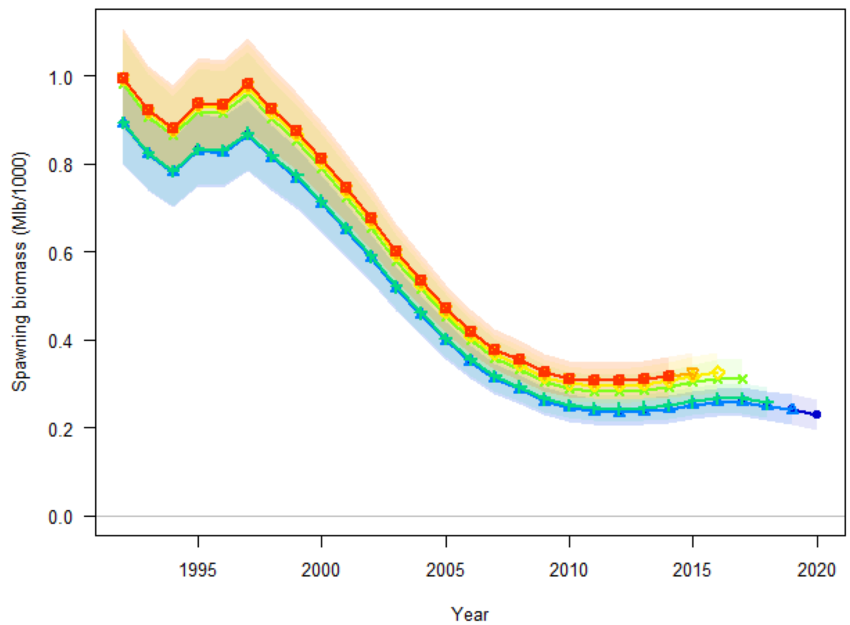

FIGURE 57. Alternative six-year retrospective analysis of spawning biomass (1st year is a projection with no data) holding commercial male selectivity scale constant based on the AAF short model.

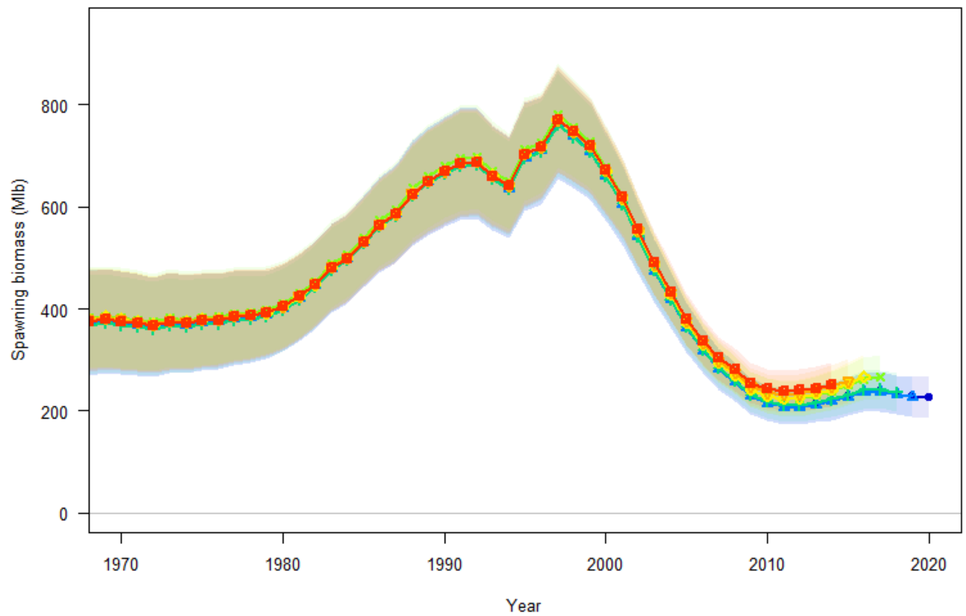

FIGURE 58. Alternative six-year retrospective analysis of spawning biomass (1st year is a projection with no data) holding commercial male selectivity scale constant based on the AAF long model.

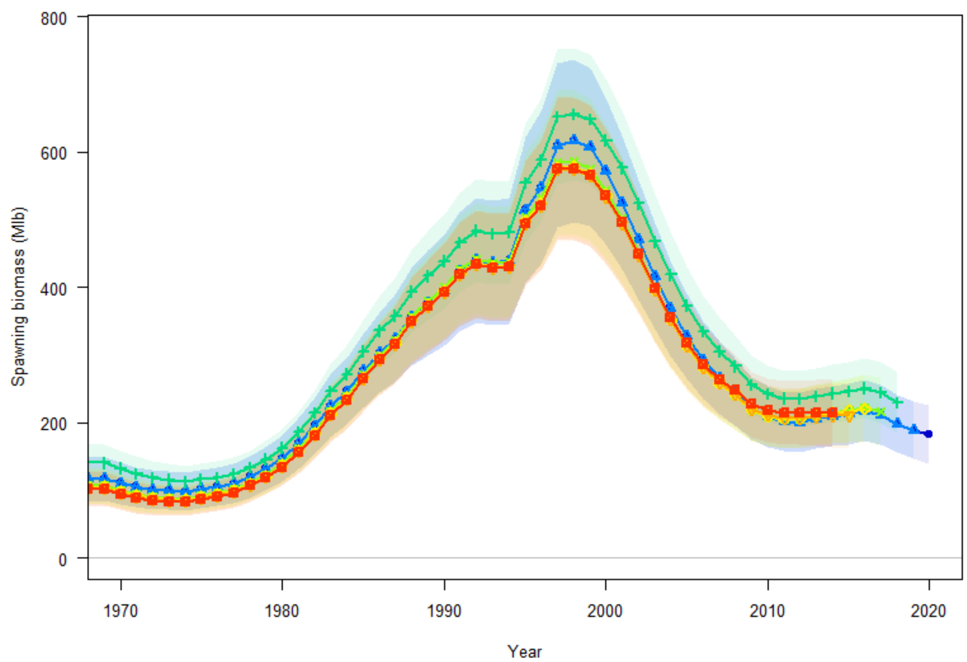

FIGURE 59. Alternative six-year retrospective analysis of spawning biomass (1st year is a projection with no data) holding commercial male selectivity scale constant based on the coastwide long model.

Bayesian analysis

Like most fisheries analyses, the models used for the Pacific halibut stock assessment have always been based on Maximum Likelihood Estimates (MLEs) and asymptotic approximations to the variance about these estimates (Fournier et al. 2012). 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 in complex fisheries models (Stewart et al. 2013b).

Recent work by Cole Monnahan, who contributed to the 2015 stock assessment (Stewart et al. 2016), has demonstrated the potential for new methods to dramatically increase the computational efficiency of Bayesian models implemented in AD Model builder (Monnahan and Kristensen 2018). Similar, but not identical, results were reported for a regularized and simplified Bayesian assessment model based on the 2015 coastwide short Pacific halibut model as part of a larger evaluation of these new methods (Monnahan et al. 2019).

Previous reviews have not placed a high priority on the refinement of the models contributing to the Pacific halibut stock assessment toward a fully Bayesian implementation, but have noted some interest. For this preliminary assessment, we investigated the coastwide short time series model (the fastest running of the four) in a Bayesian context. We followed the iterative approach suggested by Monnahan et al. (2019; https://github.com/colemonnahan/bayes_assess/blob/master/demo.R) of first identifying highly correlated parameters with slow mixing during short pilot chains using the Random Walk Metropolis (RWM) algorithm in AD Model Builder, and then simplifying the model to reduce these posterior correlations. After this initial regularization, a two-step approach was used to run several parallel chains of the No-U-Turn-Sampler (NUTS) algorithm based first on the Hessian created during minimization, and then re-running longer parallel NUTS chains using a mass-matrix updated by the earlier run. Results were integrated using the ‘sample_admb’ function in R and diagnosed using the ‘launch_shinyadmb’ function.

For the coastwide short model, a small number of selectivity parameters (primarily highly correlated deviations) were removed from the model during regularization which had a very small (<3%) effect on the maximum likelihood estimate of spawning biomass, but dramatically sped up the mixing of the posterior sampling chains. Computation time for the iterative approach was approximately: 70 seconds for minimization and calculation of the Hessian matrix, 20 minutes per RWM chain, 12-14 hours for each set of preliminary NUTS chains, and 3-4 days for final NUTS chains. To compare with maximum likelihood results, seven parallel chains were run for 3000 iterations each. There were no divergences, and despite remaining parameter correlations and broad (weakly informed) posteriors for some deviation and selectivity parameters, the effective sample size was 1,217, and maximum ‘Rhat’ was 1.004 for the least well-mixed parameter. Therefore, these results appeared sufficient to draw inference on parameter distributions and quantities of management interest. These results were obtained much more quickly than previous attempts using the RWM algorithm, which still showed very low effective sample sizes after days of integration.

The short coastwide model results indicated that posterior distributions for primary scaling parameters were very close to maximum likelihood estimates (FIGURE 60). Both the female spawning biomass time-series and the recruitment time-series posteriors were also nearly identical to the asymptotic distributions with only a very slight asymmetry in the uncertainty intervals (FIGURE 61). This result suggests that the asymptotic distributions are a reasonable approximation for the full posterior distributions in this model, and also that the process of regularizing the selectivity parameters, and removing some deviations prior to 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., Stewart and Monnahan 2017).

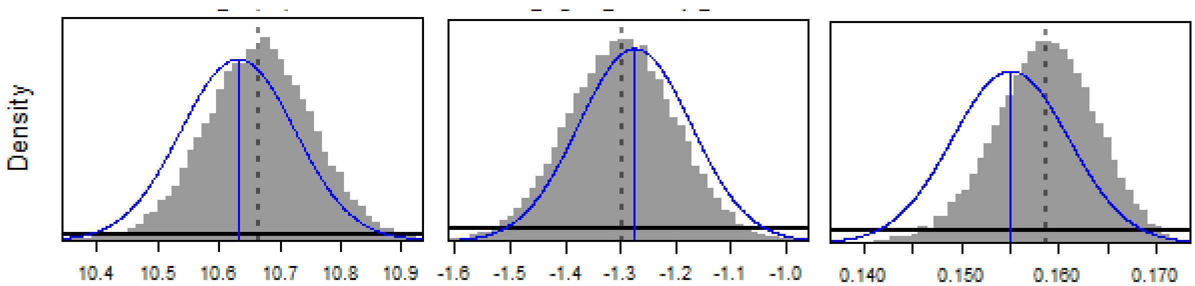

FIGURE 60. Distributions for select model parameters for the coastwide short model: ln(R0) (left panel), initial offset to ln(R0) (center panel), male natural mortality (right panel). Dark horizontal line represents the prior likelihood, symmetric distribution the MLE (vertical line) and asymptotic distribution, and histogram the posterior distribution with dashed line indicating the median.

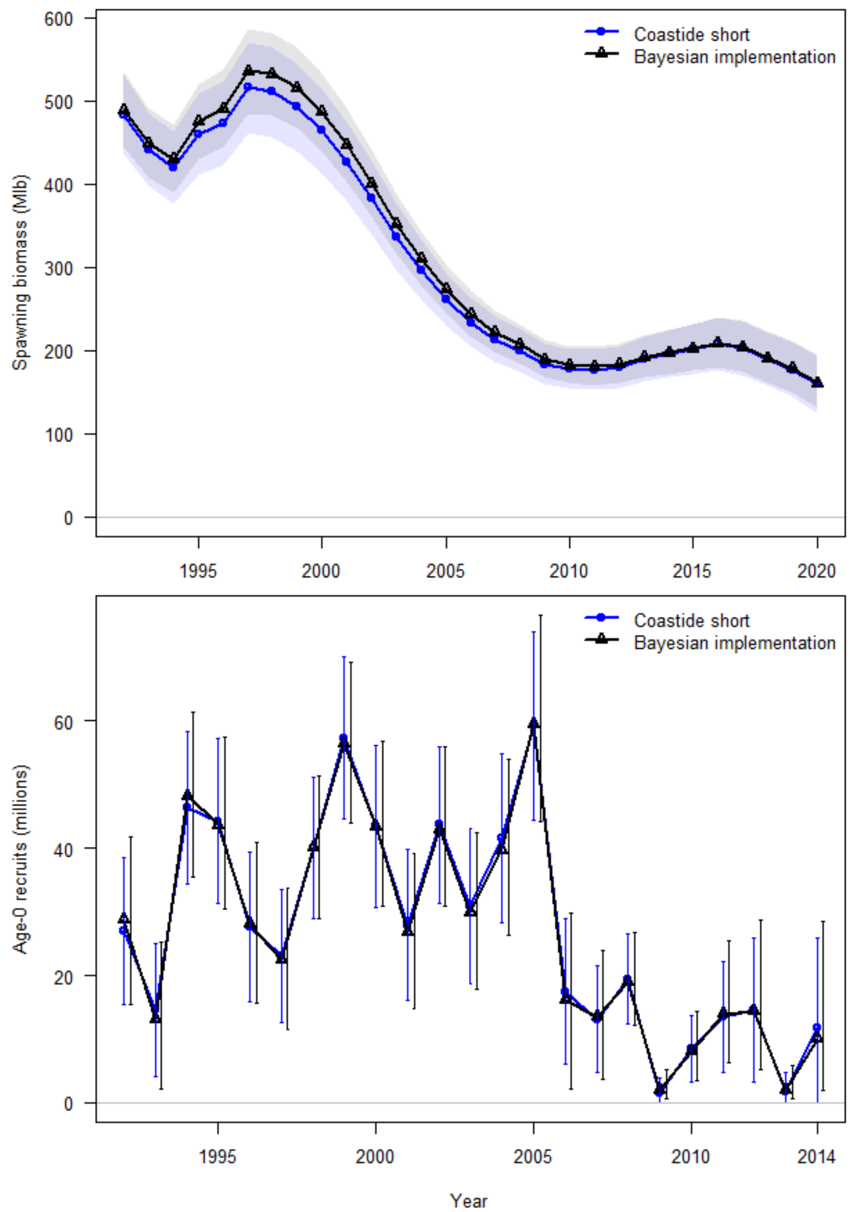

FIGURE 61. Comparison of the coastwide short model maximum likelihood estimates and asymptotic intervals to Bayesian posteriors from a regularized version of the model.

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, age-dependent 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 future weight-at-age (although the sensitivity of this was investigated at SRB request in 2016 and found to be low), future selectivity, and projected mortality.

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 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 output for stock assessment results, as well as the basis for decision table probabilities (Stewart and Hicks 2019). 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. For the results presented below n_m for all models was set equal to five million, this generated equal weight for each model and was found to be sufficient to create extremely smooth distributions, with little to no sign of Monte-Carlo error even in the extreme tails of the distributions. Although this choice could potentially be optimized, current integration code (in R) takes only seconds to run, and does not represent a constraining step in the analysis. For each element in the vector a random normal value with mean and standard deviation equal to the estimates from that model was created. Summary statistics for the integrated distribution were saved for reporting and plotting.

Since the 2005 assessment, the IPHC has transitioned to using SPR as the primary metric to measure fishing intensity, and as the basis for the harvest control rule (Hicks and Stewart 2019). Similar to spawning biomass, SPR is a direct output quantity from stock synthesis including a variance estimate for each year. Thus, random draws can be created without additional inputs other than the model report files.

Previous calculation of relative spawning biomass for use as the reference points in the IPHC's harvest control rule was structured to match the assumptions of the IPHC's harvest strategy as closely as possible. 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 $SB_{20\%}$. Historically, relative biomass was defined relative to poor recruitment conditions and relatively good size-at-age, and the constants defining these conditions were fixed (no variance) and had been estimated through historical analyses that could no longer be recreated (Clark and Hare 2002b, 2006). These reference points could be approximated in the long time series models, but relied on fixed constants in the short time-series models (Stewart and Martell 2016).

For consistency with current MSE analyses informing management decisions about the scale of coastwide mortality (Hicks and Stewart 2019), and better propagation of variance, the calculation of relative spawning biomass is updated in this assessment. The IPHC's ongoing MSE and other research has highlighted the value of dynamic reference points (representing current, rather than average or period-specific historical conditions) when strong direction shifts in productivity occur (Berger 2018). The dynamic estimate of 'unfished' biomass is calculated for each year of the time-series in stock synthesis. 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.

The only challenge to using the dynamic unfished biomass in the calculation of status and reference points is that it is not currently calculated as an 'sd_report' variable in stock synthesis, and thus has no variance or covariance associated with the point estimate. Therefore, for all relative spawning biomass calculations as simple approximation was used that included the estimates and variance of the estimates of spawning biomass in each year and the point estimate of dynamic unfished biomass in each year. The approach can be summarized in the following steps:

1) Extract the estimate and variance of spawning biomass in each year.
2) Convert the quantities in (1) to a coefficient of variation (CV).
3) Use CV of spawning biomass as a proxy for CV of the dynamic unfished spawning biomass, thereby accounting for the scale difference in the two quantities.
4) Assume a correlation between spawning biomass and dynamic unfished spawning biomass of 0.75.
5) Simultaneously draw random correlated multivariate normal values for spawning biomass and unfished dynamic spawning biomass in each year in order to calculate relative spawning biomass (the ratio of the two).

Two avenues were explored in order to evaluated how appropriate this assumption was, and how sensitive is was to alternative levels of correlation. The first of these was to make a general comparison with the results of the current MSE operating model where both quantities are simulated over many years. The MSE operating models indicated that correlations of around 0.75 were reasonable. Second, results were recalculated under differing assumptions of correlation ranging from 0.35 to 0.95. For current relative spawning biomass in the preliminary 2019 ensemble there was only a 1% difference observed in the median estimate, and less than 5% difference in the tails corresponding to the 95% interval. This calculation includes at least an approximation for more components of the variance in relative spawning biomass than previous methods, and appears to provide a reasonable proxy until revisions can be made to stock synthesis to extract the variance and covariance of the dynamic unfished spawning biomass and estimated spawning biomass for each year.

To calculate the ratio of projected future spawning biomass estimates to current values (e.g., spawning biomass current vs. spawning biomass three years in the future), conditioned on alternative input projected catch streams, both the variance and covariance estimates are directly available. The correlation is included in the calculation of this ratio as well: instead of drawing a vector of independent random normal values for each spawning biomass, the draws are multivariate normal, including the estimated covariance. The decision table also includes a metric reporting the probability that the harvest rate in the upcoming year will exceed to target harvest rate using the estimates and projections of SPR. The ratio of the projected harvest rate to the target rate (modified by the median spawning biomass relative to the $SB_{30\%}$ and $SB_{20\%}$ references points via the 30:20 control rule) is then computed. The proportion of values greater than 1.0 thus represents the probability of exceeding the target. The remaining model-integrated results are the fishery trend metrics. These report the probability that applying the current harvest policy in a future year (one and three years hence) would result in a lower TCEY (Total Constant Exploitation Yield; essentially the mortality limit set by the IPHC each year including all sources except bycatch of small, <26 inch, fish) than the value specified for that row of the decision table. This calculation first creates a distribution of SPRs, then finds the target harvest rate accounting for the spawning biomass relative to the harvest control rule and creates a distribution of future TCEYs.

Preliminary results for 2019

Comparison of the 2020 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 (TABLE 13, FIGURE 62). However, these differences are much smaller than the uncertainty in historical biomass levels in the 1990s (FIGURE 63). Recent recruitment time-series clearly reflect the differences in the various estimates and fixed values of natural mortality, but show very similar relative trends across all four individual models (FIGURE 64).

TABLE 13. Summary of individual model and ensemble distributions for 2020 spawning biomass (millions of pounds).

	Percentile		
Model	2.5%	50%	97.5%
Coastwide Long	141	184	227
Coastwide Short	125	159	193
AAF Long	185	227	269
AAF Short	194	230	265
Ensemble	135	203	261

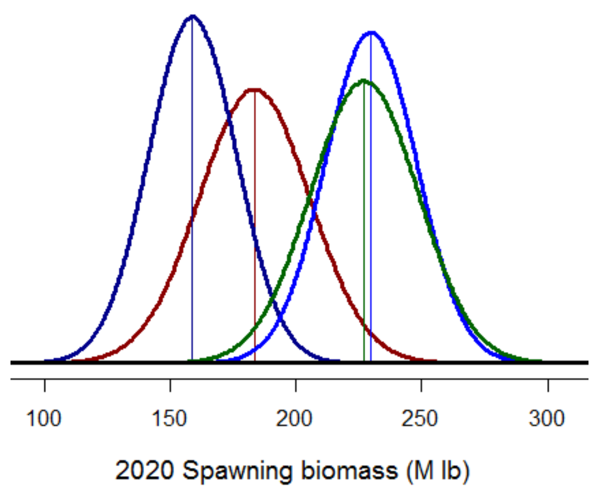

FIGURE 62. Comparison of 2020 spawning biomass distributions (asymptotic approximations) from each of the preliminary models contributing to the 2019 ensemble.

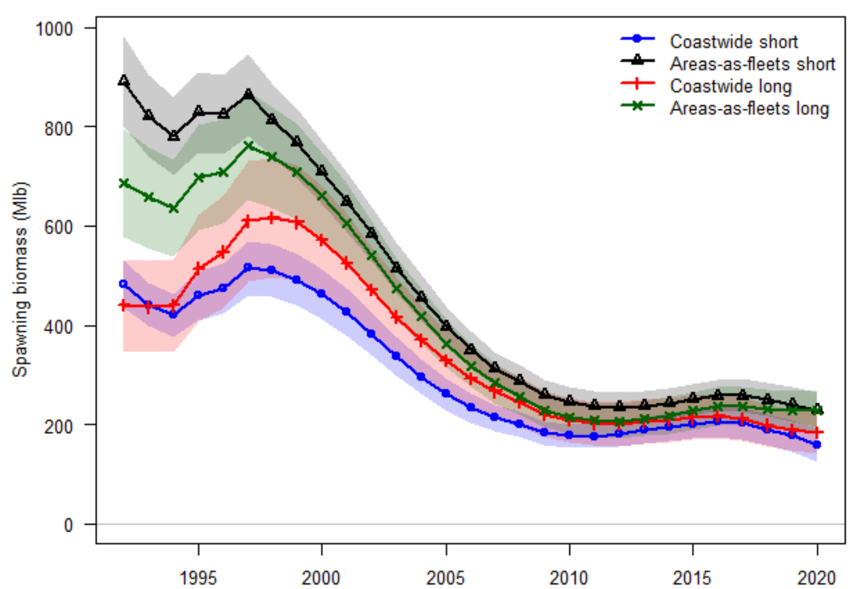

FIGURE 63. Comparison of spawning biomass time series (shaded regions indicate asymptotic approximations to the 95% confidence interval) from each of the preliminary models contributing to the 2019 ensemble.

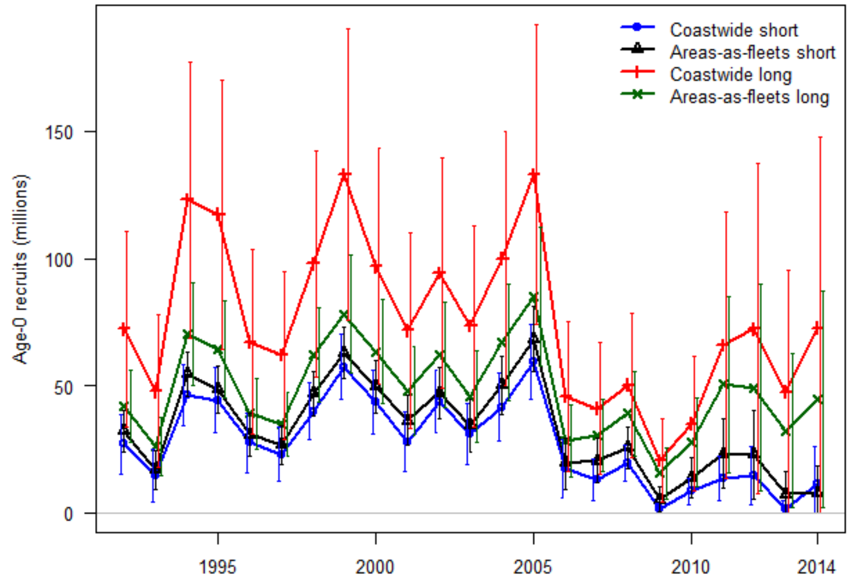

FIGURE 64. Comparison of recruitment time series (vertical lines indicate asymptotic approximations to the 95% confidence interval) from each of the preliminary models contributing to the 2019 ensemble.

As in recent assessments, the four stock assessment models comprising the ensemble are equally weighted. Comparison of ensemble results for the time-series of spawning biomass with recent stock assessment indicates a slightly larger median spawning biomass at the end of the series, but a very similar trend to previous results (TABLE 14, FIGURE 66). Fishing intensity (via SPR) is estimated to be somewhat higher since 2003 (TABLE 14). Because the mortality inputs to the assessment models have not changed and the biomass is larger, this clearly illustrates the effect of an increased fraction of females in the commercial landings, and therefore a greater effect of the lifetime spawning output of the stock.

The median relative spawning biomass (as calculated above) at the beginning of 2020 was estimated to be 31% (95% interval from 20-44%), with a probability of being below $SB_{30\%}$ of 44%, and a probability of being below $SB_{20\%}$ of 2%. Given the change in the calculation of these reference points from the fixed historical inputs to the dynamic calculation, a series of comparisons were made in order to clearly determine how much of the change in status from the 2018 assessment was due to the additional year of projection, the calculation methods, and the new data and updated models. The following reference points were constructed from the 2018 stock assessment and the preliminary 2019 results:

- From the 2018 stock assessment: median relative biomass in 2019 (based on the previous reference points) was estimated to be 43% (95% interval from 27-63%), with a probability of being below $SB_{30\%}$ of 11%, and a probability of being below $SB_{20\%}$ of <1%.
- Extending the 2018 stock assessment assessment time series, but not making any changes to the data or calculations: median relative biomass in 2020 (based on the previous reference points) was estimated to be 38% (95% interval from 22-51%), with a probability of being below $SB_{30\%}$ of 25%, and a probability of being below $SB_{20\%}$ of <1%.

- After updating the assessment to the preliminary 2019 configuration: median relative biomass in 2019 (based on the updated calculations) was estimated to be 32% (95% interval from 23-44%), with a probability of being below $SB_{30\%}$ of 38%, and a probability of being below $SB_{20\%}$ of <1%.

Thus, a portion of the change in status is due to the change in reference points, but the majority of the change (7% of the 12%) is due to the addition of new data and updating of the individual models comprising the ensemble. The considerable uncertainty in these estimates leads to overlapping confidence intervals in all reference point comparisons.

TABLE 14. Summary of ensemble distributions from the 2018 stock assessment and this preliminary analysis.

	2018 assessment				2019 preliminary			
Year	Spawning biomass (Mlb)	95% interval (Mlb)	Fishing intensity ($F_{XX\%}$)	95% interval	Spawning biomass (Mlb)	95% interval (Mlb)	Fishing intensity ($F_{XX\%}$)	95% interval
1992	NA	NA	NA	NA	555	380-950	44%	30-54%
1993	NA	NA	NA	NA	541	376-875	44%	29-54%
1994	NA	NA	NA	NA	535	374-831	45%	30-55%
1995	NA	NA	NA	NA	607	420-882	53%	37-63%
1996	503	398-737	51%	37-66%	632	437-877	52%	37-63%
1997	546	432-762	45%	32-62%	690	477-918	46%	32-58%
1998	543	424-727	43%	30-61%	682	474-864	44%	31-57%
1999	530	406-681	41%	29-60%	663	457-815	42%	30-56%
2000	500	377-633	41%	29-60%	621	430-755	41%	31-56%
2001	461	344-580	38%	28-58%	570	394-691	38%	29-53%
2002	416	307-525	34%	26-55%	510	354-622	34%	26-50%
2003	368	266-467	31%	23-52%	449	310-549	30%	24-46%
2004	327	233-417	28%	22-49%	397	273-487	27%	23-43%
2005	290	204-370	26%	21-48%	348	240-426	25%	22-42%
2006	260	181-332	26%	21-48%	307	214-376	25%	21-41%
2007	238	165-302	26%	21-48%	275	196-336	24%	21-41%
2008	222	154-284	26%	21-48%	252	183-310	24%	20-41%
2009	202	140-260	27%	21-49%	225	167-281	25%	20-42%
2010	194	134-250	27%	21-49%	212	161-265	25%	20-42%
2011	190	132-246	33%	25-53%	205	158-258	29%	25-47%
2012	190	133-247	38%	27-57%	204	160-255	34%	29-51%
2013	196	139-254	41%	29-58%	210	167-258	36%	30-53%
2014	202	142-263	46%	31-61%	216	172-264	42%	33-56%
2015	208	145-275	47%	31-61%	222	176-273	42%	33-56%
2016	215	149-288	48%	31-62%	229	180-281	43%	32-57%
2017	213	144-292	48%	29-61%	227	175-282	42%	30-55%
2018	205	134-288	49%	28-62%	216	163-273	42%	29-55%
2019	199	125-287	47%	NA	209	152-266	39%	24-54%
2020	NA	NA	NA	NA	203	135-261	NA	NA

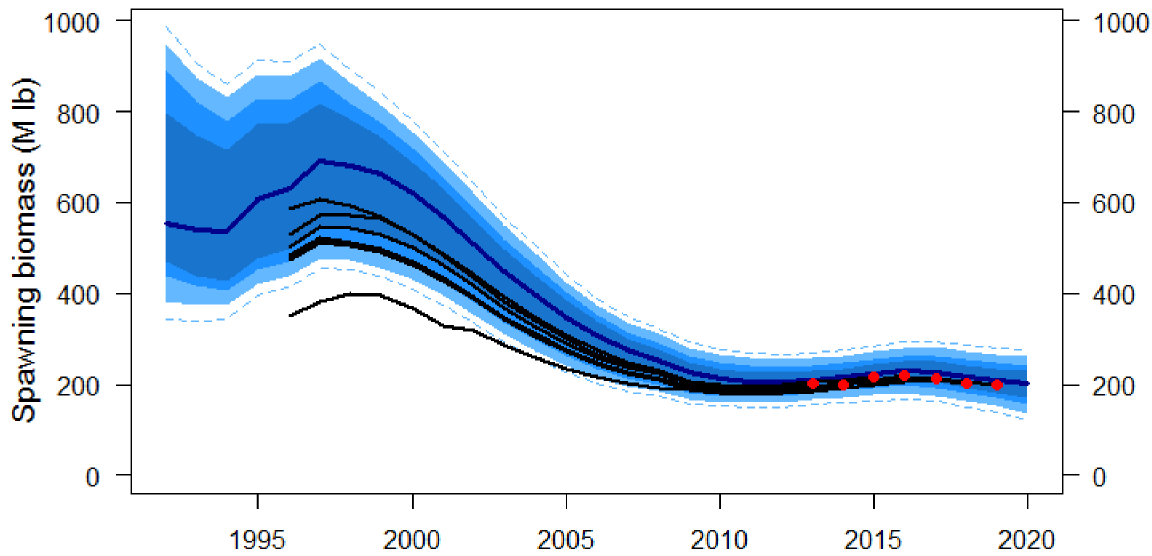

FIGURE 65. Comparison of estimated biomass time series for the preliminary 2019 ensemble (shaded region, colors indicate quantiles) and recent ensembles from 2013-2018 (black lines; red points indicate terminal estimates).

Future development

Several extensions to this preliminary assessment will be possible for the final 2019 analysis. These include:

- Responses to suggestions and comments generated from the external and SRB reviews to be conducted in June, 2019.
- Addition of all 2019 data, extending existing time series (mortality, indices, ages, etc.).
- The sex-ratio of the 2018 commercial fisheries landings may be available to be included in the final 2019 stock assessment.

In addition to the research priorities outlined below, there are potential avenues for development within and among the four models included in the ensemble.

One of these would be further investigation of the specific data sources that were relatively downweighted in one or more of the individual assessment models (see sensitivity analyses above). Alternative parameterizations of the underlying selectivity relationship and the temporally varying components of selectivity may allow for a model configuration that fits the particular data source better (less pattern in the residuals), allowing increased weighting, and perhaps improved fits to other data sources.

Other avenues for development include changes to the ensemble approach itself. Expanding the number of models included in the ensemble to better capture the uncertainty in natural mortality that is missed through using a fixed value in the two short time-series models is one such approach. Using the sensitivity analysis presented above, a comparison of the ensemble results when four additional models were added (two to each short time-series model representing higher and lower values of natural mortality). This comparison suggests that the plausible range of recent spawning biomass would be slightly wider under this expanded ensemble (FIGURE 66), but that the median value would be relatively unaffected due to the two short time-series models falling at the upper and lower ends of the range. If this approach is to be explored further, both weighting and technical efficiency should be considered. The appropriate weighting is likely to be via considering the high and low values of natural mortality for each of the short time-series models to be nested variants of a single model, and therefore each would get one-third of the weight assigned to the nested group consistent with traditional multimodel inference (Burnham and Anderson 2002). Technical costs of adding four additional models to the ensemble (doubling the number of model runs to be conducted overall) 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 doubling the ensemble in this manner beyond slightly smoother integrated distributions.

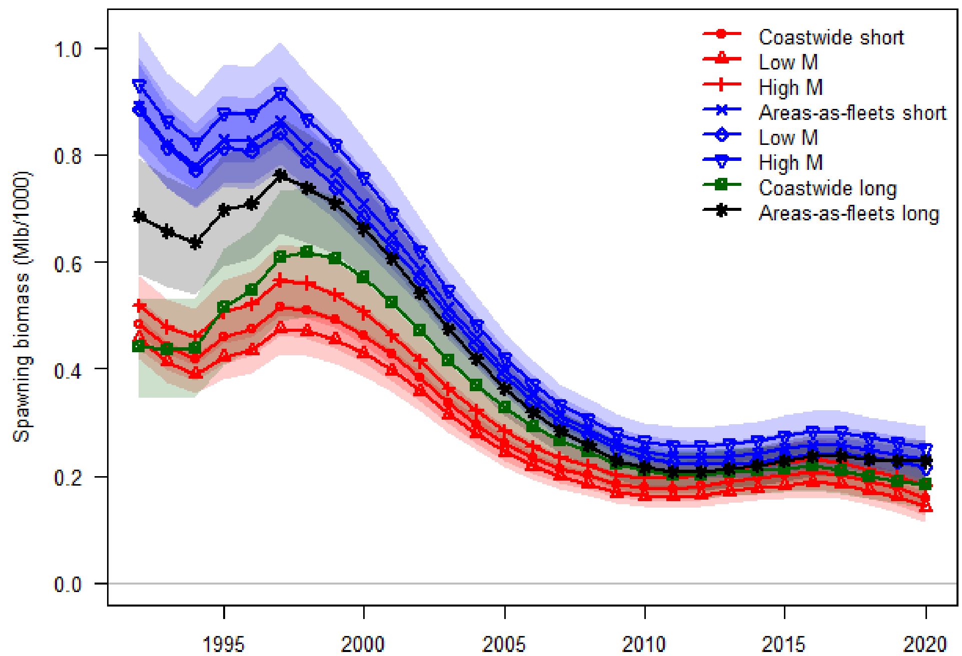

FIGURE 66. Comparison of spawning biomass time series (shaded regions indicate asymptotic approximations to the 95% confidence interval) from each of the preliminary models contributing to the 2019 ensemble with the addition of two alternative values for natural mortality for each of the short models.

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) in the absence of true random effects capability in AD Model builder. 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.

Finally, since 2015 there have been several investigations into using a revised weighting approach to the individual models contributing to the ensemble. Methods have included fit (and implied aggregate fit for the AAF models) to the coastwide survey time-series, the retrospective behaviour of each model relative to a null (simulated) distribution (Hurtado-Ferro et al. 2015), and the prospective skill of each model to predict the terminal survey index value. During 2015, 2016 and 2017, each of these methods was derived and presented to the SRB, but there was no clear support for deviating appreciably from an equal weighting approach. The benefits of such a weighting ‘rule’ could be realized if it were applied over time and annual decisions about weighting did not have to be made.

Research priorities

Research priorities for the Pacific halibut stock assessment can be delineated into three broad categories: improvements in basic biological understanding, investigation of existing data series and collection of new information, and technical development of models and modelling approaches.

Biological understanding

During the last several years, the IPHC Secretariat has developed a comprehensive five-year research program (Planas 2019). The development of the 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. Key areas for improvement in biological understanding include:

- The current functional maturity schedule for Pacific halibut, including fecundity-weight relationships and the presence and/or rate of skip spawning.
- The stock structure of the Pacific halibut population. Specifically, whether any geographical components (e.g., Region 4B) are isolated to a degree that modelling approximations would be improved by treating those components separately in the demographic equations and management decision-making process.
- Movement rates among Biological Regions 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.
- 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.

- 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.
- Improved understanding of discard mortality rates and the factors contributing to them may reduce potential biases in mortality estimates used for stock assessment.

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.

- 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. Potential methods for estimating historical sex-ratios from archived scales, otoliths or other samples should be pursued if possible.
- The work of Monnahan and Stewart (2015) modelling commercial fishery catch rates has been extended to include spatial effects. This could be used to provide a standardized fishery index for the recent time-series.
- A revised hook spacing relationship (Monnahan and Stewart 2017) will be investigated for inclusion into IPHC database processing algorithms.
- Reevaluation of the historical length-weight relationship to determine whether recent changes in length-at-age are also accompanied by changes in weight-at-length and how this may change estimates of removals over time is ongoing.
- A historical investigation on the factors influencing observed size-at-age, and ageing of additional samples from key periods and areas to support this analysis is ongoing at the IPHC.
- There is the potential that trawl surveys, particularly the Bering Sea trawl survey, could provide information on recruitment strengths for Pacific halibut several years prior to currently available sources of data. Geostatistical modelling and renewed investigation of the lack of historical correlation between trawl survey abundance and subsequent abundance of Pacific halibut in the FISS and directed fisheries may be helpful for this effort.
- 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.
- NMFS observer data from the directed Pacific halibut fleet in Alaska could be evaluated for use in updating DMRs 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 need to 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.
- A space-time model could be used to calculate weighted FISS 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.

Technical development

There are a variety of technical explorations and improvements that could benefit the stock assessment models and ensemble framework. Although larger changes, such as the new data sets and refinements to the models presented in this document, naturally fit into the period full assessment analyses, incremental changes may be possible during updated assessments when and if new data or methods become available. Specifically, development is intended to occur in time for initial SRB review (generally in June), with 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:

- Maintaining consistency and coordination between MSE, and stock assessment data, modelling and methodology.

- 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 downweighting of certain data sources (see section above), evaluation of additional axis of uncertainty (e.g., steepness, as explored above), or others.

- Evaluation of estimating (Thorson 2018) rather than tuning (Francis 2011; Francis 2016) the level of observation and process error in order to achieve internal consistency and better propagate uncertainty within each individual assessment model. This could include the 2d Autoregressive smoother for selectivity, the Dirichlet multinomial, and other features now implemented in stock synthesis (Methot et al. 2019).

- Continued development of weighting approaches for models included in the ensemble, potentially including fit to the survey index of abundance, retrospective, and predictive performance (see section above).

- Exploration of methods for better including uncertainty in discard mortality and bycatch estimates 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).

- Exploration of stock synthesis features previously unavailable or unevaluated including: timing of fishery and survey observations, the fishing mortality approximation used (i.e., estimated parameters, 'hybrid' or Pope's approximations)
- An analysis of model sensitivity and statistical performance of treating the environmental relationship between recruitment and the PDO as annual deviates (+/-), a running mean, or annual values (actual PDO), or other methods that differ from the binary indicator variable currently employed.
- 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

IPHC datasets comprise a wide array of sources based on extensive sampling and reporting efforts by state and national agencies in the U.S. and Canada. The IPHC's annual stock assessment benefits from the hard work of all of its current and former employees providing high-quality data sets as comprehensive as any used for fisheries analysis. The Scientific Review Board and national science advisors have provided extensive guidance and constructive criticism of the treatment of data sources, the individual models and the stock assessment ensemble. Ray Webster leads, or contributes to, many of the supporting data analyses on which the assessment is based. Cole Monnahan has contributed conceptually to the stock assessment methods employed in this assessment, as well as technically in the implementation of Bayesian algorithms in ADMB.

References

Berger, A.M. 2018. Character of temporal variability in stock productivity influences the utility of dynamic reference points. Fish. Res. ***In press***. doi:10.1016/j.fishres.2018.11.028.

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

Clark, W.G. 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. Nonparametric estimates of age misclassification from paired readings. Can. J. Fish. Aquat. Sci. **61**: 1881-1889.

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

Clark, W.G., and Hare, S.R. 2002a. 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. 2002b. Effects of climate, stock size, and region on recruitment and growth of Pacific halibut. IPHC Report of Assessment and Research Activities 2001. p. 169-190.

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.

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.

Drinan, D.P., Loher, T., and Hauser, L. 2018. Identification of Genomic Regions Associated With Sex in Pacific Halibut. The Journal of heredity **109**(3): 326-332. doi:10.1093/jhered/esx102.

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

Erikson, L. 2019. Fishery Statistics (2018). IPHC-2019-AM095-05. 23 p.

Erikson, L., Soderlund, E., and Geernaert, T. 2019. Fishery-independent setline survey (FISS) design and implementation in 2018, including current and future expansions. IPHC-2019-AM095-06. 14 p.

Erikson, L.M., and Kong, T.M. 2018. 1.3 Sampling commercial landings in 2017. IPHC Report of Assessment and Research Activities 2017. p. 9-20.

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. Can. J. Fish. Aquat. Sci. **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. Can. J. Fish. Aquat. Sci. **74**(11): 1878-1894. doi:10.1139/cjfas-2016-0290.

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. 2019. IPHC Management Strategy Evaluation (MSE): update. IPHC-2019-AM095-12. 36 p.

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 J. Mar. Sci. **72**(1): 99-110. doi:10.1093/icesjms/fsu198.

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.

IPHC. 2019. Report of the 95th Session of the IPHC Annual Meeting (AM095). Victoria, Canada, 28 January to 1 February 2019. IPHC-2019-AM095. 46 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.

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 J. Mar. Sci. **72**(1): 70-81. doi:10.1093/icesjms/fsu057.

Magnusson, A., Punt, A.E., and Hilborn, R. 2012. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fisheries: no-no. 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. Can. J. Fish. Aquat. Sci. **70**(6): 930-940.

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.

McCarthy, O. 2015. Development of a method for marking Pacific halibut by sex onboard commercial fishing vessels. IPHC Report of Assessment and Research Activities 2014. p. 435-464.

McGilliard, C.R., Punt, A.E., Methot, R.D., Hilborn, R., and Jacobson, L. 2014. Accounting for marine reserves using spatial stock assessments. Can. J. Fish. Aquat. Sci.: 1-19. doi:10.1139/cjfas-2013-0364.

Methot, R.D. 2015. User manual for Stock Synthesis. Model version 3.24s. NOAA Fisheries, Seattle, WA. 152 p.

Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability in estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. **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. Appendix A: Technical description of the Stock Synthesis assessment program. Fish. Res. **142**: 26 p.

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

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.

Monnahan, C.C., and Stewart, I.J. 2017. The effect of hook spacing on longline catch rates: Implications for catch rate standardization. Fish. Res. doi:10.1016/j.fishres.2017.10.004.

Monnahan, C.C., and Kristensen, K. 2018. No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. PLOS ONE **May 24, 2018: 1-10.**

Monnahan, C.C., Branch, T.A., Thorson, J.T., Stewart, I.J., and Szuwalski, C. 2019. Overcoming long Bayesian run times in integrated fisheries stock assessments. ICES J. Mar. Sci. **In press.** doi:10.1093/icesjms/fsz059/5475859.

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.

Planas, J. 2019. IPHC 5-year biological and ecosystem sciences research program update. IPHC-2019-AM095-14. 7 p.

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. Fish. Res. **168**: 85-99. doi:10.1016/j.fishres.2015.04.002.

Soderlund, E., Randolph, D.L., and Dykstra, C. 2012. IPHC Setline Charters 1963 through 2003., International Pacific Halibut Commission Technical Report No. 58. 264 p.

Stewart, I., and Hicks, A. 2018a. Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2017. IPHC-2018-AM094-10. 25 p.

Stewart, I., and Hicks, A. 2018b. Evaluation of the IPHC's 32" minimum size limit. IPHC-2018-AM094-14. 44 p.

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

Stewart, I., and Webster, R. 2019. Overview of data sources for the Pacific halibut stock assessment, harvest policy, and related analyses. IPHC-2019-AM095-08. 76 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 Martell, S. 2014a. Assessment of the Pacific halibut stock at the end of 2013. IPHC Report of Assessment and Research Activities 2013. p. 169-196.

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

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

Stewart, I.J., and Martell, S.J.D. 2015. Reconciling stock assessment paradigms to better inform fisheries management. ICES J. Mar. Sci. **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 Hicks, A.C. 2017. 4.2 Assessment of the Pacific halibut stock at the end of 2016. IPHC Report of Assessment and Research Activities 2016: 365-394.

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

Stewart, I.J., and Hicks, A.C. 2018c. Interannual stability from ensemble modelling. Can. J. Fish. Aquat. Sci. **75**: 2109-2113. doi:10.1139/cjfas-2018-0238.

Stewart, I.J., Monnahan, C.C., and Martell, S. 2016. Assessment of the Pacific halibut stock at the end of 2015. IPHC Report of Assessment and Research Activities 2015. p. 188-209.

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. Fish. Res. **142**(0): 37-46. doi:http://dx.doi.org/10.1016/j.fishres.2012.07.003.

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. 2018. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. Fish. Res. ***In press.*** doi:10.1016/j.fishres.2018.02.005.

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.

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. Fish. Res. **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. 2019. Space-time modelling of survey data. IPHC-2019-AM095-07. 19 p.

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

Webster, R.A., 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. Can. J. Fish. Aquat. Sci. **70**(4): 642-653. doi:10.1139/cjfas-2012-0371.

Appendices

Appendix A: Supplementary material

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

1) Input files for each of the assessment models (implemented in stock synthesis) 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 HTML files), and the full report (text) file from each model. The report file has not been annotated; content and formats can be determined from the stock synthesis user manual (Methot et al. 2019) 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. 2019). 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 2019) and the stock assessment results (Stewart and Hicks 2019) from the 2018 analysis.

5) The documentation from the 2015 full stock assessment (Stewart and Martell 2016).

6) Recent relevant IPHC manuscripts describing the history of the halibut stock assessment (Stewart and Martell 2014b), 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 2018c).

7) Additional historical stock assessment documentation 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 (2013-2018) are available (https://www.iphc.int/meetings/calendar?category=4).