Development of the 2015 stock assessment

Draft 6/9/15

For SRB review: 16-17 June, 2015

Ian Stewart
Steve Martell

Table of Contents

Summary

Creating robust, stable, and well-performing stock assessment models for the Pacific halibut stock has historically proven to be problematic due to the highly dynamic nature of the biology, distribution, and fisheries (Stewart and Martell 2014). Although recent modelling efforts have created some new alternatives, no single model yet evaluated satisfactorily approximates all aspects of the available data and scientific understanding. In 2014, an 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 used to describe the range of plausible current stock estimates. Each of these models (and many alternatives explored during development) has shown a similar historical pattern: a stock declining from the late 1990s, with several years of relative stability at the end of the time-series.

For 2015, an extensive effort was made to address many previously identified shortcomings in the input data. This included of a complete reprocessing of all inputs, and the addition of several new sources of information. Important improvements included: generating weight-at-age by geographic area (Areas 2, 3, 4, and 4B separated; Fig. 1) for the AAF models, improving the weight-at-age calculations for young halibut (< age-7) rarely encountered in the setline survey using data from NMFS trawl surveys, summarizing index variances and age composition sample sizes, particularly by area for the AAF models, adding age-information to directly inform the selectivity curves for bycatch, sport, and sublegal discard removals, and extending all age-data arrays to include ages 2-25 (instead of 6-25, used in historical analyses).

Although the basic approach to each of the four assessment models used in 2014 remains unchanged, several modeling aspects were explored more deeply than in previous analyses and improved where necessary. These improvements included: updating the constraint on recruitment deviations (σ_r) for consistency with stock-recruitment assumptions, updating relative data weighting to reduce the potential influence of outliers and strong residual patterns, and updating the constraints on time-varying parameters to better reflect degree of estimated variability. In addition, a much greater number of sensitivity analyses, alternate model configurations, and diagnostic evaluations were completed than in previous assessments. The 2014 assessment highlighted a difference in trend for the most recent years between the two aggregated coastwide models (the long and short time-series), and the two AAF models for the last several years. For this preliminary analysis, all models were extended to 2016 (using the projected removals for 2015, but no new observed data from that year) in order to better evaluate the recent differences among the four models in both estimated scale and trend.

Preliminary results were very consistent with projections from the 2014 assessment and indicate relatively flat trends in the coastwide models and slightly increasing trends in the AAF models. The terminal (2016) biomass estimate is uncertain both within each model and among models based on the integrated distribution. Looking forward, it is not clear whether addition of the 2015 data for the final assessment will begin to reconcile the differences in recent trend between the coastwide and AAF models. Several recent studies (e.g., Hurtado et al. 2014, Punt et al. 2015, Li et al. 2015, McGilliard et al. 2014) have evaluated the performance of aggregated vs. disaggregated approaches to catch-at-age modelling in the presence and absence of variability among fisheries and movement among areas within the stock. In some cases the AAF approach appears to be an improvement over aggregated methods, and in others (particularly Punt et al. 2015) it does not. The primary conclusion from simulation-based studies is that if the true underlying process is well-represented, then models tend to perform well. In the halibut assessment, it is likely that none of these models accurately represents the complex spatial

dynamics. In light of this uncertainty, the same equal weighting among models is retained for this analysis, and the preliminary integrated results provided for quantities of management interest.

Two primary uncertainties continue hinder our current understanding of the Pacific halibut resource: 1) the sex-ratio of the commercial catch (not sampled due to the dressing of fish at sea), which serves to set the scale of the estimated abundance (identified in the 2013 assessment; Stewart and Martell 2014b) in tandem with assumptions regarding natural mortality, and 2) the treatment of spatial dynamics and movement rates among regulatory areas, which are represented via the coastwide and AAF approaches, have very strong implications for the current stock trend. In addition, movement rates for adult and juvenile halibut (roughly ages 0-6, which were not well-represented in the PIT-tagging study), particularly to and from Area 4, are necessary for parameterizing a spatially explicit stock assessment. Ongoing research on these topics may help to inform our understanding of these processes in the long-term, but in the near-future it appears likely that a high degree of uncertainty in both stock scale and trend will continue to be an integral part of the annual management process.

Data sources

This section is not intended to duplicate the more detailed summary of data sources provided in Stewart (2015), but instead focuses on the improvements that have been made specifically for the 2015 assessment. Development and refinement of methods for aggregating raw data collected by individual regulatory area (Fig. 1) into larger areas that can be treated as separate fleets, as well as coastwide values, has been an ongoing effort over the last several years. Much progress was made during 2015 with remaining data processing challenges summarized below (see research priorities). This description is divided into two sections, the first dealing with the improvements to the treatment of weight-at-age, and the second with improvements to the index and age data used in the development of 2015 assessment models.

Weight-at-age

Historical halibut assessments have used various extrapolation and smoothing methods to assign weight-at-age to fish that are younger than those observed in the setline survey, which provides the most detailed source of sex-length-age information. These calculations are not critically important to the treatment of commercial fishery or survey information, as few very young fish are observed in those data sets; however, accurate depiction of the removals from other sources, such as recreational fisheries and bycatch in non-target fisheries requires representative weight-at-age for all fish captured, particularly ages 2-6. For 2015, average weight-at-age by year and sex was summarized from the NMFS trawl surveys in Alaska. Age and length data were available for all years since 1998, although mean values were somewhat variable due to limited sample sizes (Fig. 2). To reduce the effect of sampling variability (there is no easy way to account for observation error in the treatment of weight-at-age), raw values were smoothed across years within age (Fig. 3). Only a small subset of the 2014 trawl survey ages had been entered into the IPHC's databases at the time of this analysis, so values for that year are more variable at present, but will be revised to include the full datasets for the final 2015 assessment. These data were used to augment the weight-at-age inputs calculated from ages 7+ in the setline survey and commercial fishery (as described in Stewart, 2015).

A second important improvement to the treatment of weight-at-age, separating the trends by geographic areas (2, 3, 4, and 4B) was explored for use specifically in the AAF models. Due to the unknown sex-ratio of the commercial data, and the lack of comprehensive historical survey observations, a method was developed during 2013 to use the relative trends in weight-at-age observed for ages 8-16 in the commercial fishery, to scale the recent observations of sex-specific weight-at-age for fish collected by the setline survey. When the method was developed, it was applied only to data aggregated at the coastwide level. The coastwide trends among ages 8-16 in those data showed very similar historical patterns, despite differences on an absolute scale (Fig. 4), suggesting that temporal changes in relative weight-at-age have been relatively conserved across different ages. When this approach was duplicated for data by geographic area, the patterns for Areas 2 and 3 were quite similar to those observed at the coastwide level (Fig. 5), and appeared coherent enough to warrant summary into area-specific trends. When rescaled (relative to the value observed in 1997, the first year of comprehensive survey data), it is more obvious that the coastwide and Area 2 trends have been less pronounced than the very large increase in fish size observed for Area 3 from the 1950s through the 1990s (Figs. 6-7).

Summarized fishery data for Area 4 (including regulatory areas 4A and 4CDE) were available only beginning in 1945. The estimated trends for ages 8-16 in Area 4 showed a markedly different historical trend than for Areas 2 and 3, with fish not much larger during the historical period than in the early 1990s (Fig. 8). The relative scalar for Area 4 is therefore only slightly above a value of one for most of the historical period, and the smallest values occur in the most recent years (Fig. 9).

No historical data predating the setline survey were available from the commercial fishery in Area 4B. The Area 4 weight-at-age arrays were therefore used as model input for both Area 4 and Area 4B. All final input weight-at-age values used in each model can be found in the supplementary background material included with this document.

Index and age data

When the AAF models were assembled in 2014, data sets were parsed out into separate fleets, but not all fleet-specific index variances and samples sizes were derived. For the indices of abundance, current variance estimates are based on the among-sample variability within each regulatory area. In the case of survey data from Area 4CDE, this includes variability in both setline survey and NMFS trawl survey observations, but currently not the uncertainty in the calibration of the two series (Webster 2014). Combined index values by geographic area are weighted by the relative spatial bottom area in each regulatory area, and variances were summed, accounting for the square of the weights. For each geographic area, the annual index variance (σ^2) was converted to log Standard Error (SE) for model input via the relationship with each annual index value ($\bar{x}$):

$$\log(SE) = \sqrt{log\left(1 + \left(\frac{\sigma}{\bar{x}}\right)^2\right)}$$

Prior to 2001, there were individual regulatory areas and portions of regulatory areas missing from the coastwide and geographic area indices from the setline survey (Soderlund et al. 2012). To account for the associated missing variance components, the average log(SE) from the time-series with complete coverage was doubled (survey indices from 1997-2000 include a variety of calibrations for the index, but not the variance; Webster et al. 2015). For years prior to 1997 (except Area 3 in 1996) the coastwide log(SE) was doubled again to reflect the increasingly poor spatial coverage relative to that scale (both catchability and selectivity were also configured in the models to allow for the spatial changes).

Final input setline survey indices (numbers-per-unit-effort), coastwide and by geographic area, all showed a very clear increase in catch-rates associated with the switch from “J” to circle hooks in 1984 (Table 1). Surveys have been very precise, based on the limited variance components included in the estimates, with precision decreasing from Area 3 (log(SE) values around 0.03), to Area 2, and Area 4; Area 4B had the least precise survey with log(SE)s around 0.1 (Table 1).

Commercial fishery catch rates were aggregated from 1984 to the present using methods analogous to the survey data. For years prior to 1984, a log(SE) of 0.10 was assumed for all geographic areas, and the coastwide index, due to incomplete coverage of individual regulatory areas, lack of raw data (only historical summaries available), or both. As has been the case in recent analyses, unverified fishery data in the terminal year (in this case 2015) was assigned twice the observed log(*SE*) (Stewart 2015; this has been done to account for the use of unverified and incomplete logbook information during the preparation of the input data each fall, the data are complete in time for the subsequent year’s assessment). In recent years the precision of the fishery index is estimated to be similar to, or slightly better than that of the setline survey, as a function of the extensive reporting and correspondingly large sample sizes. All fishery indices showed a large increase in 1984, similar to the setline survey. Trends in recent years have been much more pronounced in the fishery indices of weight-per-unit-effort (compared to survey indices of numbers-per-unit-effort), because these include the effects of both the numbers of halibut, as well as those of trending size-at-age (Table 2).

Examination of the number of survey stations contributing age data reveals that Area 3 is sampled much more heavily than Area 2; this is expected, and approximately in proportion to the spatial extent of the each due to the uniform 10 nautical mile grid design. Area 4 generally contributes around half as many samples as Area 2, and Area 4B less than 10% of the total (Table 3). A similar situation is present for the number of fishing trips sampled for age data, with the exception that Area 2 comprises more than half of the total samples in recent years (Table 4). Prior to 1964, only summarized age data are currently available, and samples sizes are assumed to be roughly half of those in later years. There are no age data available for Area 4 prior to 1945 and for Area 4B prior to 1991.

Historical halibut assessments have included age-data delineated only for fish age-6 and greater. For 2015, the age-arrays for all input data were extended to include ages 2-25, with age-2 including all observations age-2 and below (a ‘minus group’) and age-25 including all observations age-25 and greater (a ‘plus group’). This change was necessary to accommodate several data sources (description below) with appreciable numbers of age-5 and younger fish as well as to provide more detailed information from existing data sets, such as the setline survey and commercial fishery. As an example of this additional information, coastwide fishery age data contain appreciable numbers of age-4 and age-5 halibut from the 1930s through the introduction of the current 32-inch minimum size-limit in 1973 (Fig. 10). As in past assessments, age-data were still been aggregated at age-20+ for years where only surface ages are available: prior to 2002, except for the 1998 setline survey data, which was re-aged in 2013.

In historical assessment models, there have been no data representing the age-structure of the discards from the commercial fishery, bycatch in non-target fisheries, or the sport and personal use removals. In the absence of direct data, selectivity curves were assumed for each of these sources of removals. In 2015, each of these sources was re-examined, and methods for including more representative selectivity estimates were developed. This effort began with the processing of relevant data for each, as outlined below.

The calculation of wastage, or halibut captured as part of the commercial fishery, discarded, and assumed to subsequently die, has historically been performed as an external analysis to the stock assessment (Gilroy and Stewart 2015). The magnitude of the wastage estimates has been based on the rate of sublegal to legal catch rates in the setline survey. This calculation has been made independent from the fixed selectivity curve assumed in the stock assessment. For 2015, the age-distributions of sublegal female and male halibut captured in the setline survey were compiled for evaluation in the stock assessment directly. These data showed a remarkably protracted age-distribution, with both male and female halibut age-10 and greater making appreciable contributions to the total (Fig. 11). The age-distribution for the two sexes also differed importantly, with sublegal females present in appreciable numbers from roughly age 7 to 11, and sublegal males from 7 to well beyond age 15 in some years (Fig. 11). The protracted age structure of fish below the 32” minimum size-limit illustrates the recent variability in size-at-age: some fish from each cohort reaching the minimum size limit by age-6, and others (particularly males) many years later. Although the distributions derived from survey data may not be strictly representative of the age-structure of the discards in the commercial fishery, they are consistent with the calculation of wastage outside the assessment model, and allow for the direct estimation of selectivity rather than simply the assumption of a fixed curve. Summary of these data also allows for comparison, and potentially replacement with direct fishery data collected by the various observer programs when and where it becomes available.

The length-distribution of halibut caught as bycatch in fisheries targeting other species is reported to the IPHC each year by the National Marine Fisheries Service (NMFS; for Alaska and Washington-Oregon-California) and Fisheries and Oceans Canada (DFO; for British Columbia). The historical time-series of these lengths has been summarized each year by regulatory area, and also aggregated to the coastwide level (weighting by the total estimated number of halibut) for use in the annual harvest policy calculations and catch tables. In order to evaluate these data directly in the context of the stock assessment, they first needed to be converted to age-distributions. Due to the large frequency of very small (and young) halibut observed in the bycatch removals, the length-to-age relationships from neither the setline survey, nor the directed halibut fishery were applicable. Halibut of all ages are routinely sampled for length and age by IPHC samplers on the NMFS trawl surveys conducted in the Bering Sea, Gulf of Alaska, and Aleutian Islands (Sadorus and Palsson 2014, Sadorus et al. 2015, Sadorus et al. 2015b). These data contain halibut of roughly the same size-range as have been observed in the bycatch data. Annual age-length keys were produced from the NMFS survey data for the years 1997-2014. Relatively few fish greater than age-15 were present in these data; therefore, to avoid extensive smoothing or extrapolation across years, the keys were aggregated at age-15. Without earlier data available, the key for 1997 was used for all prior years. Exploration of the average length-at-age didn’t show particularly strong trends for this age range; however, it would be preferable to have year-specific key information. Coastwide aggregate bycatch lengths were summarized into predicted ages via the annual age-length keys. Estimated bycatch ages showed a mode (or modes) between age-3 and age-10, with up to one-third of the total age distributions represented by halibut age-4 or less in some years (Fig. 12).

The length data currently available for bycatch and used in this analysis is in the form of summaries, for which the methods and original data sources are unknown. It is clear from several of the year-specific age-distributions that some of the historical data must be duplicated among years (e.g., 1974-1976 in Fig. 12). Ongoing efforts by the IPHC and the NMFS during 2015 to reconcile bycatch estimates and biological data may be able to provide a more reliable time-series for the 2015 stock assessment or in the near-future. The issues to be addressed include the stratification of estimates by IPHC regulatory area and the appropriate weighting of length data within and among vessels, fisheries, and areas. In the meantime, it may be reasonable to consider these data generally representative of the age structure of the bycatch, but annual observations may not be appropriate for deriving information on cohort strengths.

The final new source of information evaluated during 2015 was from the recreational fishery. Otoliths from recreationally caught halibut in regulatory Area 3A have been routinely collected by ADF&G, and the ages read by IPHC staff. Estimated numbers-at-age for the years 1994-2006 were weighted by port within Area 3A, and summarized by Scott Meyer (ADFG, pers. comm.). These data showed a variable but generally larger proportion at ages younger than age-5, and smaller proportion greater than age-15 (Fig. 13) compared to the coastwide setline survey over a similar time-period (Fig. 14). The recreational data contained a few halibut at ages 2-3, younger than any observed in the setline survey. The observation of extremely young halibut is somewhat surprising, as trends in size-at-age indicate that some of the smallest fish for their age across the coast are currently observed in Area 3A, so that area might be expected to have relatively fewer very young fish in the recreational harvest if selectivity were similar to that of the setline survey. These data are not geographically comprehensive; however, recreational removals from Area 3A represented 52% of the coastwide recreational total in 2014. Additional age data from the 3A recreational fishery collected in 2007-2013 were made available during the completion of this document, and will be analyzed for the final 2015 assessment. Currently, there are no additional age data from the recreational fisheries in other regulatory areas, but such data could be included with those from Area 3A if they become available in the future.

Model development

Structural rationale

The Pacific halibut stock assessment model has evolved through a number of different structural configurations (Clark and Hare 2006, Stewart and Martell 2014). Perhaps the most influential of these changes in recent years was the change from area-specific 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). A number of simulation studies have found that dividing a migratory population into several discrete assessment units tends to overestimate the total biomass (e.g., McGilliard et al. 2014 and Li et al. 2014 provide recent examples).

A primary structural assessment model choice is whether or not to model growth explicitly (and often parametrically) or empirically. Many U.S. and Canadian groundfish stock assessments assert a growth function of some type. 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 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). In the face of appreciable growth variability, a great deal of complexity is 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.

This challenge has resulted in many groundfish stock assessments taking a simpler approach to growth by using empirically derived weights-at-age where there are sufficient data available to do so. An example of this is the Pacific hake stock assessment, where a large amount of historical length data has been omitted from recent analyses, in favor of the use of weight-at-age directly, due to the complexity in observed growth (Taylor, I.G., Stewart, I.J., Hicks, A.C., and Hamel, O.S. *In review*. Drowning in data: empirical vs. parametric growth in an integrated stock assessment model. *Fisheries Research*). The simplicity of 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. 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 current fishery sampling approaches have 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 policy 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 treated 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 (as inferred from the survey catches) 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. Including four alternative assessment models in an integrated result should better approximate the uncertainty associated with the many structural choices that must be made in developing these models as well as the estimation uncertainty within each model.

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 individual model details (where they differ) below.

All 2015 stock assessment models were constructed using the generalized stock assessment software Stock Synthesis (Methot and Wetzel 2013, Methot and Wetzel 2013b, Methot 2015). The most recent version (3.24U) was used, however there were no changes made in recent versions that had any relevant impact on the Pacific halibut models as they have been developed over the last three years. This software separates the inputs into several files read in prior to 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 (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, bycatch, sport, personal use, and survey (by area in the AAF models). Table 5 summarizes the data and key features of each model. Age data were aggregated into bins representing each age from age-2 (which also includes ages 0 and 1) through age 25 (which includes all observations greater than or equal to age 25). Aging bias and imprecision were estimated externally to the stock assessment, based on multiple reads for both surface aging (all years <2002, except the 1998 survey data) and break-and-bake (all years >=2002, as well as the 1998 survey data) methods (Stewart 2015). Break-and-bake ages are assumed to be unbiased (which has been corroborated via radiocarbon methods; Piner and Wischniowski 2004) and estimated to be relatively precise, while the surface ages are increasingly biased and much less precise beyond about age-15 (Fig. 15). Each annual age composition observation was assigned the appropriate ageing method in the data file and age data were partitioned by sex (the vectors for each year contain females, then males), where this information was available. 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%. The model was also set up to add a very small constant (0.0001) to all age proportions in order to stabilize the computation.

All 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 and the vector of maturity-at-age (Stewart 2015).

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 6 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. However, an important aspect of the treatment of the stock-recruitment relationship for all models is that they were structured to recreate the time-series of recruitments, not to estimate reference points such as *MSY* (this is discussed further in the context of each model below). This means that the output in the report file and automatically generated figures in the background material pertaining to *MSY* are not meaningful. However, this does not apply to the calculation of the Spawning Potential Ratio (*SPR*) as it is calculated on a per-recruit basis.

The double-normal selectivity parameterization (option #20) is used extensively 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 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. 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, which was iteratively adjusted such that the resulting variability in the deviations was similar to, but less than the value for σ (essentially the 'Thompson and Lauth method'; Annex 2.1.1 in Thompson and Lauth 2012).

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 described above, several new age data sets were available for evaluation in 2015 including the sublegal halibut captured by the setline survey, the estimates from the bycatch length frequencies, and the recreationally caught halibut from Area 3A. Rather than assume a fixed selectivity curve, as has bene done in the past for discard mortality and bycatch, for 2015 these curves were estimated in the assessment models.

There are currently no options for age-based discarding (selectivity plus a retention function) available in Stock Synthesis. Therefore, as has been the case for all historical halibut assessments, discards are treated as a separate fishery. This treatment of discard removals (sublegal wastage) was substantially improved in 2015. First, sex-specific selectivity curves were estimated in each model based on 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, and the relative scale of females to males estimated directly. Both sexes were allowed to be 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, preliminary 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.

The second improvement related to modelling discarded halibut in the directed fishery was to implement a way to quickly and easily evaluate the assumed 16% Discard Mortality Rate (DMR). Using the features readily available in Stock Synthesis, the approach was based on the existing length-based implementation. Briefly, a retention function (constant across all lengths; option #2 in Stock Synthesis) was added, with the retention parameter (ρ) set to 0.0016. Input removals in the data file were divided by 100, and the DMR parameter (ψ) was set to 0.15864, the result being an identical quantity of dead discards to the previously assumed value, but with a mortality parameter could be adjusted to correspond to different hypotheses. For example, 1,000 pounds of handled halibut, with an assumed DMR of 16% would result in an estimated 160 pounds of wastage. As implemented: 1.6 pounds of input catch, implies 1,000 pounds handled $\left({}^{1.6}/_{\rho}\right)$, 158.4 pounds of additional handling mortality $\left(\left({}^{1.6}/_{\rho}\right) * \psi\right)$, for a total of 160 pounds of wastage, or an implied discard mortality of 16% (160/1,000). This approximation, where a DMR parameter value of 15.9% approximates an actual value of 16% is effectively linear (to less than 0.2%) across a range of relevant DMRs.

Sensitivity analyses could be performed on the assumed DMR parameter directly (rather than simply adjusting the wastage calculated outside the model as was done in 2013), or uncertainty could be integrated directly into the model results via an informative prior. This particular configuration of settings (interacting with empirical weight-at-age) had apparently not yet been closely evaluated in Stock Synthesis, and a minor reporting error was discovered in the code, such that the report file contains accurate numbers but inaccurate biomass values for discarded fish. The Synthesis code was updated to fix this issue (R. Methot, personal communication) and the updated code was subsequently tested on the halibut files to verify that the dynamics were being correctly calculated; estimates of discards in the report files provided in the background material contain this reporting discrepancy. A similar approach could be taken for bycatch DMRs, however identification of a single DMR and the range of factors contributing to uncertainty in bycatch are far more complex (see discussion of uncertainty below).

Due to the unknown origin and quality of the bycatch length frequencies, and the additional uncertainty associated with using the NMFS trawl survey-based age-length key to convert those lengths to ages, it did not seem reasonable to take these data as informative about the population dynamics. However, they should contain some information about the shape of the selectivity curve, potentially more reasonable than the fixed curve assumed in previous assessments. As with the sublegal discard data, down-weighting to a very small input sample size eliminated appreciable effects on model results, but still allowed for a sex-aggregated (and time-invariant) selectivity curve to be estimated. Also like the treatment of discards, in all models this curve was allowed to be dome-shaped given the relative frequency of younger halibut in the distributions, and the general observation that large halibut are not efficiently captured by trawl gear, which comprises the majority of the bycatch removals.

Where historical assessments assumed that recreational removals were subject to the same selectivity as the setline survey, age data collected from Area 3A suggested a greater proportion of young and fewer old halibut (see above). These data were introduced to all four models, and down-weighted such that selectivity parameters could be estimated (with a commensurate contribution to uncertainty), but little signal would be imparted to the modelled dynamics. Because of this down-weighting, 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 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). Where both sources of error are accounted for but unknown, they cannot be freely estimated simultaneously.

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 (Tables 3 and 4). 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 (the fishery ages from Area 4B in the AAF short model were the sole exception, and these already represented one of the smallest values). In no cases were the input values increased from those derived from the number of trips or stations represented in the data.

The degree of process error had been evaluated in the 2012 and 2013 stock assessments, where the σ parameters (defining the random walk in selectivity parameters) were adjusted such that the change in model results using larger values was not appreciable. Specifically, the 2012 assessment found that retrospective bias was substantially reduced by allowing selectivity to vary over time for both the fishery and survey (the survey was particularly sensitive), and that this bias decreased as the temporal variability in selectivity increased. Those analyses also suggested that model estimates of stock size and trend responded to changes in the σ parameters, when the input values were small, but as they were increased little additional change was observed. When the models were extended in 2014, the σ parameters for each fleet in each model were not revisited. Evaluation during 2015 revealed that in some cases the σ parameters were appreciably larger than the variability in the resulting estimated deviations.

As a general rule, a logical approach to the treatment of observation and process error would be to adjust model structure to fit the data as well as possible, increasing or decreasing the constraint on time-varying processes (e.g., the σ values for each fleets selectivity or catchability) to be consistent with the resulting variability, then adjust the observation error to be roughly consistent with the resulting lack-of-fit. For the coastwide models, repeating this process one to two times resulted in a reasonable balance between process and observation error that minimized residual and diagnostic patterns and did not appear to re-introduce dramatic retrospective patterns. However, for the AAF models, this process did not rapidly reach a stable solution: the trade-off among fleets might lead to more processes or observation error in one iteration and less in the next. However, the trend and scale of the solutions tended to be reasonably robust to the adjustment of process and observation error for particular fleets, once both were within a reasonable range. Specific diagnostics are provided for each model below.

For model integration and calculation of decision table metrics, three-year projections were needed from each assessment model. Projected catches were assumed to be known, and treated as inputs to the forecast file. Projected selectivity for each fleet was assigned the average of the most recent three years of estimated values. This is accomplished dynamically in the model code, which serves to propagate some uncertainty (about the mean), but not all of the uncertainty associated with a future year's selectivity. Weight-at-age is projected using the most recent year's observed values, as this was found to have little effect in previous comparisons.

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 (1996). 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 (R_1) was also estimated. The effect of these two approaches was to essentially decouple the numbers-at-age in 1996 from any equilibrium assumptions.

Due to the short time-series, and for consistency with previous halibut assessments, there was no explicit stock-recruitment function imposed on the coastwide short model. To achieve this, the equilibrium recruitment level (R_0) parameter was estimated setting the scale of the stock-recruit relationship. Steepness (h) was fixed at a value of 0.75; however, recruitment deviations were implemented with no zero-centering constraint (simple deviations in ADMB, option #2 for the type of deviation in Stock Synthesis), which means that the central tendency of the stock-recruit function is unimportant (and uninterpretable) because the deviations are not necessarily centered on the curve. Without zero-centered deviations, calculation of equilibrium based reference points (i.e. *MSY*) cannot be performed internally with this model. In evaluating the 2014 model configuration, the degree of recruitment variability (σ_r) was found to be mildly constraining to the estimated deviations, so it was increased to a value of 0.9, appreciably above the resulting level of variability (the RMSE of the maximum likelihood estimates for constrained deviations will always have a negative bias relative to the appropriate σ, as the deviations are not being integrated as they should be in a Bayesian or random effects context). A summary of the number of estimated parameters contributing to each aspect of the model is provided in Table 6.

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, 1996 for the fishery, and 1997 for the survey. No deviation was estimated for the terminal year (2015), because the data were not yet available (this means that the catches in 2015 may have a different effect on the projections when they are 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 are informed by the diffuse 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 6).

Based on exploration in 2015, the scale of male selectivity for both the survey and fishery were made more flexible by allowing it to also vary over time as a random walk. With only sex-aggregated commercial fishery age compositions, 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. A specific sensitivity test of this change was explored and is reported below.

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 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, and the current harvest policy (Clark and Hare 2002 and 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, 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:

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

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 use of *MSY* estimates that might be used for informing the harvest policy; however, the calculation of *SPR* is independent of steepness and can be compared to harvest-policy based 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 highly unstable (when allowed to be freely estimated, the value often went to 1.0, inconsistent with available information about likely sex-ratios in the fishery), and was therefore fixed at the estimated offset for the setline survey. When this behavior was first identified in the 2013 stock assessment, the sensitivity in the scale of the estimated stock size was highlighted and reported as a major source of uncertainty. This continues to be the case, and is re-illustrated in the sensitivity analyses reported below. Assigning the survey value for the scale of male selectivity from the survey to the fishery does not imply the same sex ratio in the catch for all ages, only those beyond the peaks of both female and male selectivity which represent only a subset of the total removals. 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).

Areas-As-Fleets 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 Area 2 and Area 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 set equal to the analogous estimated parameters for the setline survey in that area. This was an iterative process, as changes in the fishery selectivity influenced the estimated survey selectivity; however, the values usually converged to within one or two model runs. Estimation of some or all of these male scale parameters could be evaluated in future efforts. Temporal variability in selectivity parameters occurred over a slightly longer range of years in the AAF short model, as there were area-specific survey data available for the entire time-series from Areas 2 and 3.

Areas-As-Fleets 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.

Coastwide short model

Diagnostics

Predictions of both the fishery and survey indices of abundance fit the observed data very well in the coastwide short model (Fig. 16). The predicted aggregate age distributions also matched the observed distributions quite well, indicating that the selectivity approach was generally capturing differences in both the age-structure and sex-ratio among the model fleets (Fig. 17). 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 multiplier estimated via the Francis method did not suggest further reductions (Table 7).

Fit to the annual setline survey age compositions were good, although some patterning was visible in the standardized residuals (Fig. 18). 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 (Fig. 19). 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.

Results

Estimated selectivity for the discard fleet differed appreciably for males and females, with females less selected than males overall and declining beyond about age-11, where males were fully selected until about age-16 before becoming highly domed (Fig. 20). These estimates are very consistent with the observed dimorphic growth and its likely interaction with the 32-inch minimum size-limit in the commercial fishery. Estimated selectivity for the bycatch fleet was quite similar to the fixed curve used in historical assessments, (suggesting that fitting to the length data may have been the method used to generate the original). Halibut of ages 2-7 were much more strongly selected by the bycatch fleet than any other in the coastwide short model, with full selectivity occurring at ages 4-5 (Fig. 20). Estimated selectivity for the recreational fishery 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. Neither the survey nor the fishery selectivity was estimated to have a highly variable ascending limb over the short time-series (Figs. 21 and 22). The fishery selectivity estimated a trend toward increasing selection of males in recent years, 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. Because the addition of temporal variability in the scale of male selectivity was new for 2015, an alternate model with time-invariant male scaling was also investigated (see sensitivity results below).

The degree of variability in fishery catchability was much smaller than that implied by the input σ over a broad range of starting values. Reducing this sigma until it was commensurate with the observed variability in the deviations resulted in little model change. Fishery catchability showed a trend toward increasing values in the more recent years, however the scale of this change was trivially small (Fig. 23). The sensitivity to assuming strictly constant catchability was explored, and is reported below.

Male natural mortality was estimated to be slightly less (0.138) than the fixed value assumed for females of 0.15 (Table 8, Fig. 24). The difference in natural mortality, combined with lower overall selectivity for male halibut, suggests highly skewed sex ratios that are increasing somewhat as the larger cohorts of the 1980s leave the population (Fig. 25).

In aggregate, all the updates and improvements made to the coastwide short model in 2015 had the largest influence on the spawning biomass estimated for the early portion of the time-series (Fig. 26). Additional figures of the coastwide short model results, in addition to the entire report file containing all parameter estimates, are included in the background materials. However, note that many of the plots produced automatically are not relevant to the specific model configurations used here (e.g., biology plots, stock-recruit plots for the short time-series models, etc.).

Coastwide long model

Diagnostics

Both the fishery and survey indices were fit well, with breaks in catchability to accommodate the change from “J” to circle hooks very conspicuous in both series (Table 8, Fig. 27). In aggregate, the predicted age compositions matched the observed data well (Fig. 28); 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 (Figs. 29 and 30). Fishery data fit reasonably well for the entire time-series, with patterns in the residuals corresponding to relatively small differences with observed distributions (Figs. 31 and 32). Harmonic mean effective sample sizes were much larger than adjusted inputs (Table 7). The Francis multipliers suggested slightly more weight to the fishery data, and less to the survey, but this seemed to be inconsistent with the residual patterns and scale in the recent part of these time-series.

Results

Older halibut were more represented in the bycatch age data prior to 1996, and therefore the estimated selectivity had a higher selectivity asymptote than was estimated in the coastwide short model (Fig. 33). Due to the unknown quality of the currently available bycatch age distributions, not attempt was made to allow the bycatch selectivity to change over time, although this could be explored if and when data thought to be more reliable can be included. Recreational and discard selectivity estimates were relatively similar to those from the coastwide short model. Estimated survey selectivity showed a pattern of decreasing relative values for younger halibut through the mid-2000s and then an increase at the end of the time series (Fig. 34). This may be consistent with changes in both the age-structure of the stock and the spatial distribution. Fishery selectivity generally showed a pattern toward selecting fewer younger fish over a longer historical period, but a similar trend to the setline survey in the most recent years (Fig. 35). Fishery catchability showed a very large (unconstrained) increase associated with the change from “J” to circle hooks, and a similar trend from the late-1990s through the end of the time series as was estimated in the coastwide short model (Table 8, Fig. 36).

Female natural mortality in the coastwide long model was estimated to be higher (0.202) than for males (0.156; Table 8, Fig. 37). The environmental link parameter (β) was estimated to be positive (0.308), with very little density below a value of 0.0 (Table 8, Fig. 37). However, the time series of estimated recruitment deviations 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 residuals, respectively (Fig. 38).

The net change to the time-series estimates from all the updates and changes made for 2015 was minor and had the largest influence on the peak biomass values (Fig. 39).

Areas-As-Fleets short model

Diagnostics

The AAF short model fit the observed trends in Areas 3, 4, and 4B relatively well, but not the trend observed in Area 2 (Fig. 40). None of the configurations evaluated for either AAF model were able to capture the full extent of the recent increase in Area 2, and the continued decline in Area 3 at the same time. If the mismatch in trends for Area 3 and Area 2 are actually spatial in nature (halibut with similar demographics are moving from Area 3 to Area 2), then there is little chance of capturing both trends simultaneously with any approach that is not explicitly spatial, even using separate fleets as in the AAF models. Trends in the fishery catch-rate indices were also fit reasonably well, including in Area 2 (Fig. 41); this was achieved via changes in catchability (see AAF short model results below).

Fit to the aggregate age data for each model fleet clearly illustrated the differences in age structure among them (Fig. 42). The biggest differences between female and male halibut occurred in the Area 3 survey, and generally Areas 4 and 4B were predicted (and observed) to have the greatest fraction of older halibut, particularly males. The fit to the annual setline survey data generally captured these patterns (Figs. 43 and 44); however, there were some relatively strong patterns in the residuals and the fits to the data from Area 4B were noisy (Figs. 45 and 46). Although the input sample sizes were substantially below the harmonic mean effective sample sizes by fleet, the Francis multipliers suggested further reduced emphasis on the survey age data (or perhaps increased process error in the selectivity deviations; Table 7). The AAF models, due to the complexity of tuning constraints on the deviations of selectivity and catchability, as well as the scale of the male selectivity were not tuned extensively, but rather a few iterations were made to bring the scale of residuals and σ parameters generally in line with the diagnostics. Fits to the fishery age data (Figs. 47 and 48) were somewhat better, however there were still clear residual patterns (Figs. 49 and 50). Perhaps the most clear of these patterns was the lack of fit to the very strong 1987 cohort apparent in the Area 4 fishery data (Fig. 50, upper panel). No model configurations evaluated during model development were able to fit the peak observations of this cohort observed in Area 4 (and to a lesser extent in Area 4B), which may be a reflection of the spatial nature of the dynamics not well approximated by an AAF approach.

Results

Male survey selectivity was estimated to be shifted much more strongly to the right relative to females, in Area 3 compared to Area 2 (Figs. 51 and 52). The surveys in both Area 4 and Area 4B were assumed to have asymptotic selectivity, with Area 4B showing a greater amount of temporal variation in the estimated ascending limb, and much younger males selected than in Area 4 (Figs. 53 and 54). Estimated fishery selectivity showed generally similar patterns, but with somewhat less temporal variation (Figs. 55-58). Bycatch, sport and discard selectivity estimates were similar to those from the coastwide short model.

Estimated fishery catchability showed differing temporal patterns and scale by Area (relative to the coastwide population dynamics), with the observed increasing trend in Area 2 corresponding to increasing catchability for the fishery in that area (Fig. 59). Temporal change estimated for Areas 4 and 4B were much smaller than in Areas 2 and 3, and this was also the case in preliminary analyses where even weaker constraints were placed on the deviations.

The estimate of male natural mortality in the AAF short model (0.129) was slightly lower than in the coastwide short model (Table 8, Fig. 22).

In aggregate, the result of all the changes and improvements to the AAF short model led to a small increase in the scale of the spawning biomass estimate which was observed across the scale of the whole period (Fig. 60).

Areas-As-Fleets long model

Diagnostics

Like the AAF short model, the AAF long model fit the survey trends relatively well, with the exception of Area 2 (Fig. 61). The fishery index in Area 3 (also similar to the fit in the AAF short model) predicted an increase at the end of the time-series despite continued decline in the observations (Fig. 62). If this pattern represents a spatial trend, then fishery catchability would appear to be the only way for a non-spatial model to begin to fit these trends.

Aggregate fits to the survey age composition data showed similar patterns to those observed in the AAF short model (Fig. 63). Generally, the fit to the survey data improved over the time series the poorest fit to the age data occurring in Area 4B (but that Area also had considerably lower average sample size; Table 7, Figs. 64-66). Residual patterns appeared to indicate temporal changes in the sex ratio, especially in Area 2 and Area 3, that were not fit by the time-invariant parameterization employed in this model (Figs 67 and 68). These patterns might be explored further in the AAF models by allowing the scale of male selectivity to vary over time on a fleet-by-fleet basis. Fits to the sexes-aggregated fishery data were reasonably good for Areas 2 and 3 (Figs. 69 and 70), although some patterns were still apparent in the residuals (Fig. 71). As was observed in the AAF short model, the fits to the Area 4 age data failed to capture the peaks of dominant cohorts (Fig. 72) leading to strong diagonals in the residuals (Fig. 73).

Results

Bycatch, discard and recreational selectivity estimates were similar in the AAF long model to those estimated in the coastwide long model. For each survey fleet, the temporal pattern of selectivity is shown (Figs. 74-77). Because the changes in selectivity for the Area 4 and 4B surveys only occurred at the end of the time-series, cropped contour plots are also presented (Figs. 78 and 79). Fishery selectivity is shown in Figures 80-83; for Area 4 and 4Bcontours are also shown to make the trends more visible (Figs. 84 and 85). Generally the estimated fishery selectivity shows a gradual pattern toward older fish in all areas, somewhat different than the variable temporal trends estimated for the survey data. Fishery catchability was estimated to be strongly increasing in Area 2 and decreasing in Area 3 at the end of the time series (Fig. 86). There was little change estimated for Areas 4 and 4B, but all areas showed a large offset associated with the change from “J” to circle hooks, as was estimated in all four models (Table 8, Fig. 86).

Female natural mortality was estimated to be only slightly less than 0.15 (0.148) and higher than the estimated value for males (0.129) in the AF long model (Table 8, Fig. 87). The environmental link coefficient was estimated to be somewhat stronger (0.522) than in the coastwide long model. Investigation of the predicted sex-ratio over time suggested that the ratio of males to females is highly dynamic, responding to both exploitation and year class strengths (Fig. 88).

The net change from the 2014 to preliminary 2015 model results were less pronounced for the AAF long model than any of the other three models (Fig. 89), with most of the change occurring in the early time-series.

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. Comparison of the two short-time-series models illustrated that the uncertainty intervals from either of one these models alone would be grossly insufficient to represent a reasonable risk assessment (Fig. 90). Comparison of the two long time-series models illustrated the effects of differing assumptions about domed vs. asymptotic selectivity for the early portion of the time period where the fishery was focused primarily in Areas 2 and 3 (Fig. 91). This aspect of the two long time series models was explored further as a sensitivity analysis (reported below). Although the recruitment time series for the two long models was similar in trend (Fig. 92), the scale, especially of the larger recruitments, reflects the large difference in estimated natural mortality rate between the two models. 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 (Fig. 93). It is not clear how additional data may or may not help to reconcile the divergence in trends in spawning biomass in the terminal years.

Convergence diagnostics

Many models were run with alternative phasing and starting values and there was no evidence that the MLE solutions were particularly sensitive to these choices. All four of these models returned a positive definite Hessian for all alternatives explored during development. Pairwise among parameter correlations were generally less than 90%. Maximum gradient components were generally less than 0.001 among alternative models explored, although the long AAF model varied between 0.001 and 1. The implementation of temporal deviations in selectivity includes a parameter for all years in the series, even when some years have no observed data. These parameters have no contribution to the dynamics (other than the indirect effect of additional change in the series) and frequently result in a value estimated to be very close to zero based solely on the constraint provided by sigma. It is unclear how or whether these parameters may influence the gradient structure.

The convergence of the coastwide short model was explored via a set of 100 sets of alternate initial parameter values created by adjusting each by a random addition of 10% of the range of the parameter bounds (from lower to upper). The goal of this type of exercise is to discover whether a very different path to convergence might identify a more global minimum in the likelihood surface. It that regard, this represents a one-sided test, capably only of proving lack-of-convergence, and it is desirable to have a high convergence failure rate in the test, which is indicative of a strong exploration. Of the 100 alternate sets of starting values, 47 produced models converged to the maximum likelihood estimate previously identified, 16 were nearly converged to that value, but had a slightly larger negative log-likelihood (these produced very similar results with regard to stock size), and 37 sets that failed to converge to a meaningful result. This suggests that the level of dispersion was sufficient to produce a reasonable test for convergence, and that the model is unlikely to be converging to a local minimum. Other models could not be run from automatically adjusted starting points due to the manual assignment of male selectivity offset parameters for the fishery fleets from the values estimated for the setline survey.

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. Both the coastwide and AAF short models showed a small retrospective trend in the scale of the spawning biomass estimates but not the trend, becoming more pronounced after five years of data had been removed (representing six model years, since there were no data yet available for 2015; Figs 94 and 95). These patterns appeared to be slightly stronger in 2015 after increasing the constraint on temporal variation (decreasing the σ parameters) to be more consistent with the level of variation estimated in the models. Original investigation of this using the 2011 model revealed the least amount of retrospective pattern both when the survey index was very strongly emphasized (effectively down-weighting the age data), and when the constraint on selectivity variation was reduced. A few alternate model configurations with substantially reduced temporal variability were explored during 2015 model development, and these suggested similar behavior. It is not clear exactly what the appropriate trade-off between flexibility in the deviations and retrospective behavior might be, but the terminal estimates from each of the recent ‘peels’ all fall within even the within-model uncertainty intervals, suggesting this is a smaller factor than others explored in the full suite of models. The two long time series models showed somewhat differing retrospective patterns, with no clear trend observed for the coastwide long model (Fig. 96) and only a slightly increasing trend as data were removed from the AAF long model (Fig. 97).

Although the coastwide short model was made more flexible in order to estimate the temporal trends in the scale of male selectivity for both the survey and the fishery, this did not improve the mild retrospective trends. However, none of these models contain data to strongly and directly inform changes in the sex-ratio of male and female halibut over time and these are changes are highly relevant to the degree of temporal variability that should be modelled, as well as to the resulting population estimates from each model.

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.
Assuming strictly proportional fishery catchability for the coastwide short time-series did not appreciably change the results (Fig. 98). Forcing the scale of the male selectivity to be time-invariant for both the fishery and survey in the coastwide short model also had little effect on the estimated time-series (Fig. 99). The scale of the estimated stock size was directly proportional and highly sensitive to the fixed value for female natural mortality (Fig. 100). The same degree of sensitivity to the fixed value for female natural mortality was also observed in the AAF short model (Fig. 101).

The fixed value of steepness (0.75) used in the coastwide long model, while extremely important for estimation of *MSY* and similar recruitment-based reference points, had a relatively minor effect on the scale of the stock size estimates compared to other sources of uncertainty. When estimated freely, the parameter estimate went to a value of 1.0 (although the model did not fit the data appreciably better), with the biggest difference in the estimated time-series occurring at the peak biomass levels at the beginning of the time-series and in the early 2000s (Fig. 102).

As was first identified in 2013, forcing the scale of male selectivity in the fishery to differ from that of the survey by +/-10% had a very strong effect on the scale of the biomass estimate (Fig. 103). This result applies to the AAF models as well, where the scale of the male selectivity is also assumed to match that of the survey on an area-by-area basis. In the absence of direct information on the scale of male selectivity in the commercial fishery there is no easy solution to this issue. It may be possible to estimate the scale parameter (and therefore propagate the uncertainty) as these models are more fully developed, however the historical period, lacking commensurate observations in the survey to balance the estimation is likely to remain problematic. Using an aggregated-sex model, such as the VPA developed in 2014 still requires an assumption of the sex-ratio to estimate female spawning biomass. The skewed and variable sex-ratios estimated in these models (Figs. 25 and 88) suggest that an aggregate approach could become highly biased if a simple assumption regarding the ratio of males to females was imposed.

Based on the differing historical trends observed in the 2014 coastwide and AAF long time-series model, an exploration of potential causes was made for 2015. Much of the historical fishery occurred in Areas 2 and 3 over the period when the two model's estimate diverged (before 1980, Fig. 91). To mimic the dome-shaped selectivity estimated for these areas in the AAF model, dome-shaped selectivity was allowed during the time-periods prior to 1958, 1959-1980, and 1981-1996 for an alternate configuration of the coastwide long model. This resulted in a substantial increase in the estimated stock size during much of the historical period (Fig. 104), and brought the results of the coastwide long model much closer to those of the AAF long model (Fig. 105). This alternate model essentially represents a hypothesis that older halibut in Areas 4 and 4B were relatively unavailable to the historical fishery. However, with domed fishery selectivity the coastwide long model also estimated a very high rate of female natural mortality (0.24), perhaps outside the plausible range for a species that is routinely observed to greater than 30 plus years, and to age-55 at the extreme. In addition, there was a substantial and abrupt change between the later domed periods and the selectivity after 1997 (Fig. 106). Conceptually, the degree of migratory connectivity among the areas should determine just how domed the early, and especially intermediate years, were as the fishery progressed to the north. The AAF long model captures this progression more naturally, but a continuously time-varying approach to the degree of domed selectivity might achieve a similar effect at the coastwide level. Further work could investigate the specific implementation of domed selectivity in the coastwide long model; refining this approach could also have implications for estimation of the scale of recent male fishery selectivity.

Other considerations

Uncertainty in the removals 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 both the degree of sublegal wastage (mortality) in the commercial fishery as well as to the magnitude of total bycatch. The scale of stock estimates was found to be relatively robust to differing levels of these removals. However, there remains considerable uncertainty in both the wastage and bycatch, although it arises from somewhat different sources in each case. In the case of wastage, the assumed static DMR of 16% could potentially scale the removals up or down, if the actual DMR differed appreciably, due to the relatively large number of halibut handled by the commercial fishery each year. Although it was not specifically investigated in 2015, the improved implementation in these updated models allows for a direct evaluation of the DMR and the potential use of an informative prior rather than a fixed assumption. However, estimation of catch in statistical catch-at-age models generally requires other stabilizing assumptions, so direct integration of this uncertainty may still prove challenging. This is a topic for future exploration.

The relevant uncertainties in both wastage and bycatch have differing components, not all of which are of equal uncertainty or potential magnitude. This is especially the case for bycatch uncertainty, where observer coverage, observer sampling, the total number of fish handled relative to the number assumed to subsequently die and the scale of the various fisheries all contribute. A qualitative comparison reveals that summarizing the uncertainty in the aggregate bycatch from many different fisheries into a single DMR or scalar is not straightforward (Table 9). For example, the major source of uncertainty in the hook-and-line fleets is likely to be the DMRs assigned to those fleets, while the a trawl fleet such as that in the Gulf of Alaska may be more uncertain with regard to the representativeness of the relatively low observer coverage.

During 2014, the uncertainty in the magnitude of bycatch was specifically addressed through the construction of alternative catch tables for the upcoming year. This process allowed for an area-by-area investigation of the sensitivity of the Blue Line (the application of the current harvest policy) to alternative levels of bycatch without the need to postulate a specific probability that alternative levels. Requested levels ranged up to the full PSC limits for Alaska. This appeared to be a helpful way to inform the management process, although it did not represent uncertainty in the historical values.

Generally, each of these models has differing but important sources of uncertainty that have not yet been and may not be easily be resolved. The coastwide short time-series model is highly dependent on the value assumed for female natural mortality. The coastwide long time-series model is sensitive to both the treatment of historical fishery selectivity as well as the scale of male selectivity in the fishery independently, and these may also be confounded given the data available. Both AAF models require a balancing of several confounded factors including: the degree of process error to allow in fishery catchability, fishery selectivity and survey selectivity, the degree of observation error to allow in fishery and survey age composition data and indices of abundance, as well as the scale of male selectivity for fully selected halibut. For the coastwide models there appeared to be more stability in the tuning of each of these factors, the AAF models with multiple fleets were much more sensitive to the allocation among error types by fleet. Heavily weighting toward observation error, led to reduced levels of process error, but tended to generate model results with very strong retrospective patterns, consistent with analyses in 2012 indicating that adding process error in selectivity was an effective tool in reducing retrospective trends. Heavily weighting toward process error did not appear to appreciably improve residual patterns in the data. It is clear that in this application (and in general), it is not possible to simultaneously estimate (or iterate toward a stable solution for) both process and observation error simultaneously. Continued development of these models may allow for estimation of the scale of male selectivity for one or more areas which would greatly improve the efficiency with which alternative weighting and error assumptions can be evaluated.

Model integration

Model-integrated quantities are used as the primary output for stock assessment results, as well as the basis for decision table probabilities. Quantities have been integrated for the recent time period (1996+, over which all four sets of model results are available) including: spawning biomass, exploitable biomass, and *SPR* (summarized as fishing intensity, $F_{XX\%}$, where the *XX%* represents the *SPR*). Decision table quantities are divided into four categories: stock trend (which is the only set of metrics that are independent of any harvest policy related assumptions), stock status, fishery trend, and fishery status. Integration is performed for all these quantities using the basic approach outlined below.

Methods

Ideally, probability distributions for each model would be obtained through Bayesian integration; however, only maximum likelihood estimates and asymptotic variance estimates are currently available. These approaches may differ importantly in both the estimates of uncertainty as well as the shape of the distributions for management-related quantities (e.g., Stewart et al. 2013). The basic approach to model integration is to create a collection of random draws from each of the four model outputs. 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* was created for each model (*m*), where the relative weight is simply the relative fraction of the total draws across all models comprised by n_m:

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

For the results presented below n_m for all models was set equal to one 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 then created. Summary statistics for the integrated distribution were then saved for reporting and plotting.

Exploitable biomass (*EB*) calculations were more complicated due to the fact that these are a product of the externally derived selectivity schedule (s) consistent with the IPHC's existing harvest policy. The exploitable biomass is the product of selectivity, the numbers-at-age estimates (*n*, by sex, *s*, and year, *y*), and the weight-at-age:

$$EB_y = \sum_{sex=m,f} \sum_{age} s_{s,a,y} * w_{s,a,y} * n_{s,a,y}$$

No uncertainty estimate is directly available for these quantities in the model output; therefore the coefficient of variation for the spawning biomass in the same year year was used to approximate the distribution for exploitable biomass. The standard deviation was then calculated from the mean and approximated CV, and a distribution was created as for spawning biomass described above. Exploitable biomass in the IPHC's harvest policy was originally a fixed function of length-based selectivity. This was converted to age-based selectivity in the 2012 model via the mean lengths-at-age observed in the setline survey. This produced a historical trend as size-at-age declined, but had been relatively constant for the terminal several years. Since 2013 the selectivity at age describing the exploitable biomass has been held constant (Fig. 107) as little change had occurred in recent coastwide size-at-age. In order to provide calculations consistent with the policy, if size-at-age changes appreciably in the future this transformation may need to be updated unless revisions to the harvest policy are conducted in the interim.

The calculation of reference points was structured to match the assumptions of the IPHC's current harvest policy as closely as possible, and to use all available information within each stock assessment model. The current harvest policy employs a control rule that reduces the target harvest rate in each regulatory area linearly from the default values (21.5% in Areas 2A, 2B, 2C, and 3A, and 16.125% in Areas 3B, 4A, 4B and 4CDE) at $SB_{30\%}$ to zero at $SB_{20\%}$. In the presence of variable recruitment and size-at-age, these reference points were originally identified with poor environmental conditions for recruitment and relatively good size-at-age (consistent with observations in the 1950s through 1970s; Clark and Hare 2002, Clark and Hare 2006) Since the long time-series models explicitly included more information on these processes than the short models, the calculation of $SB_{30\%}$ and $SB_{20\%}$ differed in the two cases. Two important quantities were not available internally to the short time-series model used in 2006 and were therefore pre-specified when the harvest policy was first applied: the historically estimated ratio of average recruitment during poor and good recruitment regimes was 4.13/13, and the historically estimated spawning biomass per age-6 recruit was 118.491. Using the short time-series models, for the same approach as originally developed, required that the average number of age-0 recruits for the period ending in 2006 (when the PDO reverted to the poor regime) were projected to age-6, accounting for natural mortality (i.e., e^{-M*6}), and the initial age structure in 1996 similarly adjusted to age-6. These calculations then produced an estimate of $SB_{x\%}$ for comparison with current and projected future biomass:

$$SB_{x\%} = x * 118.491 * \frac{4.13}{13.0} * \overline{rec}_{age-6}$$

The values for $SB_{20\%}$ and $SB_{30\%}$ were calculated using the same formula. The historically estimated but fixed quantities in these calculations did not have associated uncertainty estimates, and therefore the reference points themselves ($SB_{30\%}$ and $SB_{20\%}$) did not have uncertainty estimates. Because the quantity of interest was the ratio of current to reference point *SB*, and these values must be correlated, it would not be appropriate to add additional uncertainty to the calculation beyond that present in the current biomass estimate without including an appropriate covariance term.

For the long time-series models, this calculation is much simpler. Treatment of the PDO regime was structured such that a value of 0.0 applied to the poor phase, and this is used for the internal calculation of reference points, thus no historically-based adjustment was necessary. Similarly, the average weight-at-age for the period 1950-1980 was assigned to the internal calculation of reference points, consistent with the data available for the historical analysis producing the spawning biomass-per-recruit used in the short time-series models. This means that the ratio of current to unexploited equilibrium *SB* (sometimes confusingly referred to as 'depletion') in the long time-series models was fully internally consistent propagating both the variance and the covariance in each *SB* component.

For all four stock assessment models, current and projected future spawning biomass estimates, conditioned on alternative input projected catch streams, are available directly. The only difference with the similar *SB* calculations described above is that in addition to the asymptotic estimate of the standard deviation for each biomass (e.g., *SB* current vs. *SB* three years in the future), the correlation is also included in the calculation of the ratio defining the probability of stock decrease. Specifically, instead of drawing a vector of independent random normal values for each SB, the draws are multivariate normal, including the 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. This calculation creates a distribution of projected harvest rates by dividing the TCEY corresponding to the removals in that row of the decision table by the distribution of exploitable biomass (as described above). The ratio of the projected harvest rate to the target rate (calculated based on apportionment, 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, accounting for uncertainty in the exploitable biomass, but not the target exploitation rate itself.

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 fishery CEY than the value specified for that row of the decision table. This calculation first creates a distribution of exploitable biomass values, then finds the target harvest rate accounting for the spawning biomass relative to the harvest control rule and creates a distribution of future TCEYs. To get to the distribution of available FCEY values from the TCEY distribution, the projected removals of halibut greater than 26 inches in length (O26) not included in the FCEY calculations first need to be removed to be consistent with the calculation of catch tables (Stewart 2015b). These include static projections based on the terminal year's data (e.g., recreational removals not included in Catch Sharing Plans, CSPs), as well as O26 removals that scale with the FCEY (e.g., recreational removals included in CSPs).

Results for 2015

For this preliminary analysis, the same equal weighting among models used in 2014 is used to generate integrated results for quantities of management interest. With the additional year of projection to 2016 (and in the absence of additional data) there is considerable uncertainty in the terminal estimates of spawning biomass from each model and among models (Table 10, Fig. 108). This corresponds to a broad cumulative distribution (Fig. 109). The integrated time-series reflects this uncertainty (Fig. 110). Projected median management quantities are generally consistent with the values and trends estimated in the 2014 (Table 11) and recent assessment despite the improvements made in 2015 (Fig. 111)

Future extensions

Continued refinement of the individual models included in the integrated ensemble results could potentially include additional sources of uncertainty. Where specific probabilities can be assigned to alternative values for key inputs, such as female natural mortality, the scale of male selectivity or the magnitude of bycatch removals, these could be used to weight additional models contributing to the ensemble. This would have to be done carefully, to avoid creating too much complexity (particularly the number of different combinations of models) and also to avoid inappropriate weighting. For example, if one of the four existing models was partitioned into three inputs, each representing an alternative level of female natural mortality, those inputs would conceptually be 'nested' and should not be weighted equally to all other models in the ensemble. Specific approaches are likely a useful avenue for future consideration.

Weighting of the four models included in the ensemble could potentially be made less subjective if criteria were developed that represented the relative quality of each model. Such criteria could be based on retrospective and prospective model behavior, fit to summary data common to each (or at least capable of being summarized in each, e.g., the coastwide survey index of abundance), relative behavior in simulation experiments, and other measures of performance. However, none of these approaches is likely to clearly identify a single model (at least over the set that has been examined to date) as far superior to all others and therefore dramatically change the relative weights. Further refinement of the existing models, and continued evaluation of alternative models may be as important as the specific weighting within the ensemble. In addition, periodic comparison of ensemble results with very different approaches, such as the simple surplus production and VPA models developed during 2014 may help to better understand the dynamics in a general sense.

Additional management metrics could be added to the existing decision table (Table 12) as the need arises. Potential candidates could include metrics that are independent of the current harvest policy, such as the probability that the projected future level of fishing intensity ($F_{XX\%}$) is greater than the most recently completed fishing year. Such metrics, similar to those pertaining solely to stock trend, do not rely on the many assumptions embedded in the current harvest policy calculations. Generally, although certain metrics will likely display more or less contrast in a particular year, it may be helpful for all users of the table to continue to report a consistent set of metrics without changing the format of the table dramatically each year. A consistent set of risk metrics will also enable future evaluation of the 'risk profile' for historical decisions. This type of analysis will become more informed as multiple years of decision-making become available and may prove to be a useful input to the Management Strategy Evaluation (MSE) process.

Of current interest, in the interim before more detailed MSE results are available, are the properties of the current harvest policy. The application of this policy and its current results are outlined in Stewart (2015b). The salient result for consideration here is how the Blue Line results in the decision table are calculated. Currently, the median exploitable biomass, target harvest rates (and harvest control rule), as well as the detailed array of O26 removals are used to generate a target level of removals consistent with historical calculations. However, the IPHC's current harvest policy does not explicitly address changes in U26 mortality from those inherently included in the original simulations (Stewart et al. 2015). The implied assumption was that sources of U26 mortality would represent a minor and relatively static component of the Pacific halibut mortality over the long-term. This can introduce a lag in response if, for example, U26 mortality increases, and there is no response in harvest policy calculations until that increase in mortality is observed in subsequent year's surveys and trends. For this reason, in 2014 the projected level of fishing intensity (including all sources of mortality) was included with the decision table. If appreciable trends in the level of fishing intensity associated with the Blue Line are observed, it may be reasonable to consider a constant SPR target as a logical analog to the current Blue Line. Further, ongoing efforts to evaluate Prohibited Species Catch (PSC) in non-target fisheries as well as the effect of shifting allocations among other sectors are not easily tractable in the context of annual harvest targets without a calculation that explicitly includes all sizes and sources of mortality.

Spatially explicit model development

There are several primary motivations for constructing a spatially explicit stock assessment model which include: 1) direct use of the NMFS Bering Sea trawl data which represents a long fishery-independent time-series index and age-distributions for young halibut from Area 4, 2) a better understanding of the historical and current biomass distribution among areas, and 3) direct estimates of potential selectivity and catchability differences among areas. These insights are very relevant to understanding the importance and spatial implications of current levels of directed fishery and bycatch removals, as well as the estimates of apportionment currently based on the setline survey.

Given the development of geographic area-specific weights-at-age during 2015, the extension from the datasets used for the AAF models to a fully spatial model required further partitioning of the bycatch, recreational and wastage removals by geographic area (as they are pooled in the AAF models). These processing steps were also completed during 2015. With the data complete, there are two key processes that must be informed in the parameterization of a spatially explicit assessment model: 1) the distribution of recruitments, and 2) the rates of movement (possibly by age) among regulatory areas. Because of the lag between recruitment and subsequent observation in the setline survey of directed fishery data, it is likely that estimation of juvenile migration vs. recruitment distribution among areas will be highly confounded. Unfortunately, neither of these quantities is well understood.

The NMFS Bering Sea trawl survey data represents a unique source of information on the abundance of halibut (especially those less than age-6) in Area 4, where the setline survey coverage is weakest. The age data from the NMFS trawl survey show a clear diagonal after 2006, apparently corresponding to relatively strong year-classes; the total numbers show a rapid increase in 2006 followed by a slow decline (Fig. 112; note that age data prior to 1998 are inferred from an age-length key from that year). There may be some information in these data (trends in the Bering Sea, as well as trends in the Gulf of Alaska) with which to estimate the rate of movement out of the Bering Sea and/or the fraction of the coastwide recruitment occurring in that area.

Current understanding of adult movement rates for most areas is reasonably well understood, based on extensive historical and more recent PIT tagging studies (Valero and Webster 2012). However, previous summary of these data has been conducted by specific regulatory area, and use in a spatially explicit model structured around geographic areas (2, 3, 4, and 4B) would require re-analysis, because values are not strictly additive from more detailed summary tables (e.g., Table 13). In addition, tag releases and recoveries in Area 4 were highly dominated by Area 4A, and it is unclear how well they might also represent halibut in the Bering Sea. Detailed analysis of these data was originally based on the length of the tagged halibut (Webster et al. 2013). In preliminary re-analyses of the PIT tagging data, Webster (2015) has begun to explore estimation these rates as a function of age. Preliminary results suggest movement-at-age estimates depend on the treatment of missing ages for fish that were measured when tagged but not recaptured. However, appreciable emigration is estimated to occur from areas in the western and central Gulf of Alaska, with the highest rates were observed for young halibut leaving Area 4A (Fig. 113). The PIT tagging data include very few halibut less than age-6.

For halibut less than age-6, most of the available data come from historical studies that used trawl gear (rather than longline gear) to capture fish for tagging (Valero and Webster 2012). Hilborn et al. (1995) used data from studies conducted in the 1980s to estimate movement parameters for juveniles among specific regulatory areas within geographic Areas 2 and 3 (Table 14). These data suggest relatively high rates of 'downstream' movement to the East and South. Similar results are unavailable for Area 4 or 4B, although raw recovery rates from juvenile halibut tagged in the Bering Sea and Aleutians suggest appreciable movement to all other regulatory areas over 5-10 years of life (Webster 2015b). The lack of data from Area 4 is particularly problematic, given that this is the area where the greatest abundance of 2-4 year old halibut are observed (Sadorus et al. 2015c), and therefore assumptions about movement rates will be most important.

Based on preliminary spatial model exploration conducted during 2015, a productive path forward may be to specify a fixed array of movement rates among areas based on re-analysis of tagging data outlined above, and making assumptions about likely values for Area 4 and 4B. This should allow estimation of the annual distribution of recruitment among areas. One impediment to initial development was that Stock Synthesis does not allow for estimated recruitment distribution-by-area (the partitioning of annual recruitment deviations into area-specific components) to extend into the initial age-structure. This requires a longer modelled time-series to achieve relatively unrestricted initial model conditions. Initial set-up of this model was done beginning the time-series in 1996; however, a longer time-series may be necessary. Extending the modelled time-series to include at least the entire NMFS Bering Sea trawl survey (1982-present; the 2014 age data will soon be available), and enough prior years of estimated recruitment distributions to populate the numbers-at age in 1982, would seem to be a reasonable compromise between the complexity required for a full time-series analysis and the potential loss of information and stability associated with a very short time-series.

Although it is possible that the NMFS trawl data as well as differences in trends and age-structure among areas could provide some diffuse information to update movement parameters, a spatially explicit model may ultimately represent a tool for hypothesis exploration, rather than a robust addition to the current ensemble of models used for annual risk-analysis.

Research priorities

Although a number of research tasks have been accomplished for 2015, many of the primary research priorities remain unchanged from previous assessments. These can be divided into three general categories: new information, existing information, modelling work.

Collection of new data:

1) Continued development of methods to estimate the sex-ratio of the commercial catch. Sampling of commercial fishery trips where fish have been voluntarily marked by fishermen at-sea is being undertaken during 2015 in tandem with the development of a genetic assay (based on SNPs) to precisely determine the sex of a dressed fish from a tissue sample. These may lead to a relatively inexpensive method for collecting data from across the commercial fleet and a validation tool to understand the precision and accuracy of estimates collected in that fashion.
2) Better understanding of movement rates of adults and juveniles between the Bering Sea and other regulatory areas may improve our ability to model the stock in a spatially explicit context. This must be considered as a long-term priority which may require several avenues of research (e.g., potentially tagging, exploration of naturally occurring markers, etc.).
3) During 2015, the calibration study (last performed in 2006) to compare halibut catch rates by length for the NMFS Bering Sea trawl survey and the setline survey will be repeated. This will allow for reanalysis of the survey biomass and apportionment estimates for 2015 and for previous years. Exploration of survey index variance estimates that include calibration uncertainty be possible, although two observations is likely insufficient to develop a reliable variance component from this source.

Analysis and processing of existing data:

1) Planned collaboration between the IPHC, the NMFS North Pacific Observer Program and the North Pacific Region should improve the estimates of bycatch and length-frequency distributions by IPHC regulatory area in Alaskan waters.
2) Re-analysis of survey data prior to 1997 could yield improved variance estimates, especially if based on model-based estimators. Such an approach would also allow for propagation of uncertainty due to missing portions of regulatory areas occurring even in the current design (depth zones and small spatial areas not sampled annually).
3) There is a vast quantity of archived historical data that is currently inaccessible until organized, keypunched and formatted into the IPHC's database with appropriate meta-data. Particularly, the ability to reprocess all historical fishery landings, effort, and age samples would provide a much clearer (and more reproducible) perception of the historical period than current work using summarized information.
4) Reconstruction of historical estimates for discards in the recreational fisheries as well as personal use or subsistence harvest prior to 1991 would make these time-series more accurate, although the changes are likely to be relatively small when compared to the total removals in historical periods.

Modelling:

1) Continued development of alternative models, including an explicitly spatial model that may help to improve process understanding and/or better understand the role of spatial processes contributing to among-model uncertainty. At some point it may also be

worthwhile to consider constructing and evaluating an assessment model with explicit time-varying growth and length-based processes for comparison.

2) Continued development and sensitivity testing of existing models to better understand factors such as: the tradeoffs between data-weighting and process error variability in catchability and selectivity, the estimability of the scale of male selectivity in the coastwide long and AAF models, domed-selectivity by area and period, treatment of the stock-recruitment relationship and environmental factors, as well as other technical aspects.
3) Weighting of the individual models included in the ensemble may be of increasing importance if estimated stock trends continue to diverge between coastwide and AAF approaches. Exploration of methods for less subjective weighting could be based on: prospective and retrospective statistics, fit to summary or aggregate data series, simulation performance, and other approaches.
4) Bayesian methods for may provide improved uncertainty estimates within the models contributing to the assessment ensemble.
5) Continue to explore methods for defining and including uncertainty in wastage and bycatch estimates in both the assessment and harvest policy calculations.
6) Continued integration of assessment data and modelling with ongoing development of the harvest policy and Management Strategy Evaluation process.

Acknowledgements

The IPHC's Scientific Review Board members, Sean Cox, Jim Ianelli, and Marc Mangel, and national science advisors Robyn Forrest, and Loh-Lee Low have contributed extensively to the direction, technical approach, and presentation of recent IPHC stock assessment modelling. The work of all IPHC staff contributes to the stock assessment in many ways from data collection and processing, to evaluation of the results. Bruce Leaman and Ray Webster in particular have provided valuable conceptual discussion and ideas for improvements. Many regulatory agencies and sampling programs outside of the IPHC provide data that is essential to these analyses and their help each year is greatly appreciated.

References

Clark, W.G., and Hare, S.R. 2002. Effects of Climate and Stock Size on Recruitment and Growth of Pacific Halibut. North American Journal of Fisheries Management 22: 852-862.

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

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68: 1124-1138.

Gilroy, H.L., and Stewart, I.J. 2015. Incidental mortality of halibut in the commercial halibut fishery (wastage). IPHC Report of Assessment and Research Activities 2014. p. 47-54.

Hilborn, R., Skalski, J.R., Anganuzzi, A., and Hoffman, A. 1995. Movements of juvenile halibut in IPHC regulatory areas 2 and 3. IPHC Tech. Rep. No. 31. 44 p.

Hurtado-Ferro, F., Punt, A.E., and Hill, K.T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. 158: 102-115.

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.

Martell, S., and Stewart, I. 2014. Towards defining good practices for modeling time-varying selectivity. Fish. Res. 158: 84-95.

Maunder, M.N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. 109: 311-319.

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.

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

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

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

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

Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in Australia. Fish. Res. 168: 85-99.

Sadorus, L.L., and Palsson, W.A. 2014. Results from the Gulf of Alaska NOAA Fisheries Service bottom trawl survey in 2013. IPHC Report of Assessment and Research Activities 2013. p. 471-478.

Sadorus, L., Lauth, R.R., and Ranta, A.M. 2015. Size and age composition of Pacific halibut in NMFS Bering Sea shelf trawl surveys. IPHC Report of Assessment and Research Activities 2014. p. 627-634.

Sadorus, L., Palsson, W.A., and Ranta, A.M. 2015b. Results from the NMFS Aleutian Isalnds biennial bottom trawl survey in 2014. IPHC Report of Assessment and Research Activities 2014. p. 635-644.

Sadorus, L.L., Stewart, I.J., and Kong, T. 2015c. Juvenile halibut distribution and abundance in the Bering Sea and Gulf of Alaska, IPHC Report of Assessment and Research Activities 2014. p. 367-404.

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.J., Leaman, B.M., Martell, S., and Webster, R.A. 2013. 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. 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fish. Res. 142: 37-46.

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.

Stewart, I.J., and Martell, S.J.D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. 158: 40-49.

Stewart, I.J., and Martell, S. 2014b. 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. 2015. Overview of data sources for the Pacific halibut stock assessment and related analyses, IPHC Report of Assessment and Research Activities 2014. p. 87-160.

Stewart, I.J. 2015b. Regulatory area harvest policy calculations and catch tables, IPHC Report of Assessment and Research Activities 2014. p. 195-212.

Stewart, I.J., and Martell, S. 2015. Assessment of the Pacific halibut stock at the end of 2014. IPHC Report of Assessment and Research Activities 2014. p. 161-180.

Stewart, I.J., and Martell, S.J.D. 2015b. Reconciling stock assessment paradigms to better inform fisheries management. ICES J. Mar. Sci.

Stewart, I.J., Leaman, B.M., and Martell, S.J.D. 2015. Accounting for and managing all Pacific halibut removals, IPHC Report of Assessment and Research Activities 2014. p. 221-266.

Taylor, I.G., Stewart, I.J., Hicks, A.C., Garrison, T., Punt, A.E., Wallace, J.R., Wetzel, C., Thorson, J.T., Takeuchi, Y., and Monnahan, C.C. 2014. Package r4ss. https://github.com/r4ss.

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.

Valero, J.L., and Webster, R.A. 2012. Current understanding of Pacific halibut migration patterns. IPHC Report of assessment and research activities 2011. p. 341-380.

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.

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.

Webster, R.A. 2014. Construction of a density index for Area 4CDE. IPHC Report of Assessment and Research Activities 2013. p. 261-288.

Webster, R.A. 2015. Modelling mortality and migration as functions of age using PIT tagging data. IPHC Report of Assessment and Research Activities 2014. p. 511-522.

Webster, R.A. 2015b. Trawl tag releases of small halibut in the Bering Sea, IPHC Report of Assessment and Research Activities 2014. p. 475-510.

Webster, R.A., and Stewart, I.J. 2015. Setline survey-based apportionment estimates, IPHC Report of Assessment and Research Activities 2014. p. 181-194.

Webster, R.A., Stewart, I.J., Leaman, B.M., Sadorus, L.L., Henry, E., and Dykstra, C.L. 2015. Setline survey expansion and complementary data sources. IPHC Report of Assessment and Research Activities 2014. p. 587-602.

Tables

Table 1. Setline survey numbers-per-unit-effort and estimated log(*SE*s); assumed values in italics.

	Area 2		Area 3		Area 4		Area 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)
1977	0.60	*0.105*	2.00	*0.062*	--	--	--	--	1.47	*0.113*
1978	0.80	*0.105*	1.30	*0.062*	--	--	--	--	1.11	*0.113*
1979	--	--	1.90	*0.062*	--	--	--	--	--	--
1980	1.20	*0.105*	2.50	*0.062*	--	--	--	--	2.01	*0.113*
1981	0.80	*0.105*	3.80	*0.062*	--	--	--	--	2.67	*0.113*
1982	1.84	*0.105*	3.80	*0.062*	--	--	--	--	2.87	*0.113*
1983	2.30	*0.105*	3.40	*0.062*	--	--	--	--	2.88	*0.113*
1984	6.74	*0.105*	11.60	*0.062*	--	--	--	--	9.30	*0.113*
1985	5.65	*0.105*	11.90	*0.062*	--	--	--	--	8.94	*0.113*
1986	4.54	*0.105*	7.80	*0.062*	--	--	--	--	6.26	*0.113*
1993	5.10	*0.105*	14.50	*0.062*	--	--	--	--	--	--
1994	--	--	15.50	*0.062*	--	--	--	--	--	--
1995	5.46	*0.105*	17.74	*0.062*	--	--	--	--	--	--
1996	7.35	*0.105*	17.59	0.035	--	--	--	--	12.89	*0.113*
1997	8.15	0.061	21.72	0.035	2.68	0.088	12.17	0.093	7.78	*0.056*
1998	5.51	*0.105*	19.28	0.029	2.94	0.081	10.68	0.092	6.97	*0.056*
1999	5.43	0.058	17.93	0.026	2.32	0.086	9.59	0.099	6.26	*0.056*
2000	5.58	*0.105*	19.14	0.030	2.47	0.076	9.77	0.075	6.63	*0.056*
2001	6.47	0.063	17.89	0.034	2.39	0.078	8.08	0.101	6.38	0.028
2002	5.79	0.055	19.76	0.031	2.22	0.076	4.75	0.097	6.39	0.026
2003	5.42	0.054	17.89	0.033	1.97	0.079	4.33	0.098	5.79	0.027
2004	5.90	0.053	21.04	0.028	1.99	0.089	3.49	0.098	6.46	0.026
2005	5.61	0.056	19.59	0.033	1.63	0.084	3.78	0.086	5.93	0.027
2006	5.20	0.056	17.38	0.032	1.59	0.082	4.39	0.109	5.43	0.027
2007	5.38	0.054	18.73	0.031	1.55	0.089	4.87	0.114	5.73	0.027
2008	6.12	0.050	16.55	0.031	1.94	0.075	5.16	0.126	5.65	0.026
2009	6.18	0.048	15.07	0.031	2.14	0.083	5.26	0.096	5.50	0.027
2010	6.17	0.055	14.12	0.033	1.95	0.091	3.90	0.109	5.12	0.029
2011	5.44	0.049	14.65	0.030	1.83	0.100	3.94	0.101	5.04	0.029
2012	6.85	0.044	15.26	0.029	2.18	0.099	3.33	0.129	5.56	0.030
2013	6.47	0.044	12.01	0.030	1.86	0.122	4.69	0.109	4.73	0.034
2014	6.90	0.042	13.66	0.026	1.97	0.117	4.07	0.125	5.16	0.032

Table 2. Commercial fishery weight-per-unit-effort and estimated log(*SE*s); assumed values in italics.

	Area 2		Area 3		Area 4		Area 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 2 continued. Commercial fishery weight-per-unit-effort and estimated log(*SE*s); assumed values in italics.

	Area 2		Area 3		Area 4		Area 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	60.96	*0.100*	61.95	*0.100*	--	--	--	--	61.00	*0.100*
1975	61.97	*0.100*	66.76	*0.100*	--	--	--	--	61.00	*0.100*
1976	44.78	*0.100*	61.91	*0.100*	--	--	--	--	55.00	*0.100*
1977	63.52	*0.100*	65.57	*0.100*	--	--	--	--	63.00	*0.100*
1978	54.57	*0.100*	68.47	*0.100*	--	--	--	--	71.00	*0.100*
1979	55.99	*0.100*	67.32	*0.100*	--	--	--	--	75.00	*0.100*
1980	60.31	*0.100*	116.09	*0.100*	--	--	--	--	94.00	*0.100*
1981	75.23	*0.100*	148.86	*0.100*	137.29	*0.100*	99.00	0.078	111.00	*0.100*
1982	73.54	*0.100*	181.34	*0.100*	97.82	*0.100*	--	--	127.00	*0.100*
1984	154.98	0.045	491.33	0.046	350.32	*0.100*	161.00	0.103	291.00	*0.100*
1985	164.97	0.049	535.07	0.039	441.41	0.103	234.00	0.160	355.00	0.034
1986	140.05	0.035	506.00	0.042	325.79	0.059	238.00	0.372	318.00	0.041
1987	138.34	0.027	490.38	0.036	353.58	0.162	220.00	0.111	319.00	0.041
1988	169.56	0.052	560.55	0.042	405.71	0.105	224.00	0.122	367.00	0.035
1989	156.34	0.040	507.69	0.031	379.27	0.080	268.00	0.094	357.00	0.025
1990	195.54	0.041	403.55	0.036	362.96	0.097	209.00	0.103	318.00	0.028
1991	171.04	0.037	375.03	0.041	365.91	0.157	329.00	0.085	317.00	0.038

Table 2 continued. Commercial fishery weight-per-unit-effort and estimated log(*SE*s); assumed values in italics.

	Area 2		Area 3		Area 4		Area 4B		Coastwide	
Year	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)	Index	log(SE)
1992	169.38	0.038	413.39	0.048	324.04	0.117	280.00	0.095	318.00	0.035
1993	201.84	0.029	439.12	0.096	400.32	0.447	218.00	0.220	372.00	0.099
1994	178.56	0.026	362.77	0.049	343.23	0.333	197.00	0.101	305.00	0.072
1995	193.28	0.025	439.49	0.043	330.24	0.100	189.00	0.336	329.00	0.036
1996	210.51	0.039	505.02	0.046	427.64	0.138	269.00	0.185	391.00	0.038
1997	237.91	0.033	498.02	0.026	432.98	0.103	275.00	0.064	403.00	0.025
1998	222.15	0.027	512.60	0.036	433.56	0.084	287.00	0.058	406.00	0.025
1999	246.32	0.074	475.50	0.024	406.93	0.058	310.00	0.045	392.00	0.023
2000	228.89	0.034	494.84	0.026	415.91	0.082	320.00	0.048	401.00	0.022
2001	203.93	0.036	454.52	0.029	365.53	0.212	270.00	0.076	361.00	0.042
2002	215.97	0.030	466.46	0.025	303.98	0.080	245.00	0.081	359.00	0.019
2003	210.18	0.018	439.26	0.024	254.87	0.071	196.00	0.068	328.00	0.018
2004	194.58	0.027	425.78	0.026	242.63	0.070	202.00	0.061	318.00	0.019
2005	180.41	0.022	387.69	0.023	219.65	0.063	238.00	0.093	296.00	0.017
2006	181.05	0.023	360.69	0.022	174.23	0.066	218.00	0.111	269.00	0.019
2007	160.26	0.021	344.26	0.026	150.21	0.057	230.00	0.108	251.00	0.020
2008	141.22	0.019	318.16	0.024	162.58	0.071	193.00	0.069	232.00	0.017
2009	154.83	0.019	277.22	0.020	175.29	0.054	189.00	0.097	222.00	0.018
2010	186.45	0.035	242.31	0.024	141.55	0.081	142.00	0.063	203.00	0.020
2011	182.96	0.019	226.64	0.025	141.25	0.057	165.00	0.103	197.00	0.015
2012	197.09	0.019	213.45	0.032	136.07	0.081	149.00	0.066	194.00	0.021
2013	195.61	0.024	189.98	0.033	117.43	0.075	127.00	0.064	179.00	0.017
2014	222.36	*0.057*	180.07	*0.089*	104.56	*0.183*	168.00	*0.182*	185.00	*0.049*

Table 3. Number of sampling stations contributing to survey age data.

Year	Area 2	Area 3	Area 4	Area 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	--	--	147
1995	103	120	--	--	223
1996	188	424	--	--	612
1997	200	429	221	74	924
1998	217	507	100	42	866
1999	320	556	61	82	1019
2000	229	553	153	83	1018
2001	322	522	148	83	1075
2002	300	558	154	82	1094
2003	312	518	153	82	1065
2004	319	527	148	71	1065
2005	329	509	152	83	1073
2006	310	529	181	84	1104
2007	317	540	181	74	1112
2008	326	552	184	76	1138
2009	325	559	179	84	1147
2010	324	533	182	78	1117
2011	348	554	172	79	1153
2012	349	524	174	72	1119
2013	357	537	170	80	1144
2014	367	567	241	77	1252

Table 4. Number of sampled trips contributing to fishery age data (inputs assumed for unknown values in italics).

Year	Area 2	Area 3	Area 4	Area 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

Table 4 continued. Number of sampled trips contributing to fishery age data (inputs assumed for unknown values in italics).

Year	Area 2	Area 3	Area 4	Area 4B	Coastwide
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
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	381	224	83	14	702
2012	421	217	81	13	732
2013	459	196	73	14	742
2014	435	221	64	8	728

Table 5. General overview of each assessment model.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Modelled period*	1996-2015	1888-2015	1996-2015	1888-2015
Data partitions	N/A	N/A	Area 2, 3, 4ACDE, 4B	Area 2, 3, 4ACDE, 4B
Directed Fishery fleets	1	1	4	4
Other fishing fleets	4	4	4	4
Survey fleets	1	1	4	4
Fishery CPUE (weight)	1996+	1907+	1996+	1907+, 1915+, 1981+, 1981+
Fishery age data years	1996+	1935+	1996+	1935+, 1935+, 1945+, 1991+
Survey CPUE (numbers)	1997+	1997+	1996+, 1996+, 1997+, 1997+	1977+, 1977+, 1997+, 1997+
Survey age data years	1997+	1963+	1996+, 1996+, 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	No	B-H	No	B-H
Initial conditions estimated	R_1, N-at-age: 1-19	R_0	R_1, N-at-age: 1-19	R_0
Environmental regime effects on recruitment	No	Yes	No	Yes
Steepness (h)	0.75	0.75	0.75	0.75
$\sigma_{\text{recruitment deviations}}$	0.9	0.6	0.75	0.55
Selectivity (fishery and survey)	Asymptotic	Asymptotic	Domed (A2, A3), Asymptotic (A4, A4B)	Domed (A2, A3), Asymptotic (A4, A4B)
Scale of male fishery selectivity	Estimated	Fixed = survey	Fixed = survey, by area	Fixed = survey, by area
Bycatch selectivity	Domed	Domed	Domed	Domed
Sport selectivity	Domed	Domed	Domed	Domed
Wastage selectivity	Domed, by sex	Domed, by sex	Domed, by sex	Domed, by sex
Personal use selectivity	Mirrored to sport	Mirrored to sport	Mirrored to sport	Mirrored to sport

*Preliminary removals for 2015 are projected based on the final adopted catch limits.

Table 6. Counts of estimated parameters in each assessment model.

	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
R_1 offset	1	--	1	--
Environmental link coefficient	--	1	--	1
Fishery catchability	1	2	4	7
Fishery selectivity	5	4	10	10
Wastage selectivity	7	7	7	7
Bycatch selectivity	4	4	4	4
Sport selectivity	4	4	4	4
Survey catchability	--	4	--	4
Survey selectivity	5	5	21	21
Total static	29	34	53	61
Time-varying				
Recruitment deviations[1]	43	161	43	161
Fishery catchability deviations	18	103	72	262
Fishery selectivity deviations	54	158	144	500
Survey selectivity deviations	51	76	140	268
Total deviations	166	498	399	1,191
Total	195	532	452	1,252

[1] Recruitment deviations include estimated numbers-at-ages 1-19 in the first year of the short time-series models, as these are implemented as recruitments depreciated to each age via natural mortality. In addition, recruitment deviations for three projection years are also included in the totals.

Table 7. Sample size diagnostics for age composition data by model and model fleet.

	Average input	Harmonic mean effective	Francis weight (multipier)
Coastwide short			
Fishery	88	403	0.99
Discards	6	274	30.39
Bycatch	5	67	9.79
Sport	5	108	10.80
Survey	325	962	1.20
Coastwide long			
Fishery	95	357	1.58
Discards	6	244	24.83
Bycatch	5	27	1.15
Sport	5	127	6.02
Survey	94	191	0.88
AAF short			
Area 2 Fishery	320	590	1.49
Area 3 Fishery	159	322	0.28
Area 4 Fishery	52	65	1.13
Area 4B Fishery	24	102	2.86
Discards	6	231	20.44
Bycatch	5	48	13.24
Sport	5	116	4.53
Area 2 Survey	242	457	0.49
Area 3 Survey	210	516	0.21
Area 4 Survey	90	192	0.81
Area 4B Survey	31	132	0.64
AAF long			
Area 2 Fishery	145	254	2.07
Area 3 Fishery	83	243	0.64
Area 4 Fishery	16	42	1.73
Area 4B Fishery	19	90	2.55
Discards	6	244	19.38
Bycatch	5	29	1.54
Sport	5	105	4.30
Area 2 Survey	134	221	0.48
Area 3 Survey	102	131	0.28
Area 4 Survey	90	191	0.66
Area 4B Survey	31	131	0.52

Table 8. Select parameter estimates (final value and approximate 95% confidence interval) from each assessment model.

	Model			
	Coastwide Short	Coastwide Long	AAF Short	AAF Long
Biological				
Female *M*	0.15 (Fixed)	0.202 (0.178-0.225)	0.15 (Fixed)	0.148 (0.126-0.170)
Male *M*	0.138 (0.126-0.149)	0.156 (0.139-0.173)	0.129 (0.121-0.138)	0.129 (0.114-0.144)
Log(R_0)*	9.42 (9.11-9.74)	10.75 (10.50-11.00)	9.92 (9.65-10.20)	10.11 (9.84-10.39)
R_1 offset	-0.350 (-0.604--0.095)	NA	-0.253 (-0.466--0.040)	NA
Env. Link (β)	NA	0.308 (-0.033-0.649)	NA	0.522 (0.224-0.820)
Survey Log(q) Δ1984	NA	0.85 (0.54-1.15)	NA	*A2:*0.74 (0.56-0.92) *A3:*1.43 (1.25-1.61)
Fishery Log(q) Δ1984	NA	0.52 (0.36-0.68)	NA	*A2:0.50 (0.31-0.68)* *A3:1.07 (0.94-1.20)* *A4:0.70 (0.50-0.90)* *A4B:*0.37 (0.18-0.56)

* Log(R_0) values are not comparable for the two short time-series models as there is no constraint that this represent the central tendency of the S-R function at equilibrium. S-R related calculations were performed externally to these models.

Table 9. Qualitative sensitivity ranking of halibut wastage and bycatch estimates to sources of uncertainty, as well as the scale of the removals relative to total removals from the stock (magnitude).

Source	DMR (process)	Observer coverage (sampling)	Potential bias (non-sampling)	Magnitude
Wastage				
Alaska	High	High	High	Moderate
Canada	High	Low	NA	Low
Area 2A	High	NA	High	Low
Bycatch				
Bering Sea trawl	Low	Low	Low	High
Bering Sea H&L	High	Low	Low	Moderate
Gulf of Alaska trawl	Low	High	High	High
Gulf of Alaska H&L	High	High	High	Moderate
Canada trawl	Low	Low	NA	Low
Area 2A trawl	Low	Low	NA	Low
Area 2A H&L	High	Moderate	Moderate	Low

Table 10. Summary of individual model and integrated distributions for 2016 spawning biomass (millions pounds).

	Percentile		
Models	2.5%	50%	97.5%
Coastwide Long	131.2	182.1	233.0
Coastwide Short	145.6	185.3	225.0
AAF Long	198.4	240.8	283.2
AAF Short	231.5	278.0	324.4
Integrated distribution (equal weighting 1:1:1:1)	144.9	217.4	308.5

Table 11. Median integrated population (Mlb) and fishing intensity estimates (based on median Spawning Potential Ratio) from the 2014 and preliminary 2015 assessments.

	2014 results			2015 results		
Year	Spawning biomass	Fishing intensity ($F_{XX\%}$)	Exploitable biomass	Spawning biomass	Fishing intensity ($F_{XX\%}$)	Exploitable biomass
1996	584.6	49%	779.2	483.8	47%	655.1
1997	605.7	43%	809.6	520.3	42%	708.0
1998	591.8	42%	762.7	512.8	40%	668.1
1999	567.1	40%	746.8	496.9	38%	662.0
2000	529.5	40%	688.3	468.4	37%	614.4
2001	483.9	38%	603.0	432.4	35%	540.7
2002	434.5	34%	532.2	391.4	32%	477.6
2003	382.6	30%	460.5	346.3	29%	415.0
2004	339.5	28%	403.6	309.2	26%	365.3
2005	299.5	26%	352.6	275.4	25%	321.8
2006	266.7	26%	307.9	248.0	25%	283.2
2007	241.5	25%	266.9	227.5	25%	248.3
2008	224.4	25%	236.3	213.9	25%	222.1
2009	204.6	26%	203.9	196.5	26%	193.7
2010	197.8	27%	186.4	190.1	26%	177.8
2011	195.3	31%	175.6	188.2	31%	168.2
2012	197.2	35%	169.2	190.5	36%	163.3
2013	203.9	38%	168.8	197.4	39%	163.9
2014	208.5	43%	169.7	202.7	44%	165.7
2015	215.1	NA	180.6	209.6	51%	176.6
2016	NA	NA	NA	217.4	NA	187.4

Table 12. Decision table framework for the integrated model results.

				Stock Trend				Stock Status				Fishery Trend				Fishery Status
				Spawning biomass				Spawning biomass				Fishery CEY from the harvest policy				Harvest rate
				in 2017		in 2019		in 2017		in 2019		in 2017		in 2019		in 2016
2016	Total removals (M lb)	Fishery CEY (M lb)	Fishing intensity	is less than 2016	is 5% less than 2016	is less than 2016	is 5% less than 2016	is less than 30%	is less than 20%	is less than 30%	is less than 20%	is less than 2016	is 10% less than 2016	is less than 2016	is 10% less than 2016	is above target
No removals	0.0	0.0	0.0%													
FCEY = 0		0.0														
Blue Line																
				a	b	c	d	e	f	g	h	i	j	k	l	m

Table 13. Movement rates for halibut in the IPHC's PIT tagging study; reproduced from Valero and Webster (2012; Table 14).

Area in	Area in year i+1							
year i	**4D**	**4B**	**4A**	**3B**	**3A**	**2C**	**2B**	**2A**
4D	**0.924**	0.000	0.062	0.000	0.003	0.001	0.010	0.000
4B	0.002	**0.967**	0.000	0.000	0.021	0.005	0.005	0.000
4A	0.000	0.014	**0.792**	0.053	0.097	0.016	0.025	0.003
3B	0.000	0.000	0.003	**0.887**	0.101	0.005	0.004	0.000
3A	0.000	0.000	0.000	0.046	**0.947**	0.003	0.003	0.000
2C	0.000	0.000	0.000	0.000	0.024	**0.898**	0.067	0.012
2B	0.000	0.000	0.004	0.000	0.004	0.009	**0.970**	0.014
2A	0.000	0.000	0.000	0.000	0.000	0.008	0.110	**0.882**

Table 14. Movement rates estimated for juvenile halibut in Areas 2 and 3; reproduced from Hilborn et al. (1995; Table 14).

Movement parameters(m_{ij}):

	3B	3A	2C	2B	2A
3B	0.708				
3A	0.231	0.894			
2C	0.061	0.080	0.698		
2B		0.026	0.302	0.942	
2A				0.058	1.000

Figures

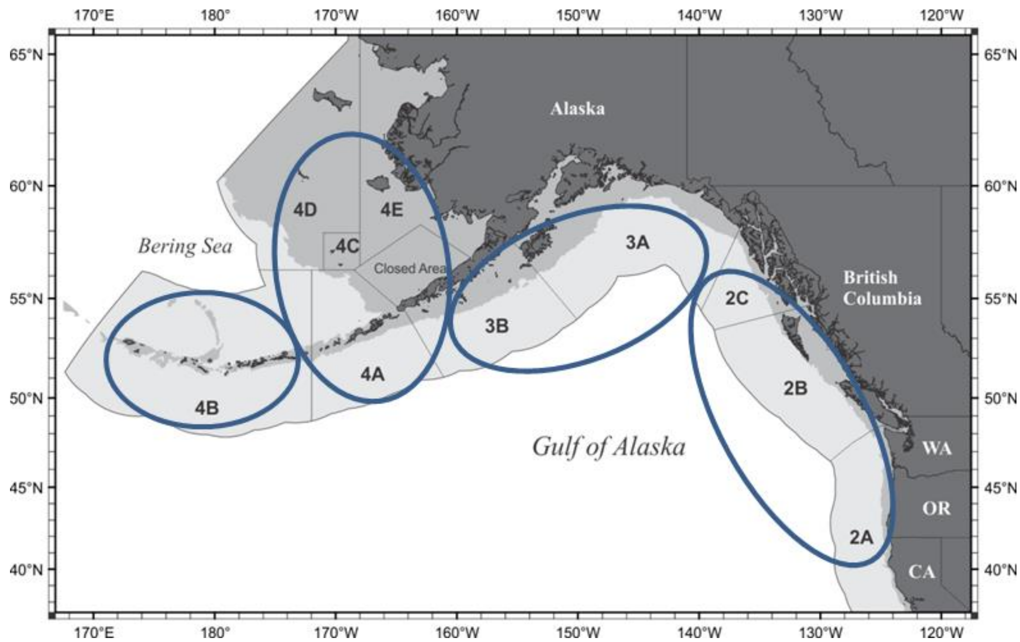


Figure 1. Current IPHC regulatory areas. Shaded region denotes the Exclusive Economic Zones of the U.S. and Canada. Circles denote aggregated geographic areas (2, 3, 4, and 4B) used to partition the data in the AAF models.

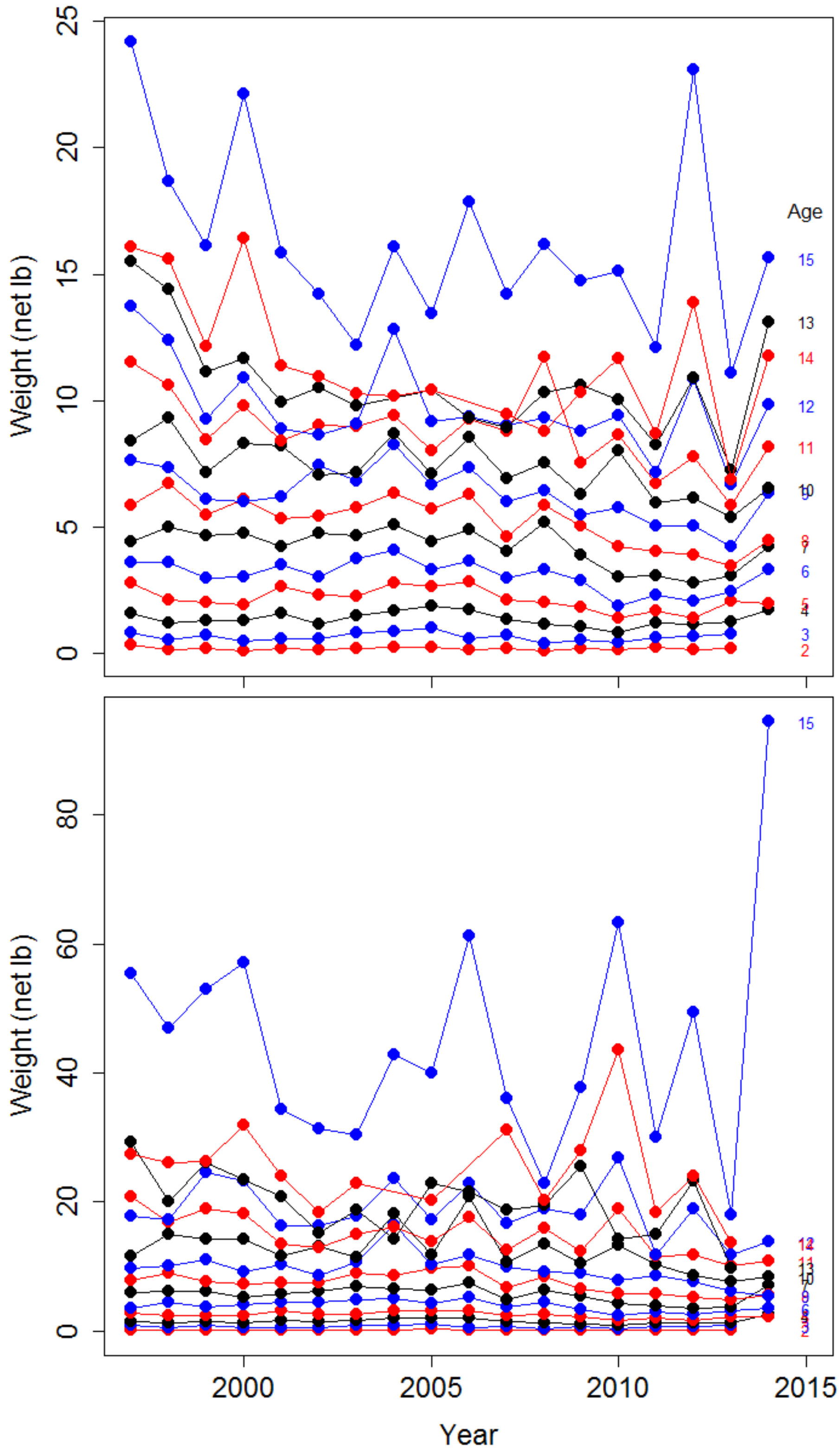


Figure 2. Raw average weight-at-age for male (upper panel) and female (lower panel) halibut, age 2-15+, caught by NMFS trawl surveys from 1997- 2014 (2014 was in the process of being finalized, and contained only a very small number of samples).

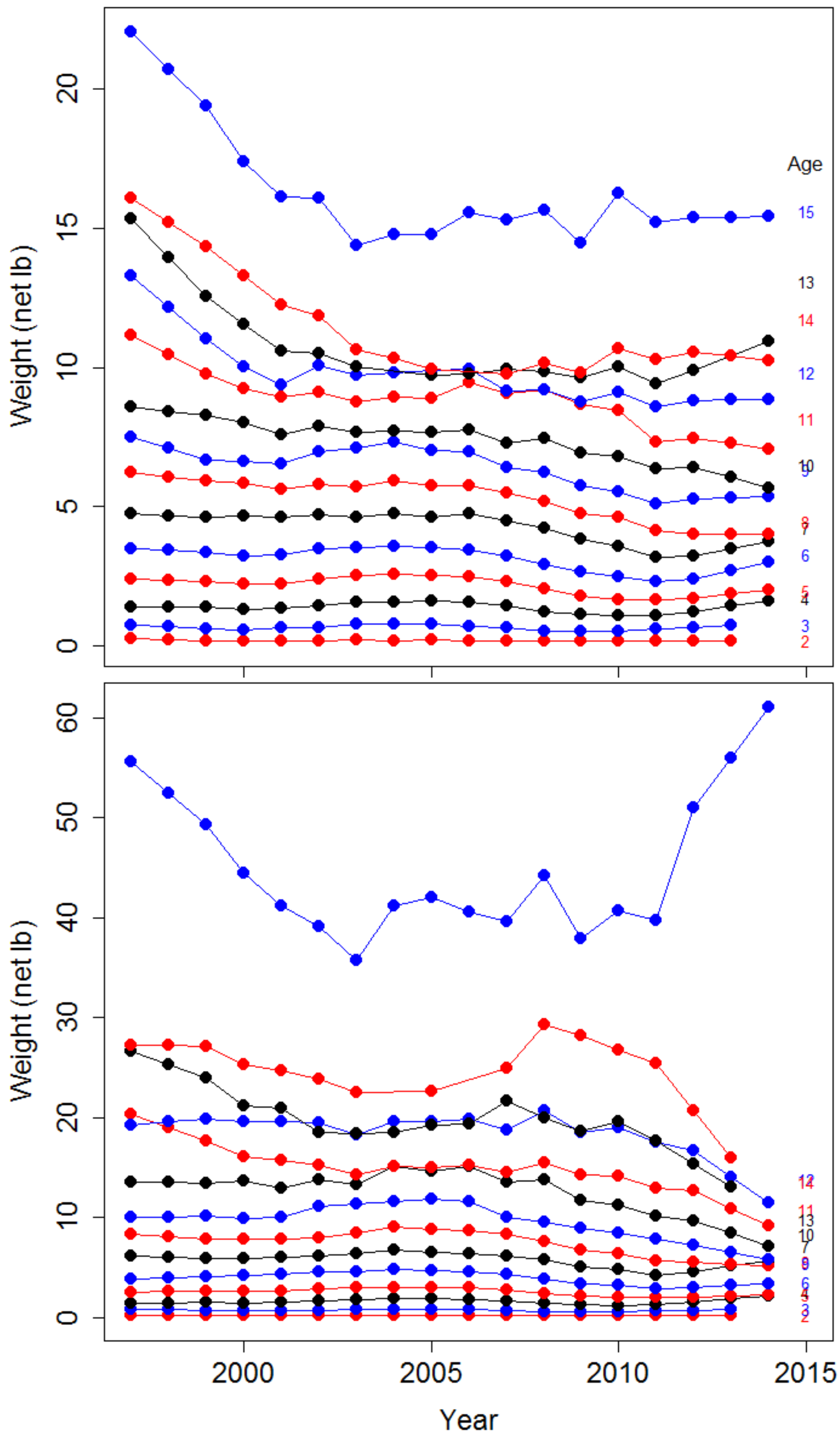


Figure 3. Smoothed average weight-at-age for male (upper panel) and female (lower panel) halibut, age 2-15+, caught by NMFS trawl surveys from 1997- 2014 (2014 was in the process of being finalized, and contained only a very small number of samples).

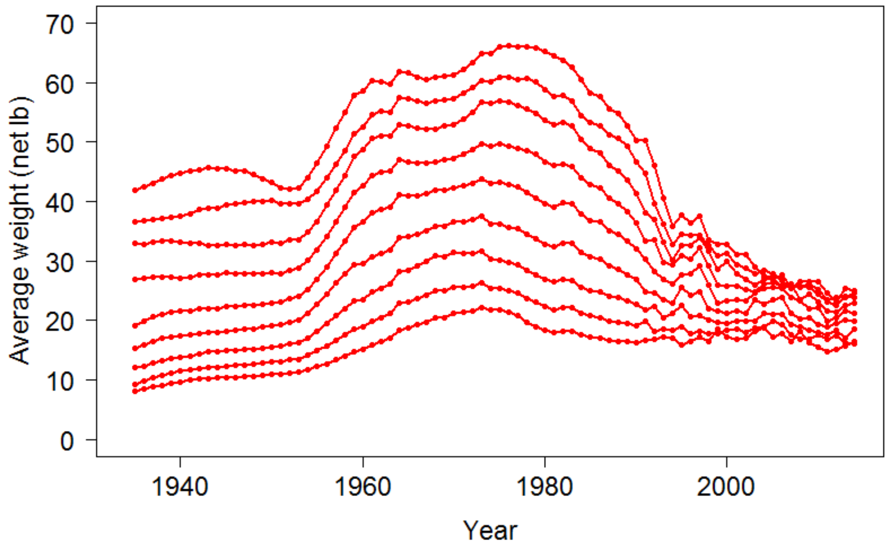


Figure 4. Estimated coastwide trends in weight at age for ages 8-16 from 1935-2014.

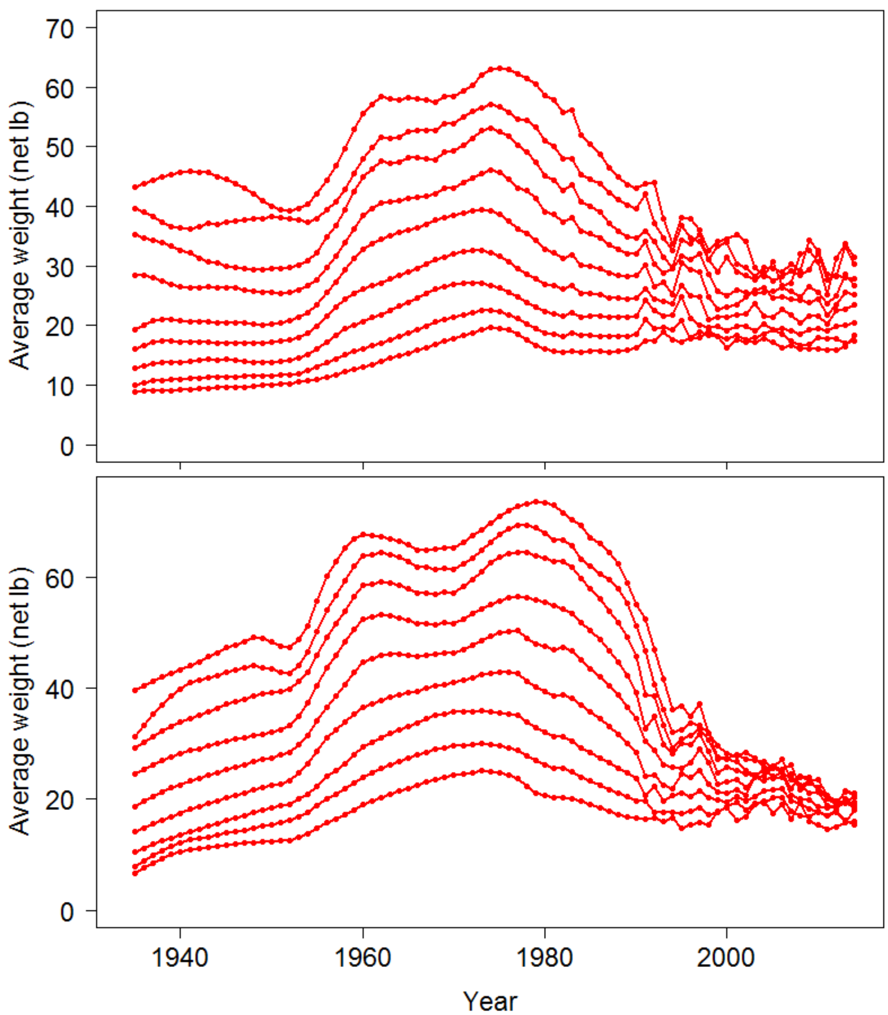


Figure 5. Estimated trends in weight at age for ages 8-16 for Area 2 (top) and Area 3 (bottom) from 1935-2014.

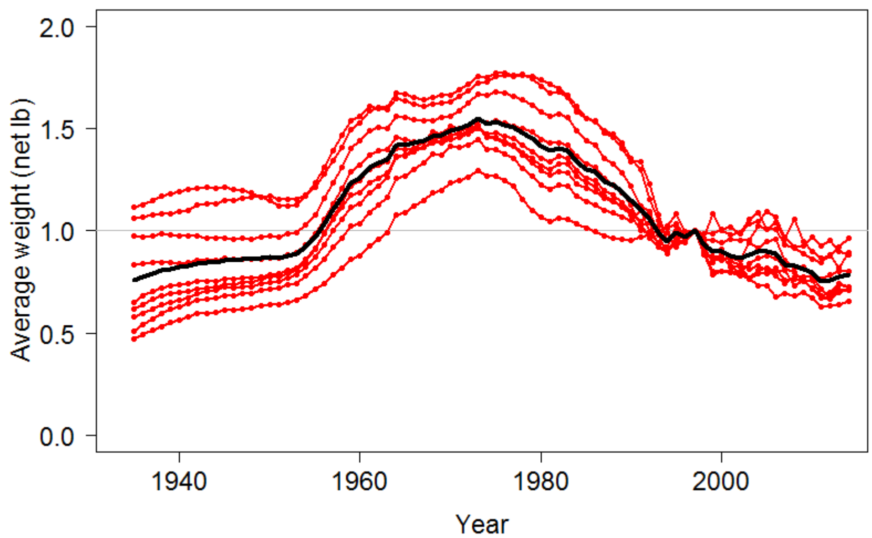


Figure 6. Estimated coastwide relative trends in weight at age for ages 8-16 (relative to 1997) from 1935-2014.

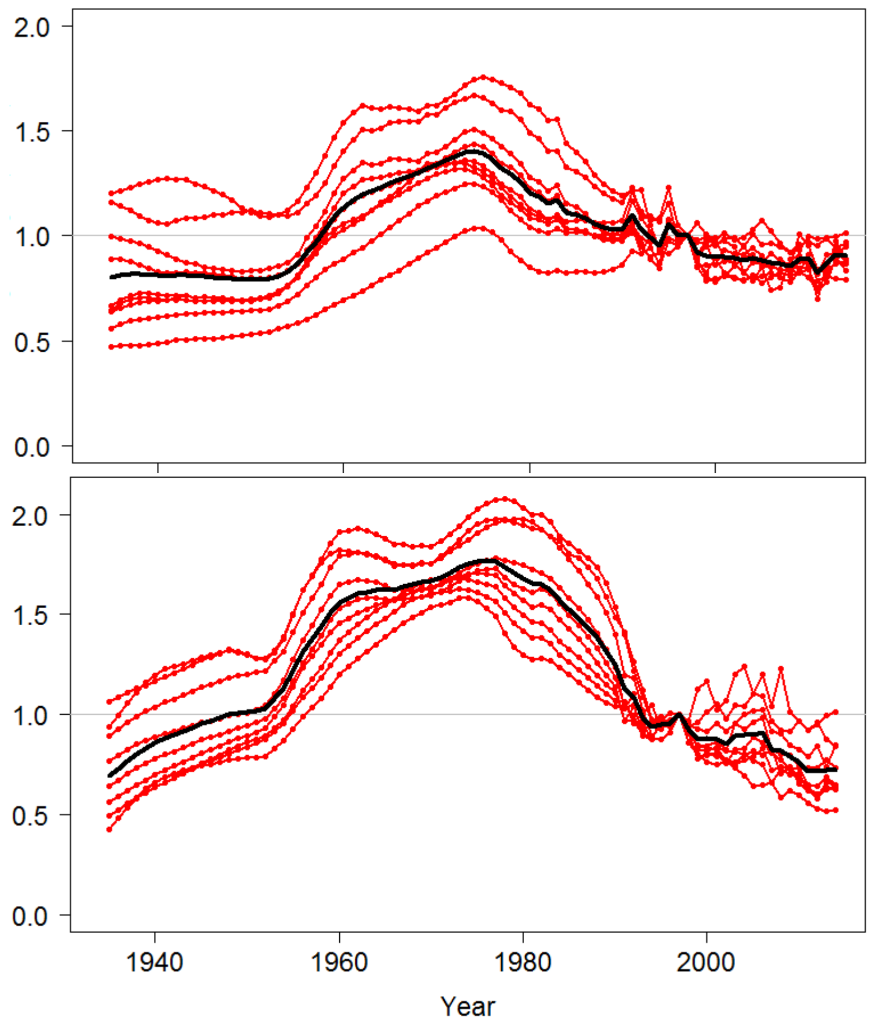


Figure 7. Estimated relative trends in weight at age for ages 8-16 (relative to 1997) for Area 2 (top) and Area 3 (bottom) from 1935-2014.

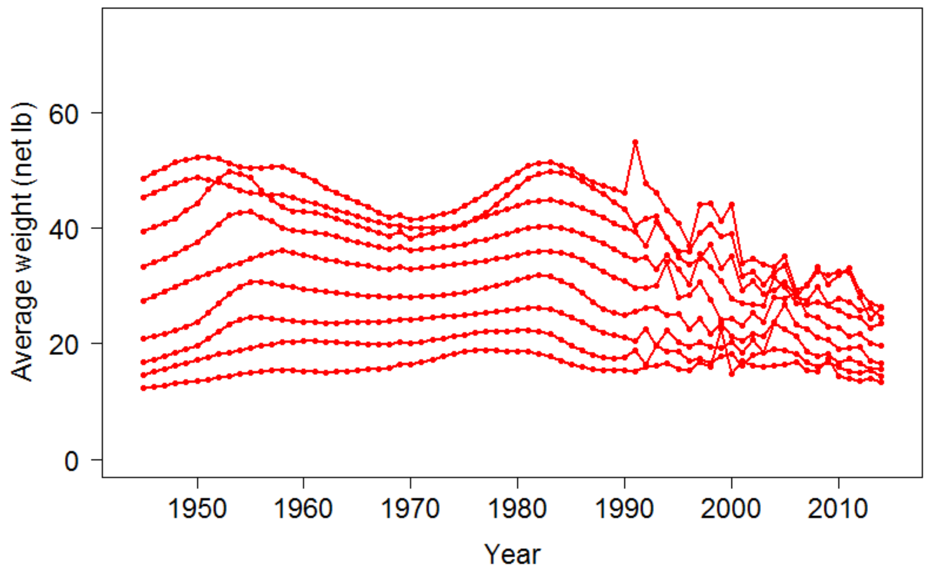


Figure 8. Estimated trends in weight at age for ages 8-16 for Area 4 from 1945-2014.

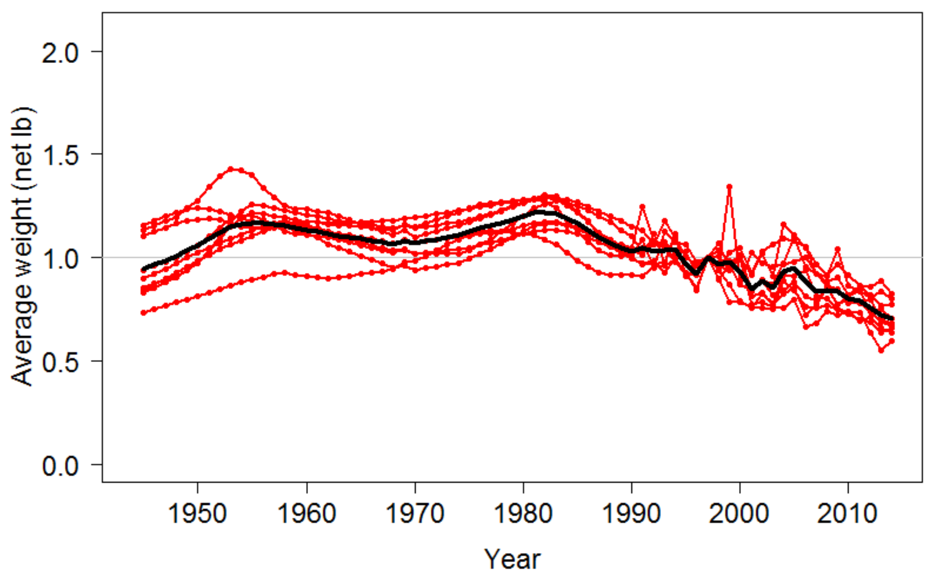


Figure 9. Estimated relative trends in weight at age for ages 8-16 (relative to 1997) for Area 4 from 1945-2014.

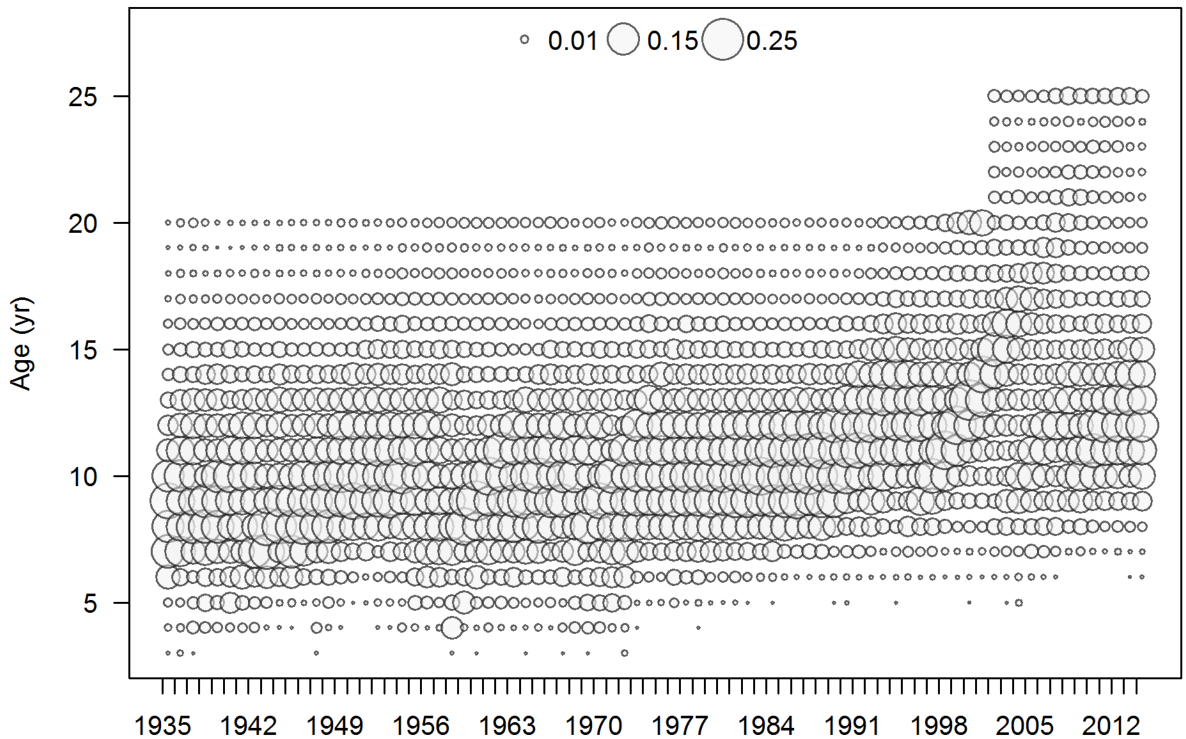


Figure 10. Coastwide age composition data for halibut captured by the commercial fishery.

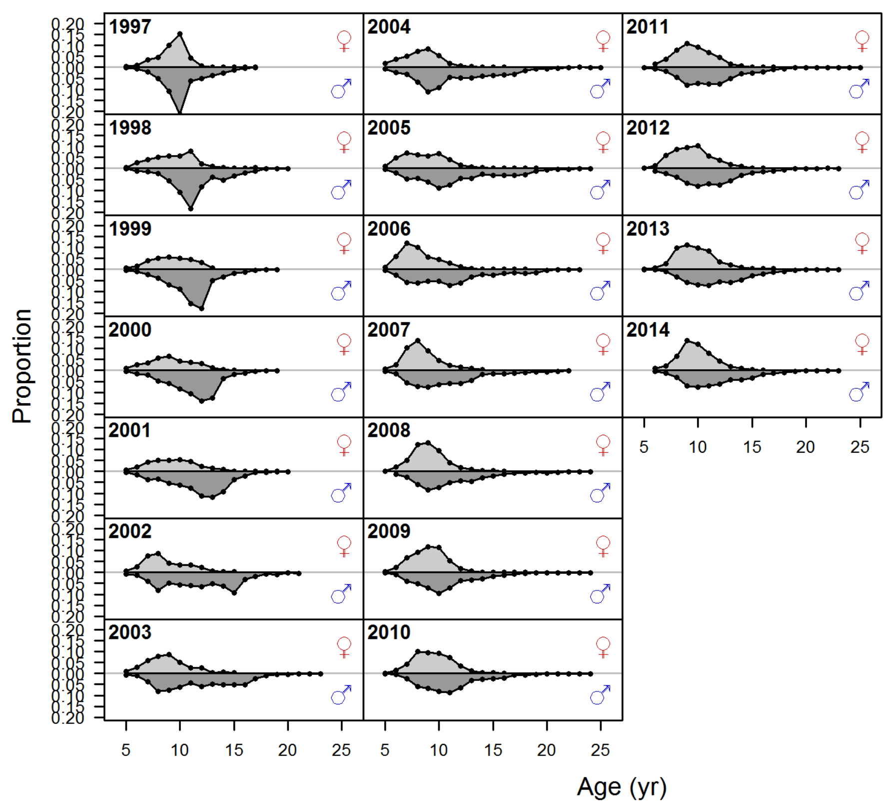


Figure 11. Coastwide age composition data for sublegal halibut captured by the setline survey. Female proportions are displayed as positive values on the y-axis, males as negative values.

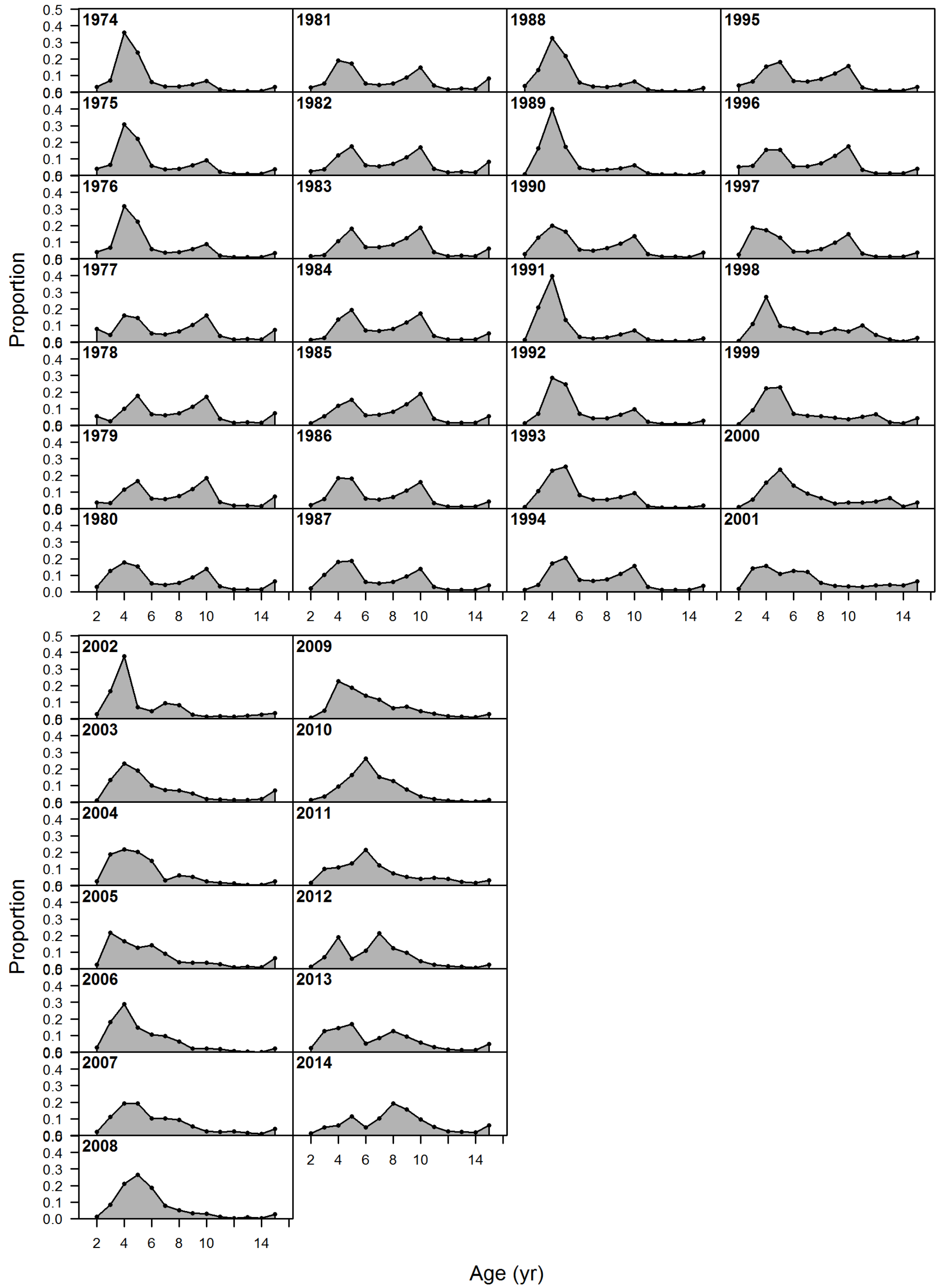


Figure 12. Coastwide age composition data created from bycatch lengths (1974-2014) and the age-length key based on NMFS trawl surveys. Some years (e.g., the early 1980s) appear to contain replicated distributions.

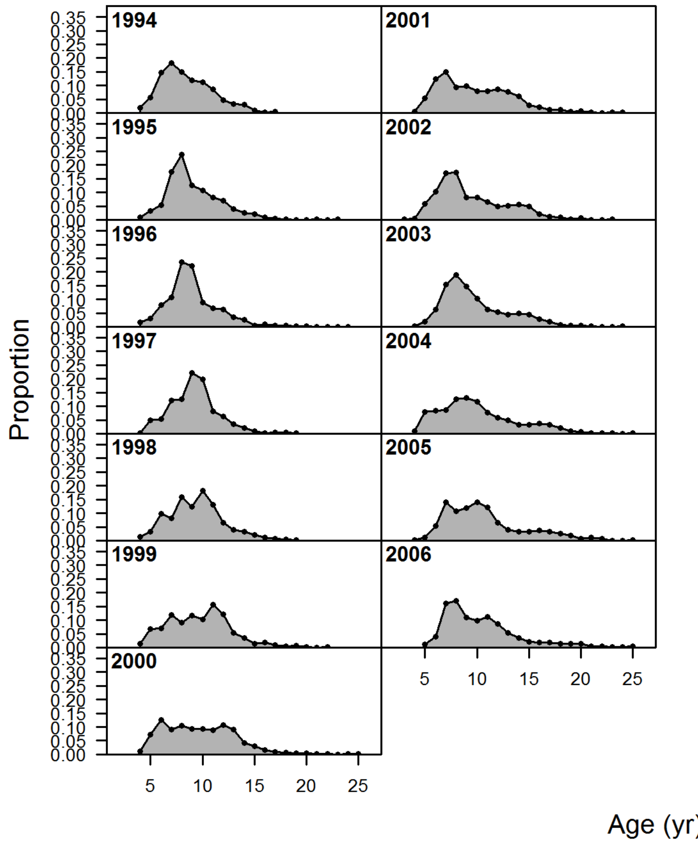


Figure 13. Age composition data from the recreational fishery in Area 3.

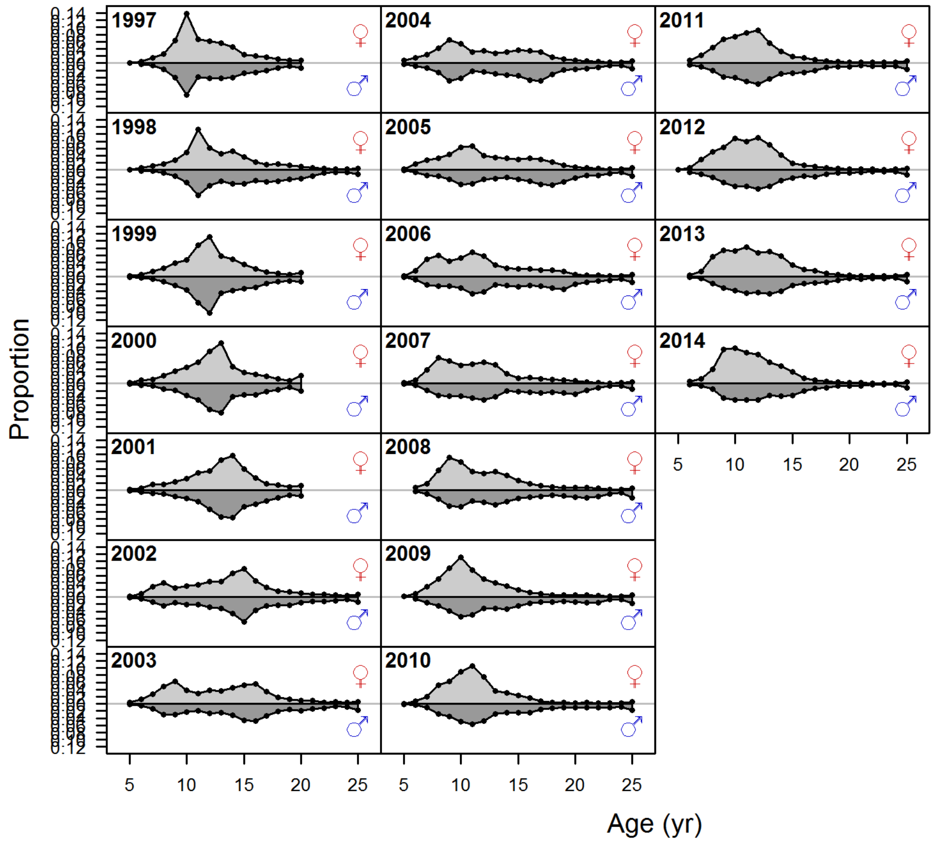


Figure 14. Coastwide age composition data from the setline survey.

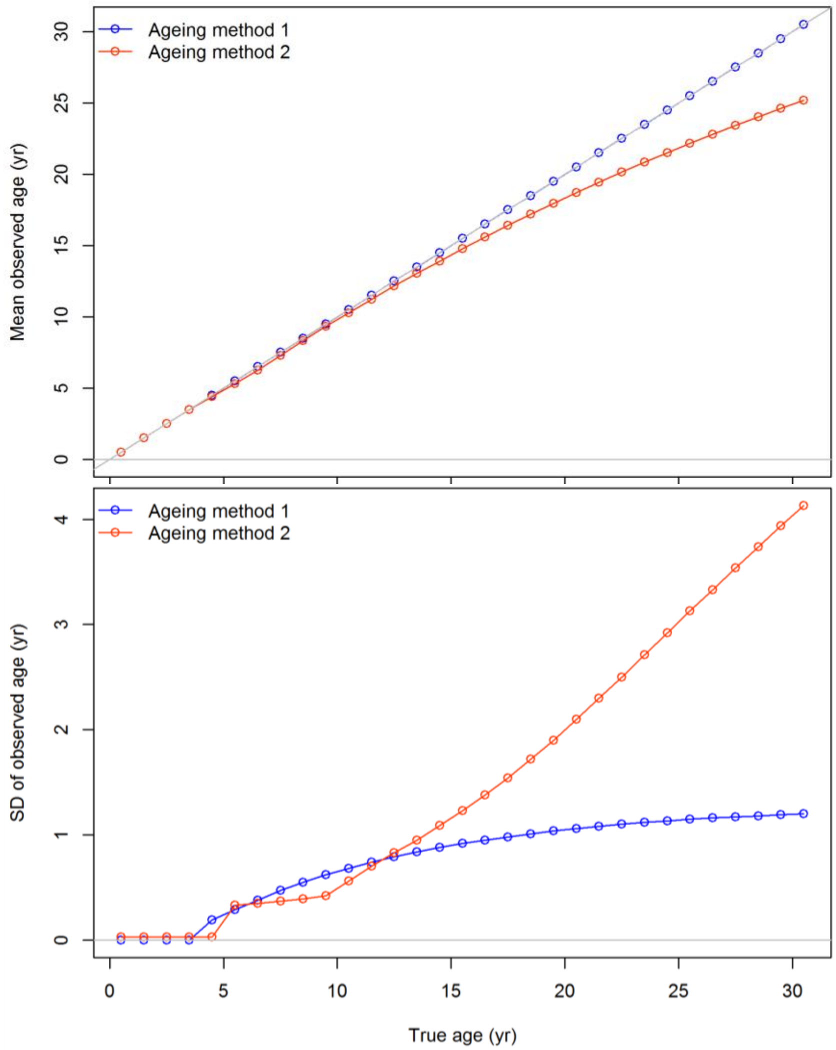


Figure 15. Degree of bias (upper panel) and imprecision (lower panel) estimated externally and assigned to break-and-bake (method 1) and surface ageing (method 2) in all assessment models.

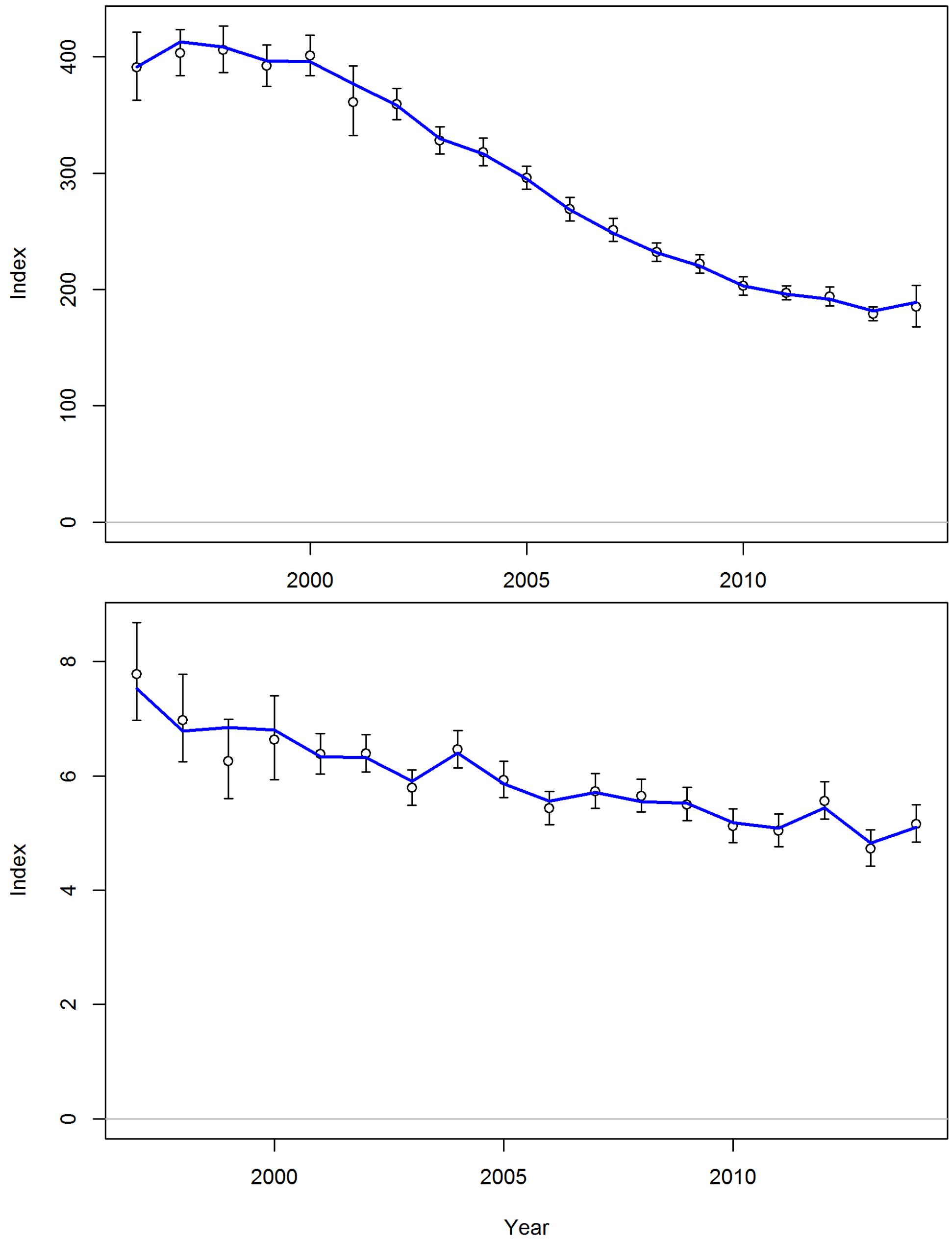


Figure 16. Fit to the index of abundance from the commercial fishery (upper panel, weight-per-unit-effort) and the setline survey (lower panel, numbers-per-unit-effort) in the coastwide short model.

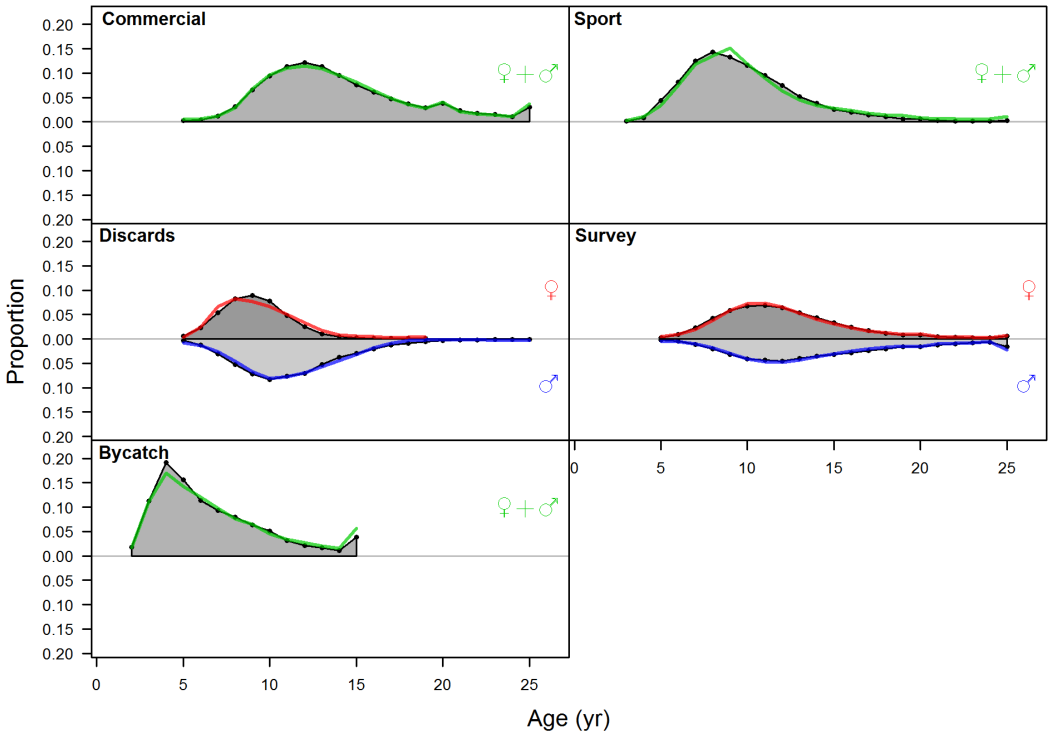


Figure 17. Aggregate fit (observed and expected values summed across all years) to the age data for each fleet in the coastwide short model.

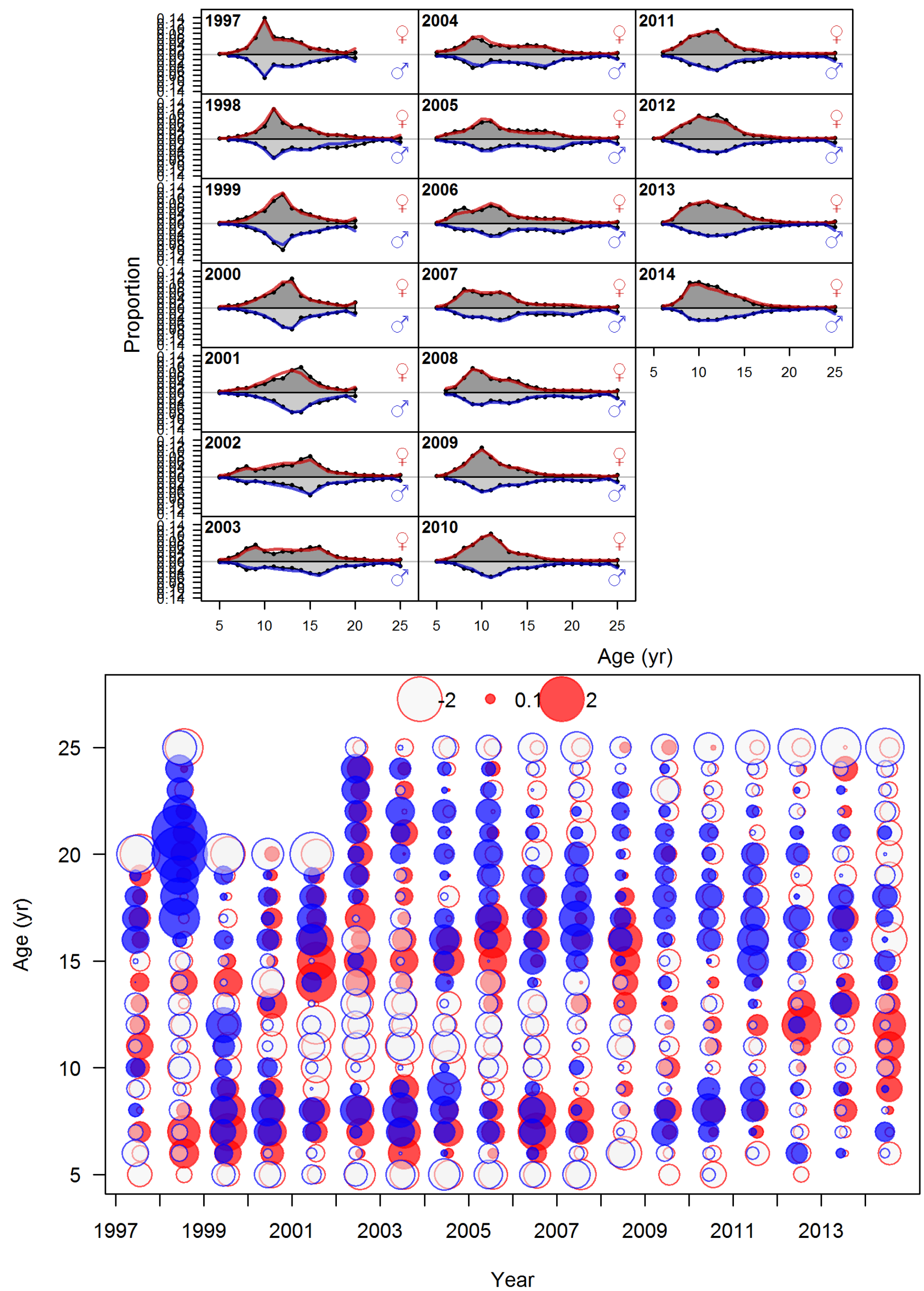


Figure 18. Fit and standardized residuals to the age data for the setline survey in the coastwide short model. Red and blue denote female and male residuals respectively.

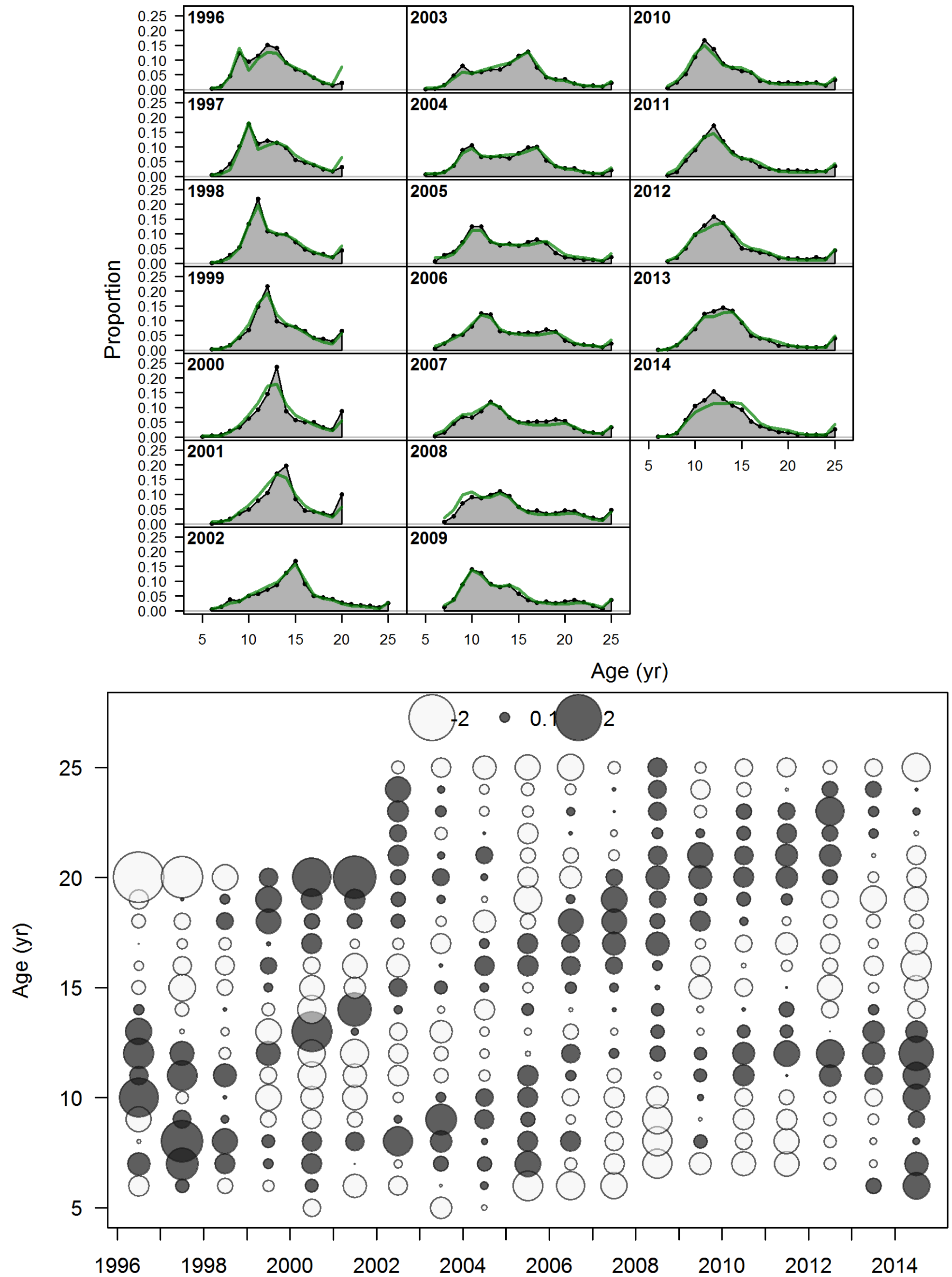


Figure 19. Fit and standardized residuals to the age data for the commercial fishery in the coastwide short model.

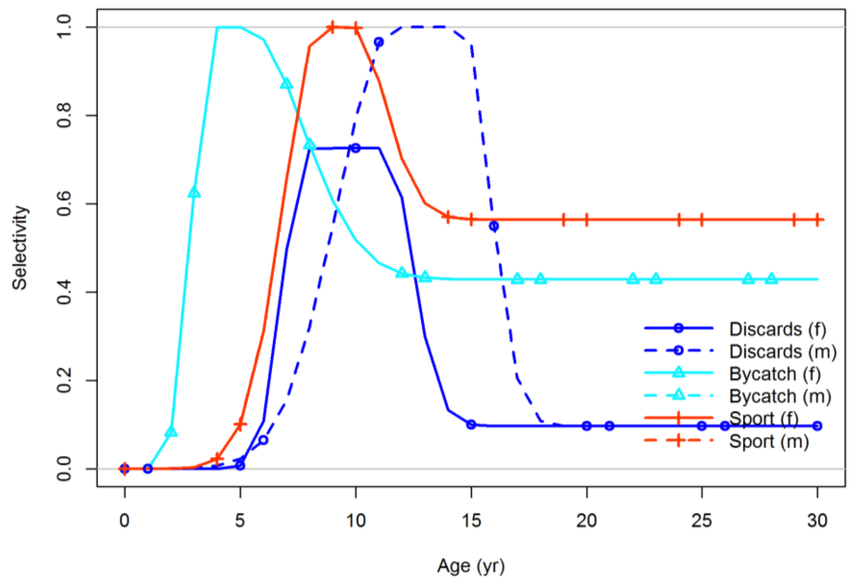


Figure 20. Selectivity patterns estimated for the discards (male and female curves), sport and bycatch fleets in the coastwide short model.

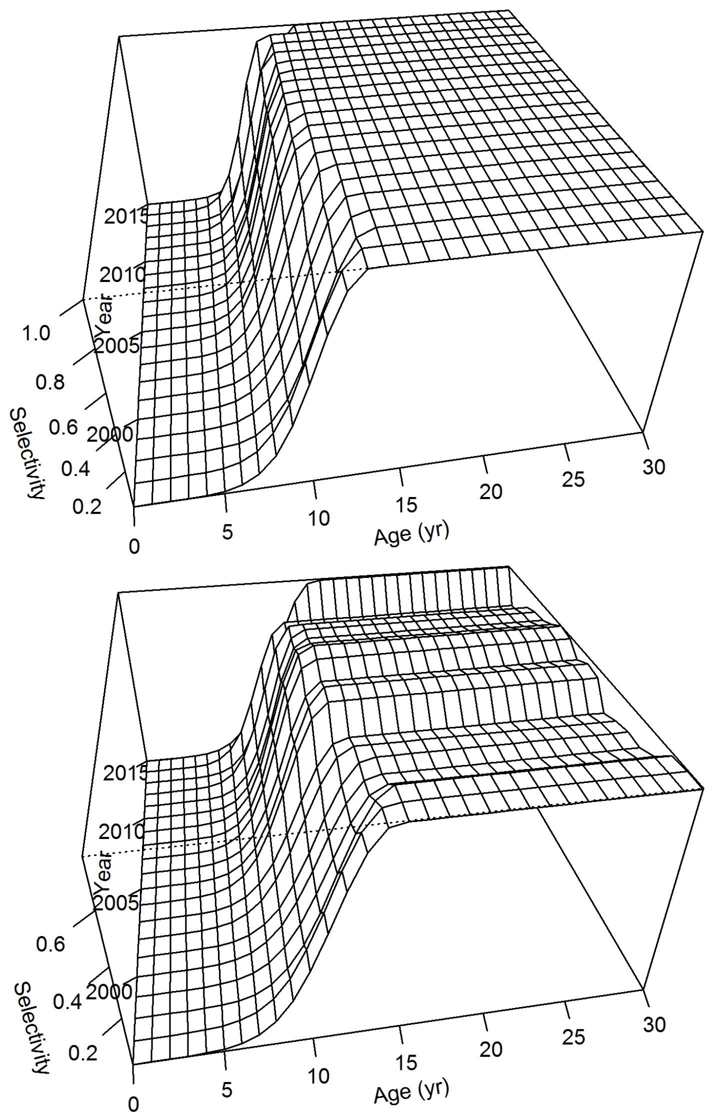


Figure 21. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the setline survey in the coastwide short model.

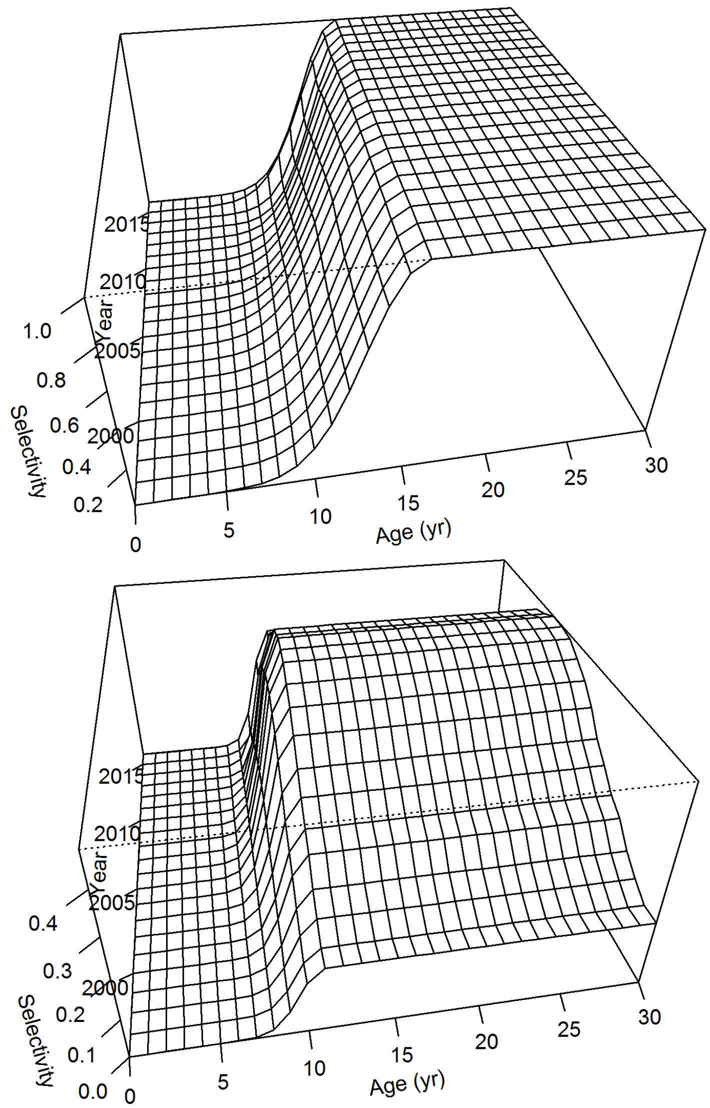


Figure 22. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the directed commercial fishery in the coastwide short model.

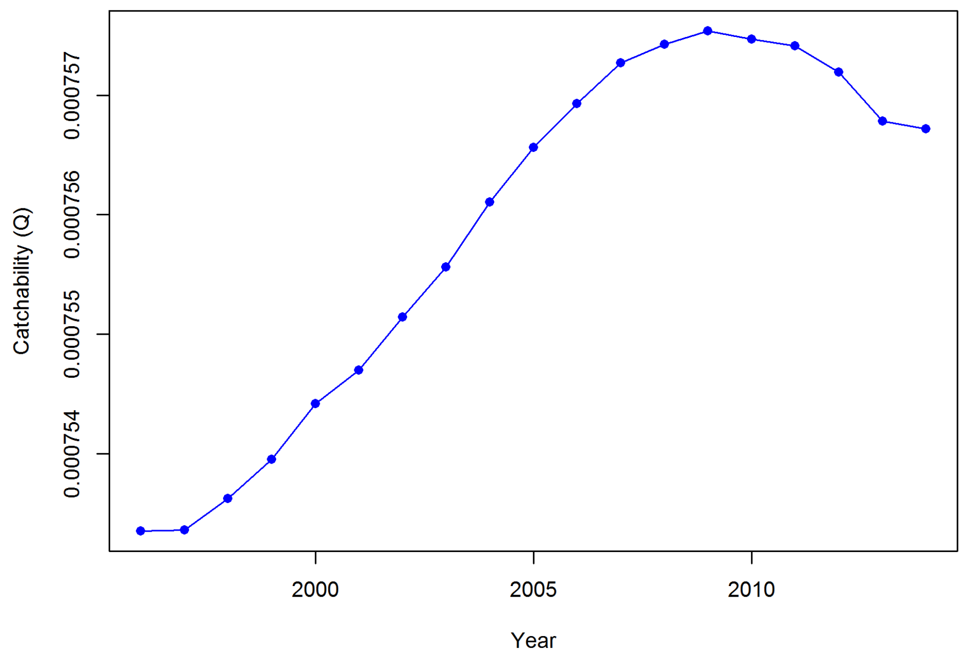


Figure 23. Estimated trend in fishery catchability in the coastwide short model.

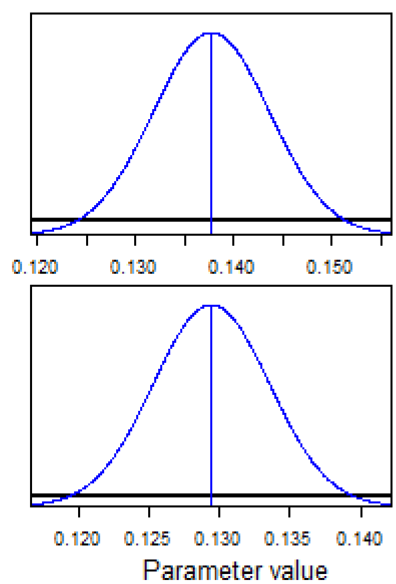


Figure 24. Prior (thick line) and MLE-based distributions for male natural mortality from the coastwide short model (top panel), and the AAF short model (bottom panel).

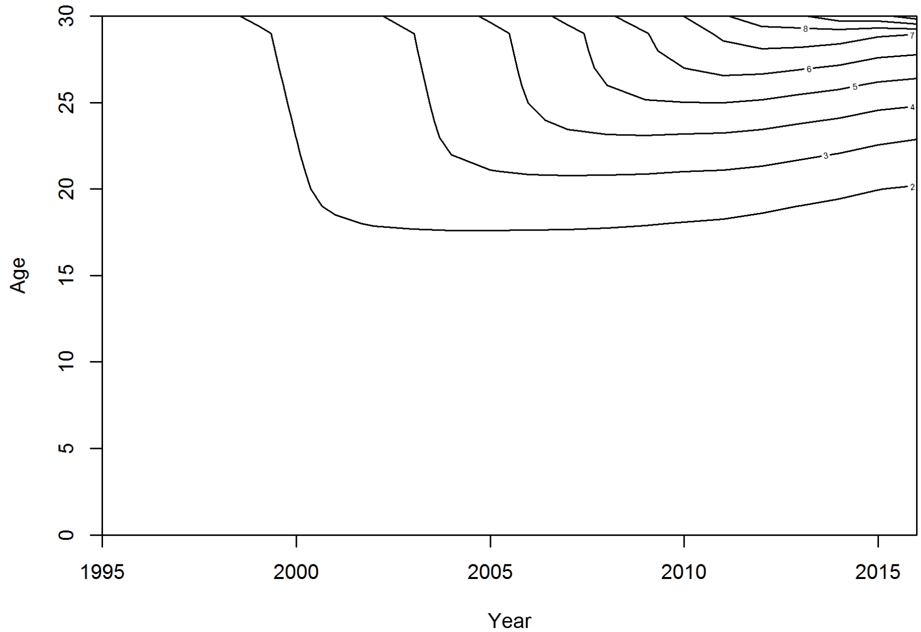


Figure 25. Population sex ratio (numbers of males/females) contours estimated by the coastwide short model.

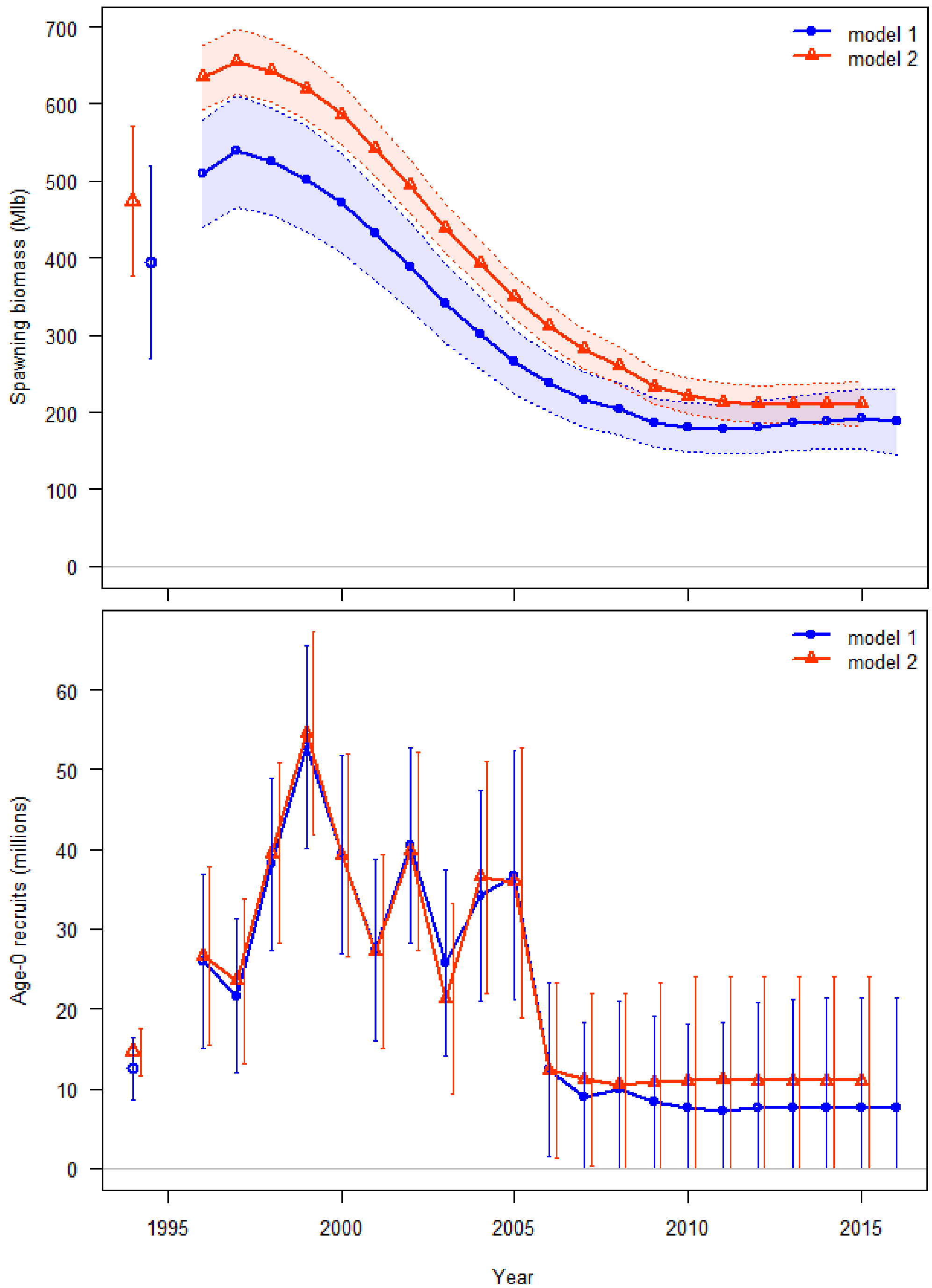


Figure 26. Comparison of the 2014 result, and the preliminary 2015 result including all updates to data and model structure for the CW short model.

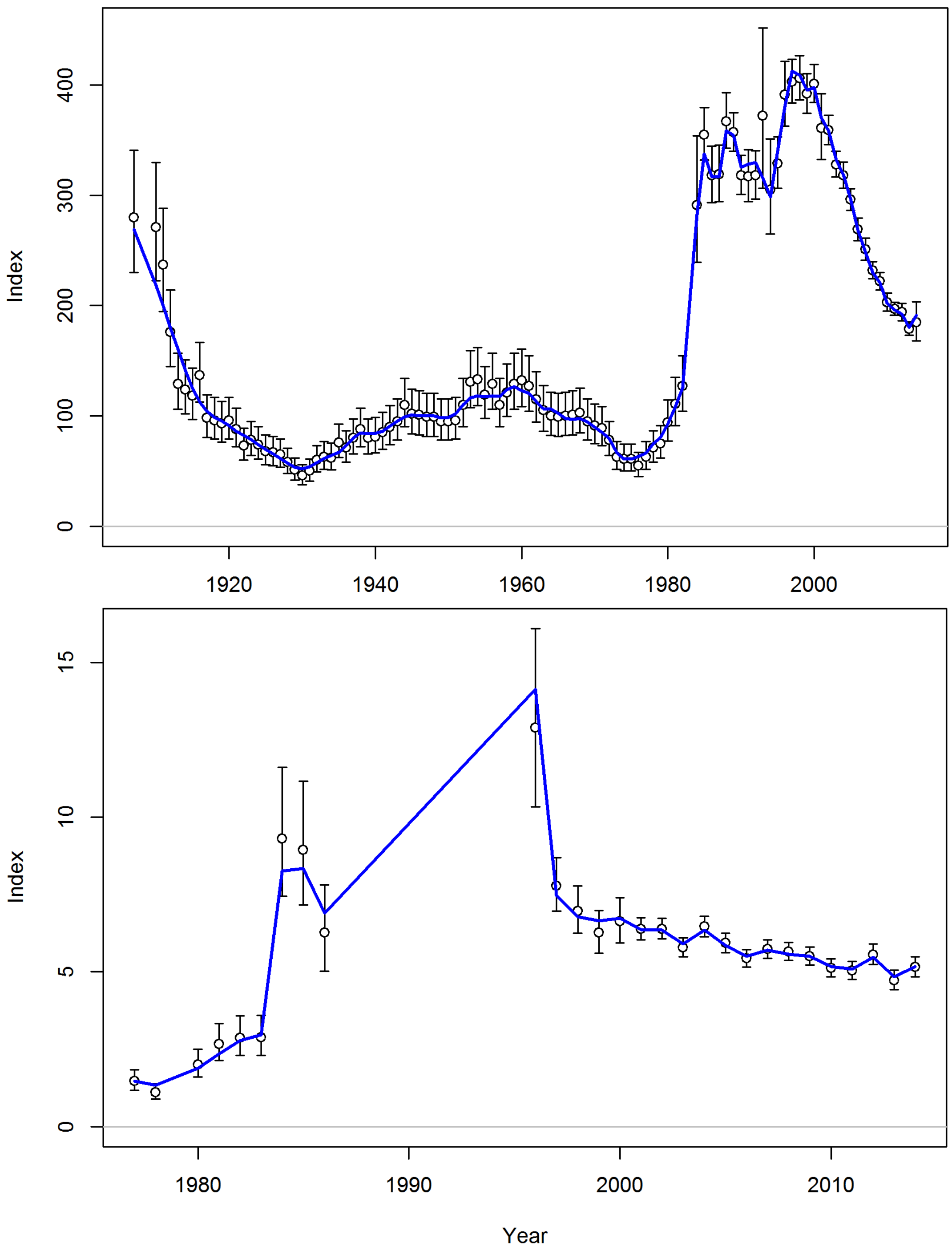


Figure 27. Fit to the index of abundance from the commercial fishery (upper panel, weight-per-unit-effort) and the setline survey (lower panel, numbers-per-unit-effort) in the coastwide long model. Note that there are unconstrained catchability breaks for the fishery in 1984, and the survey in 1982, 1984, and 1997.

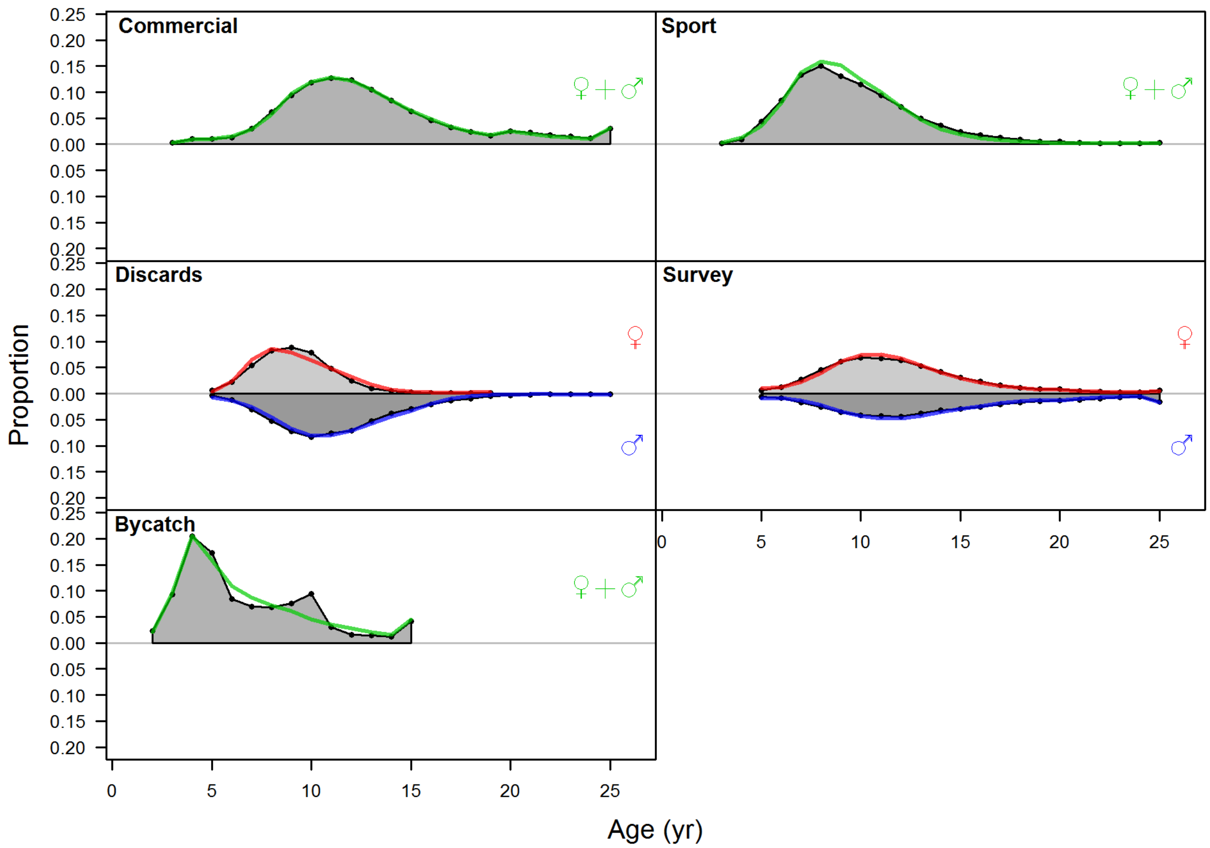


Figure 28. Aggregate fit (observed and expected values summed across all years) to the age data for each fleet in the coastwide long model.

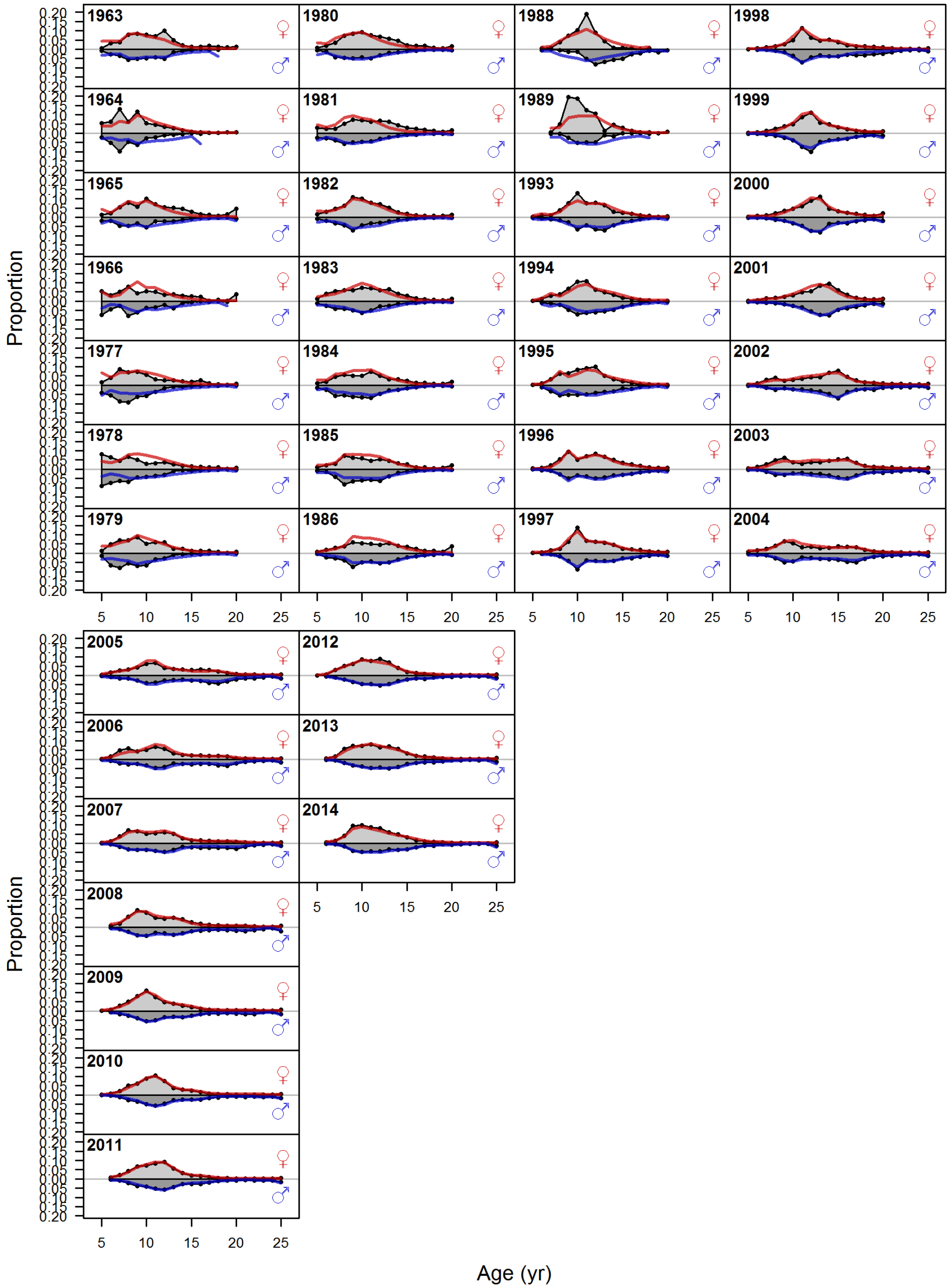


Figure 29. Fit to the age data for the survey in the coastwide long model.

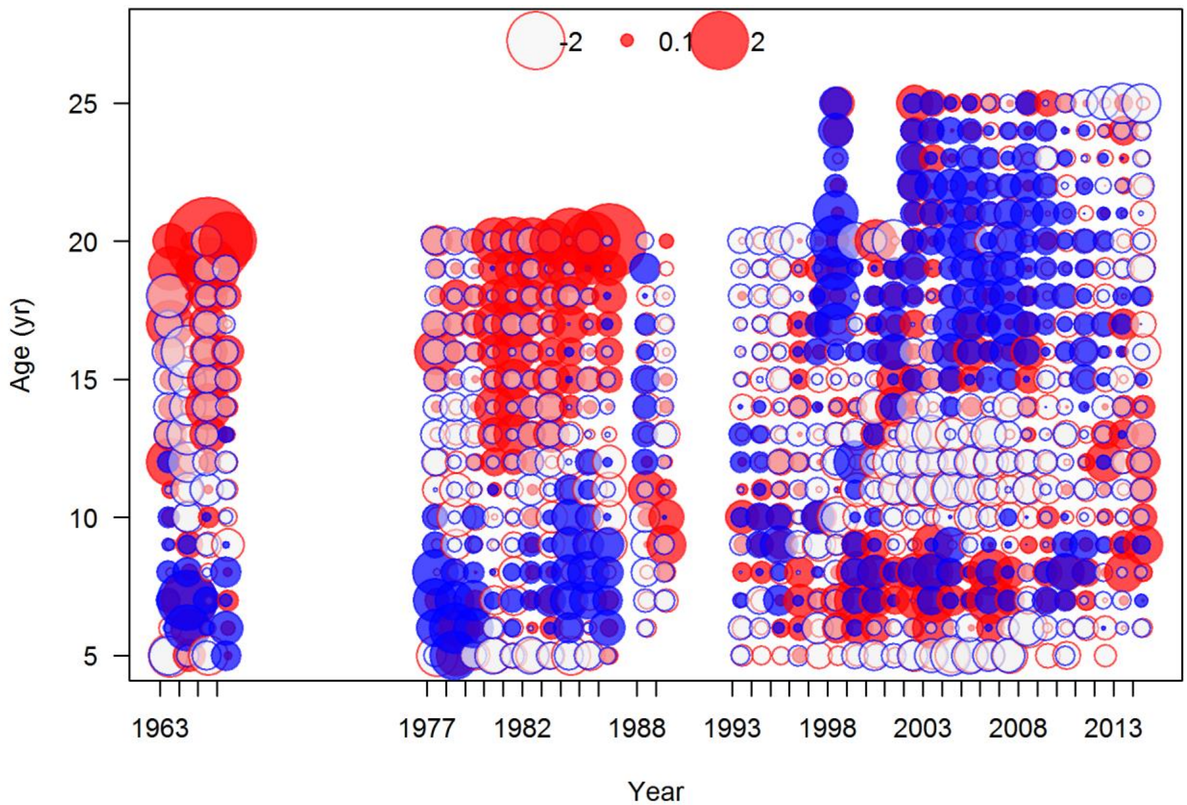


Figure 30. Standardized residuals to the age data for the survey in the coastwide long model. Red and blue denote female and male residuals respectively.

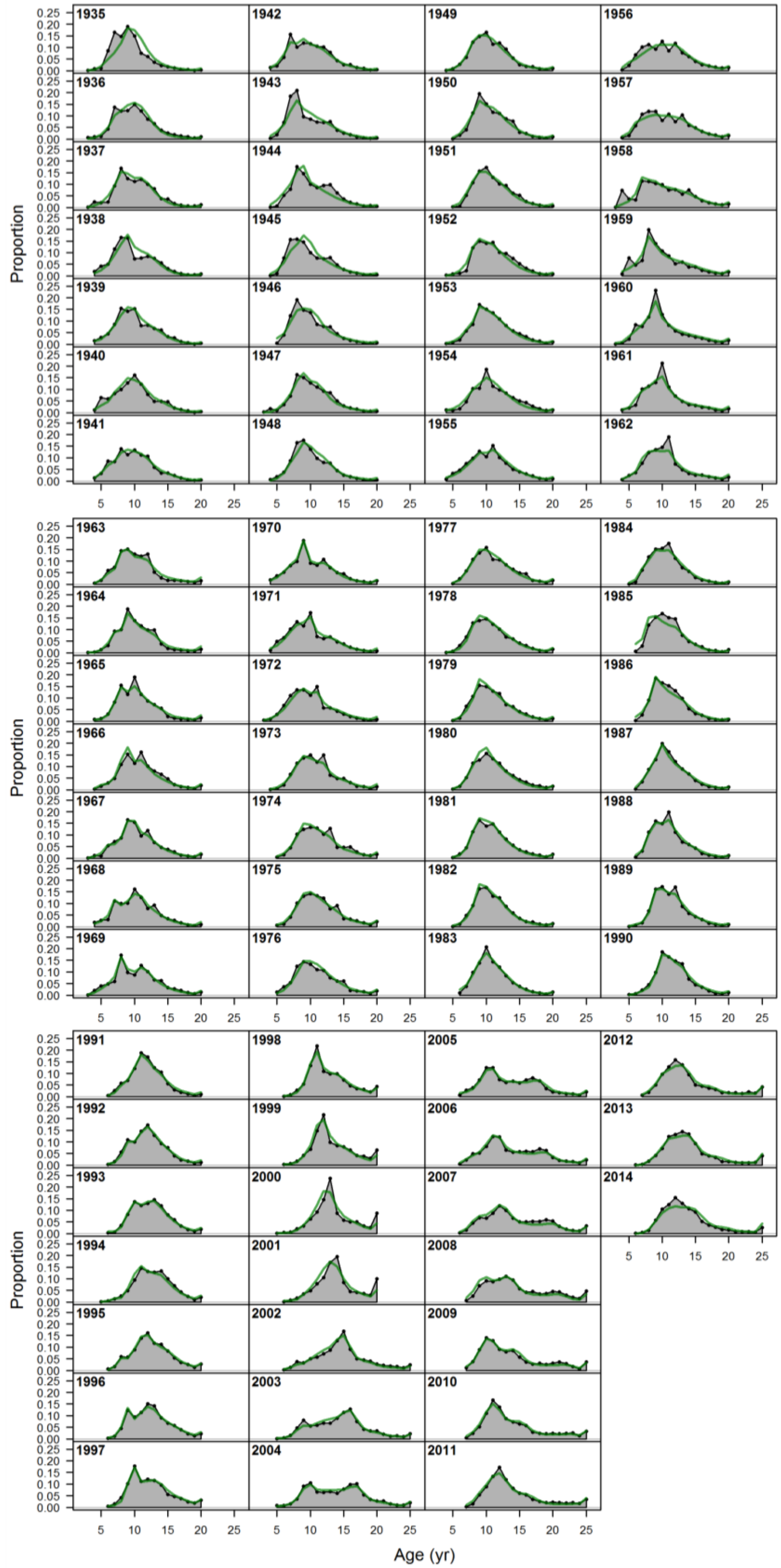


Figure 31. Fit to the age data for the commercial fishery in the coastwide long model.

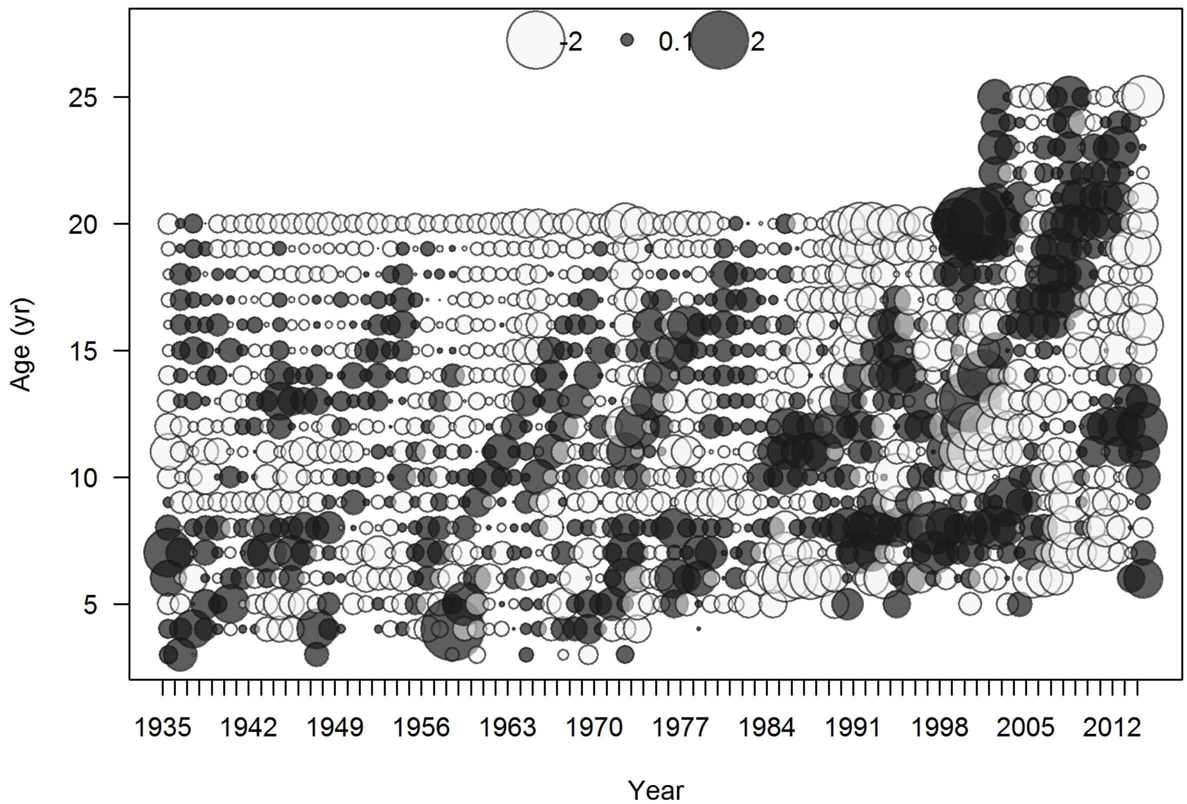


Figure 32. Standardized residuals to the age data for the commercial fishery in the coastwide long model.

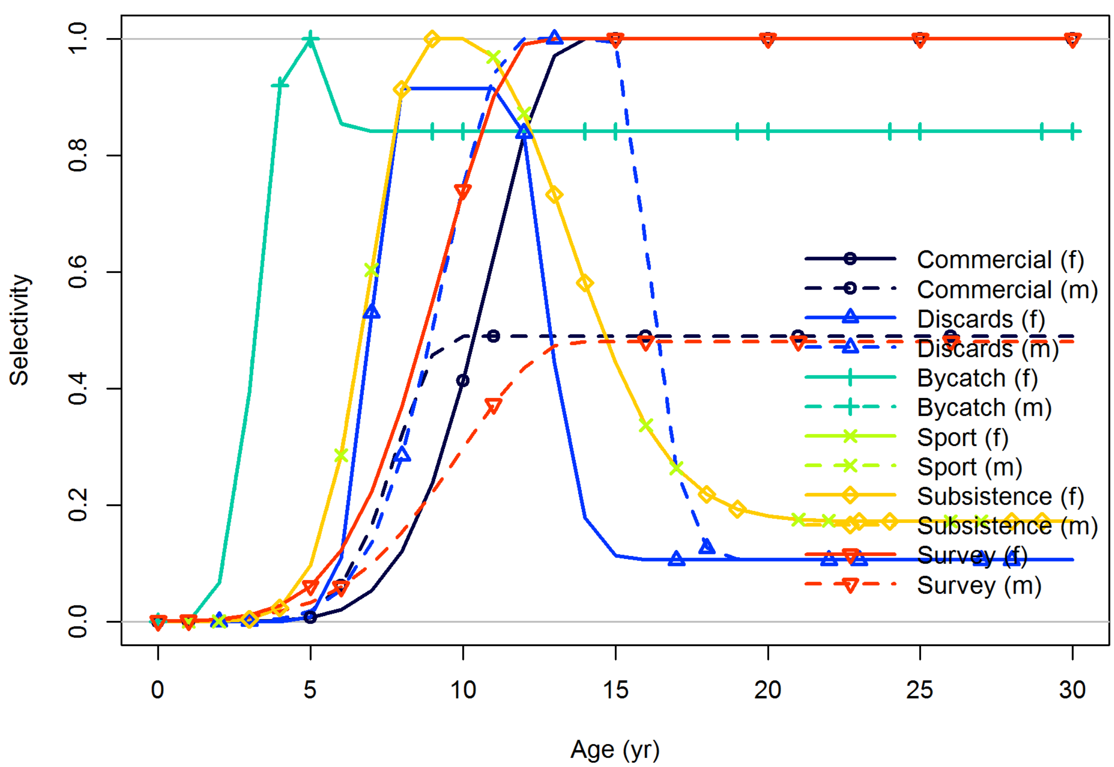


Figure 33. Terminal year selectivity patterns from the coastwide long model.

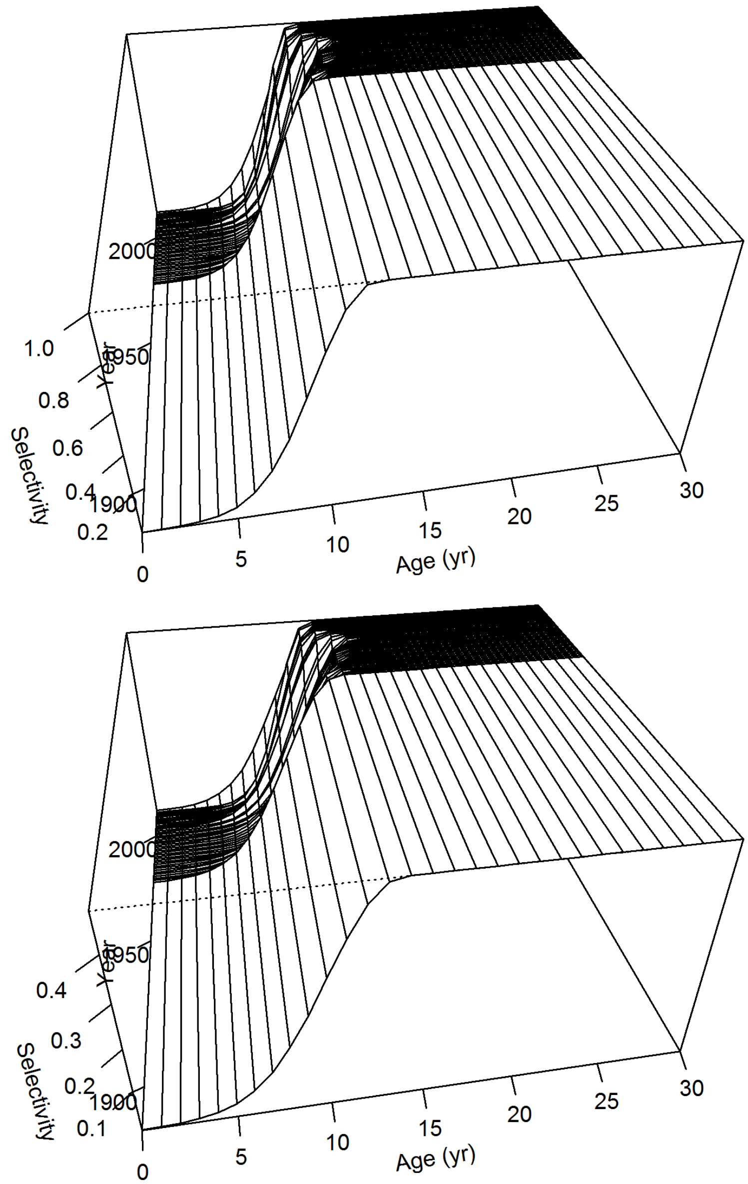


Figure 34. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the survey in the coastwide long model.

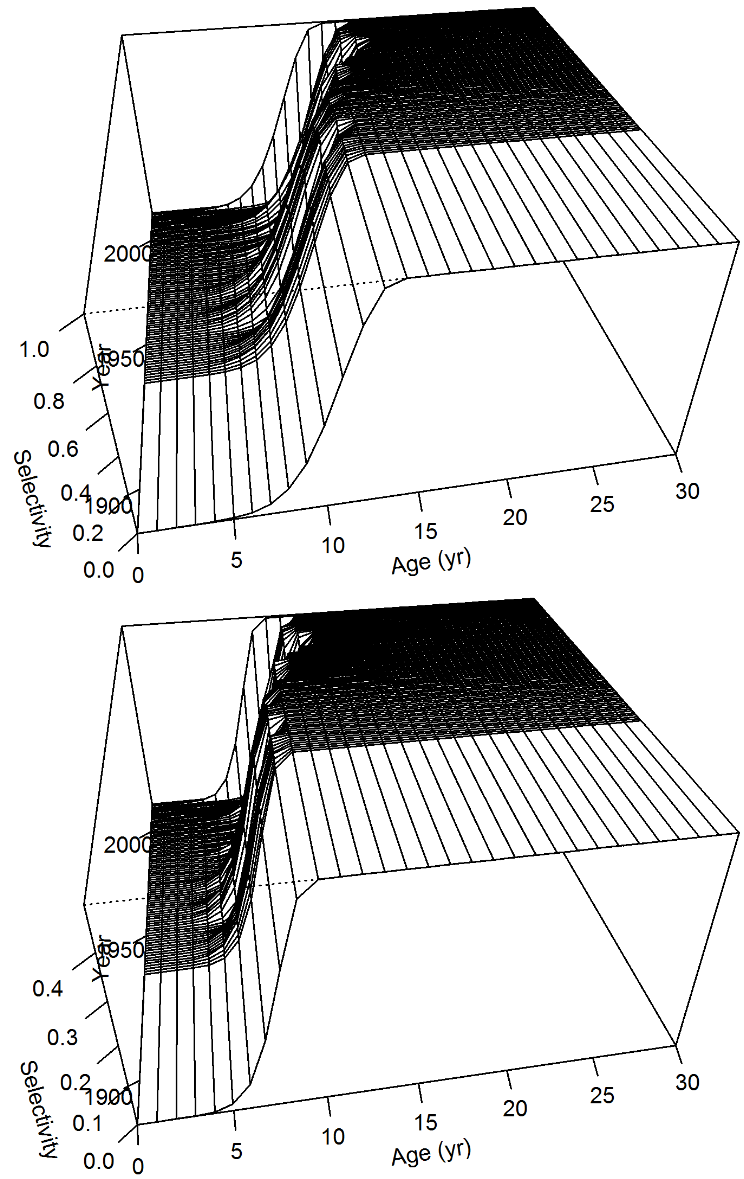


Figure 35. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the directed commercial fishery in the coastwide long model.

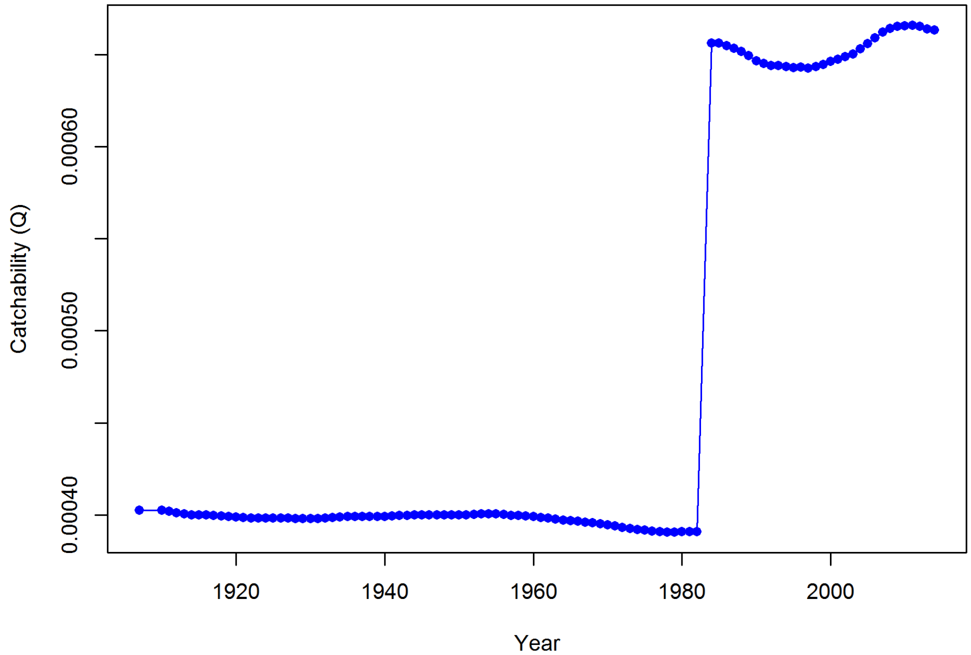


Figure 36. Estimated trend in fishery catchability in the coastwide long model.

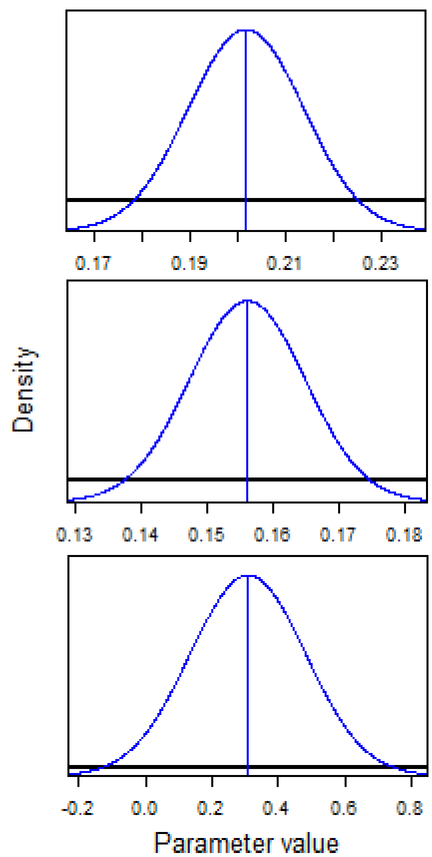


Figure 37. Prior (thick line) and MLE-based distributions for female natural mortality (top panel), male natural mortality (middle panel) and the environmental link coefficient (bottom panel) from the coastwide long model.

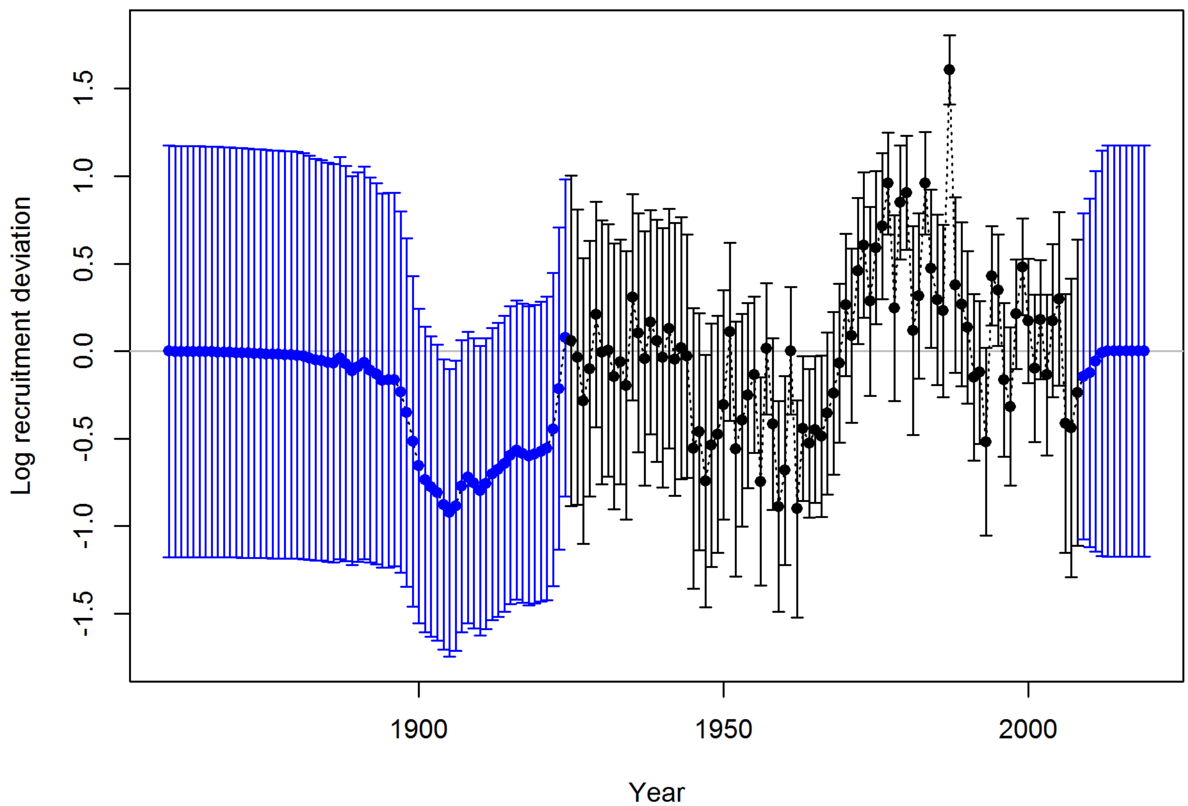


Figure 38. Recruitment deviations from the coastwide long model.

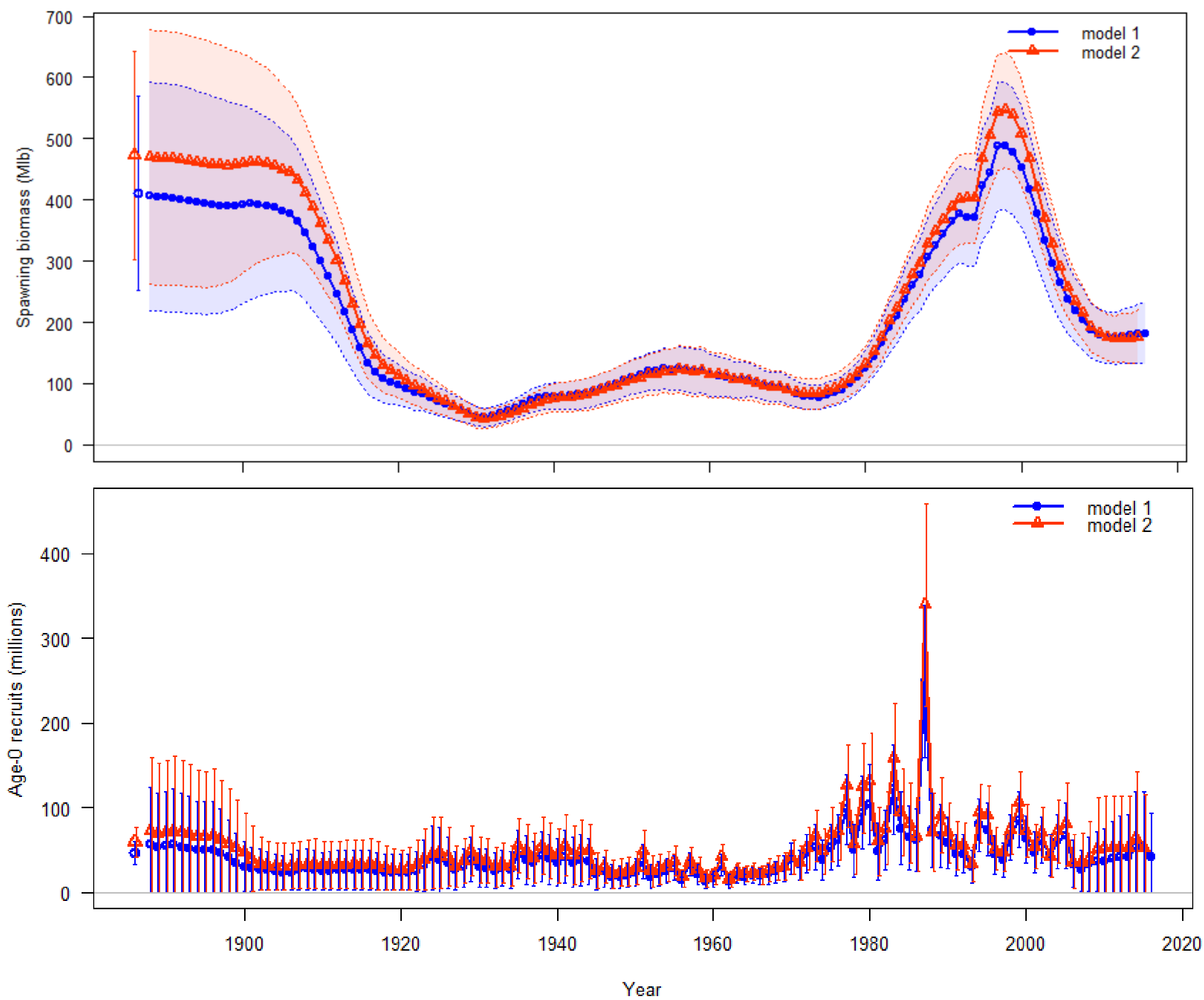


Figure 39. Comparison of spawning biomass (upper panel) and recruitment (lower panel) from the 2014 and preliminary 2015 coastwide long models including all updates to data and model structure.

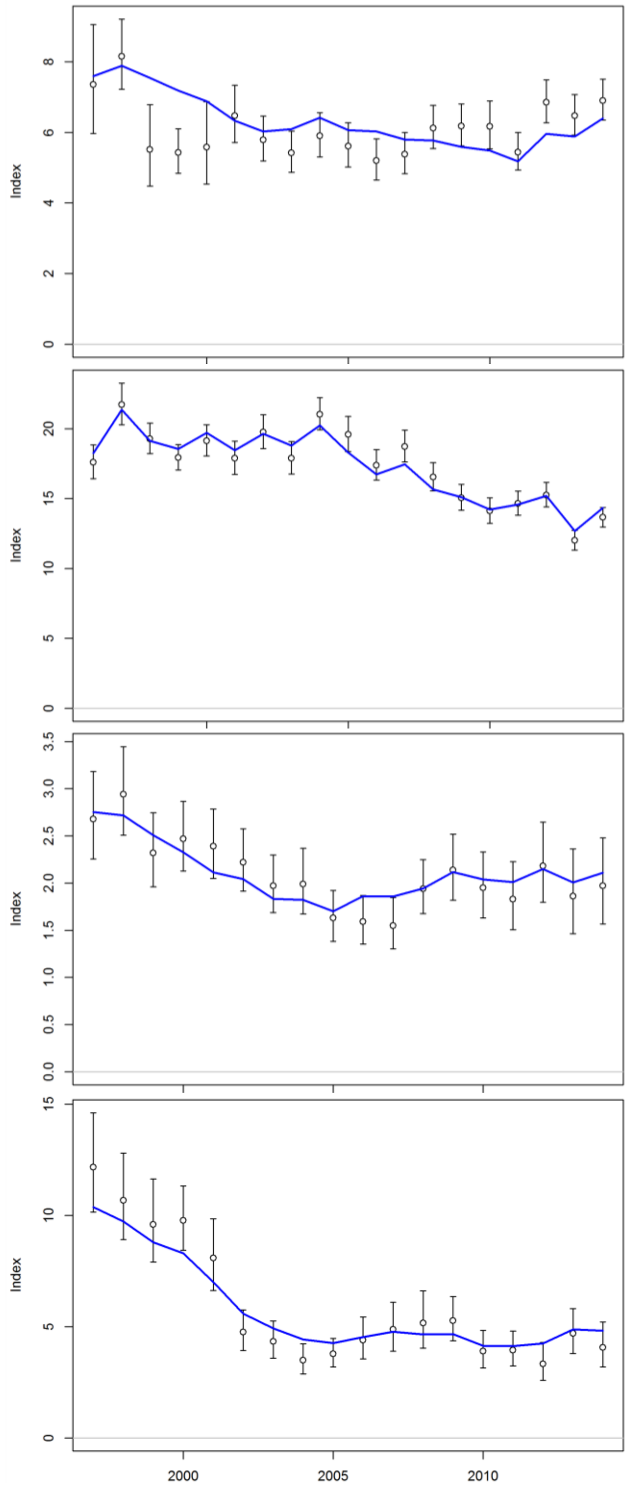


Figure 40. Fit to the index of abundance from the survey (numbers-per-unit-effort) for Area 2 (upper panel), Area 3(second panel), Area 4 (third panel) and Area 4B (bottom panel) in the AAF short model.

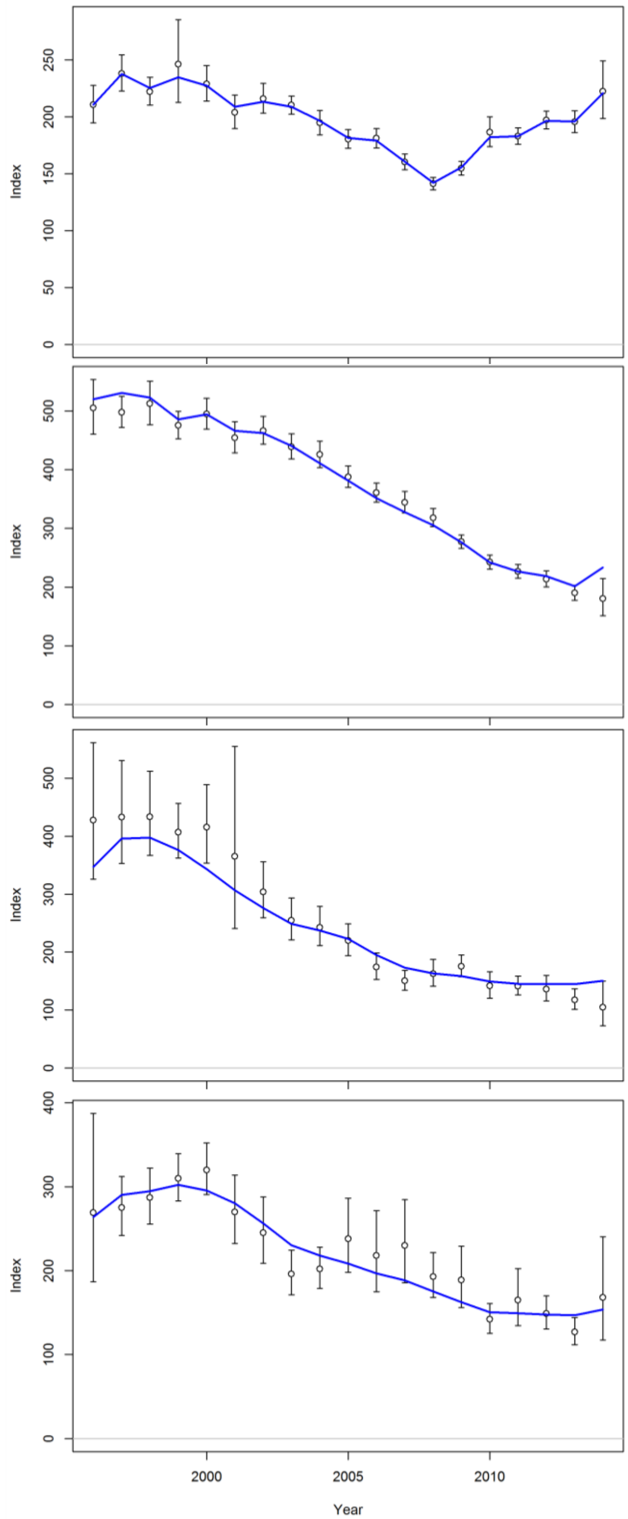


Figure 41. Fit to the index of abundance from the commercial fishery (weight-per-unit-effort) for Area 2 (upper panel), Area 3(second panel), Area 4 (third panel) and Area 4B (bottom panel) in the AAF short model.

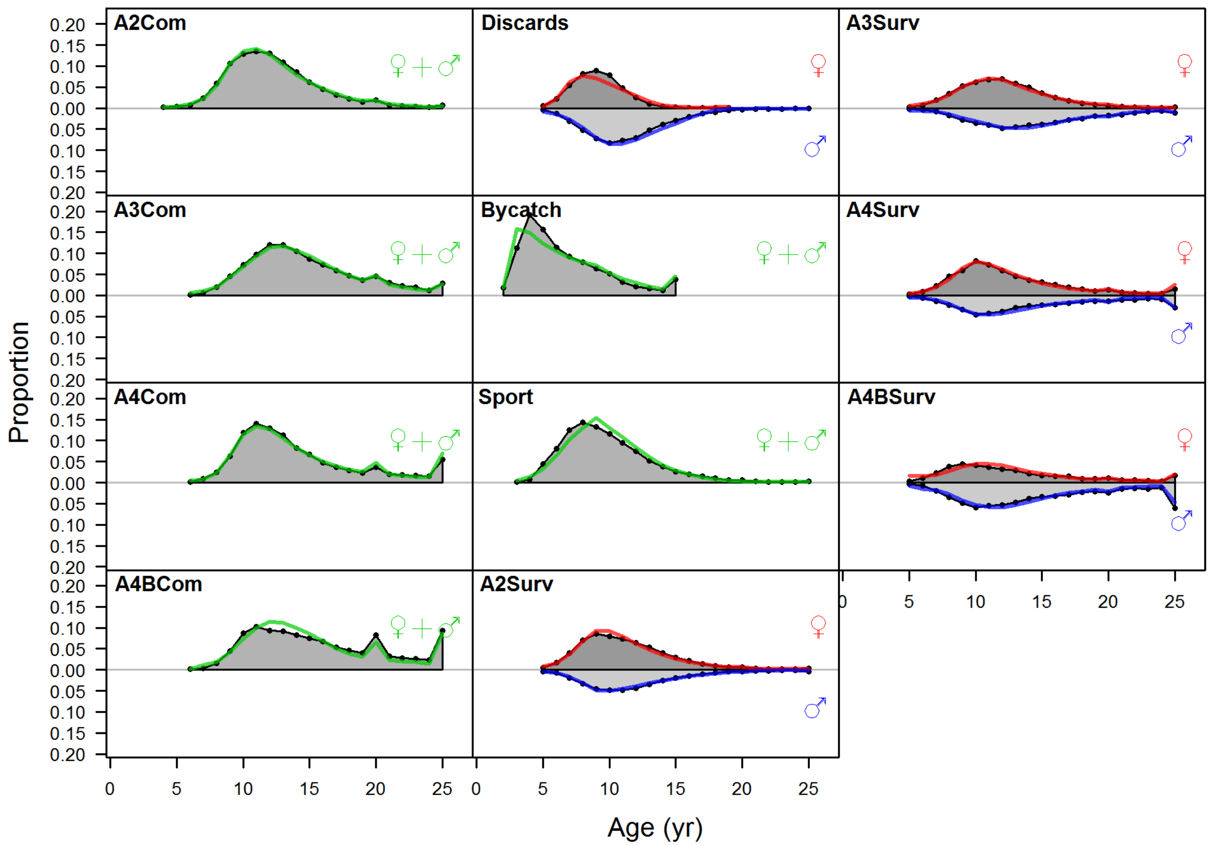


Figure 42. Aggregate fit (observed and expected values summed across all years) to the age data for each fleet in the AAF short model.

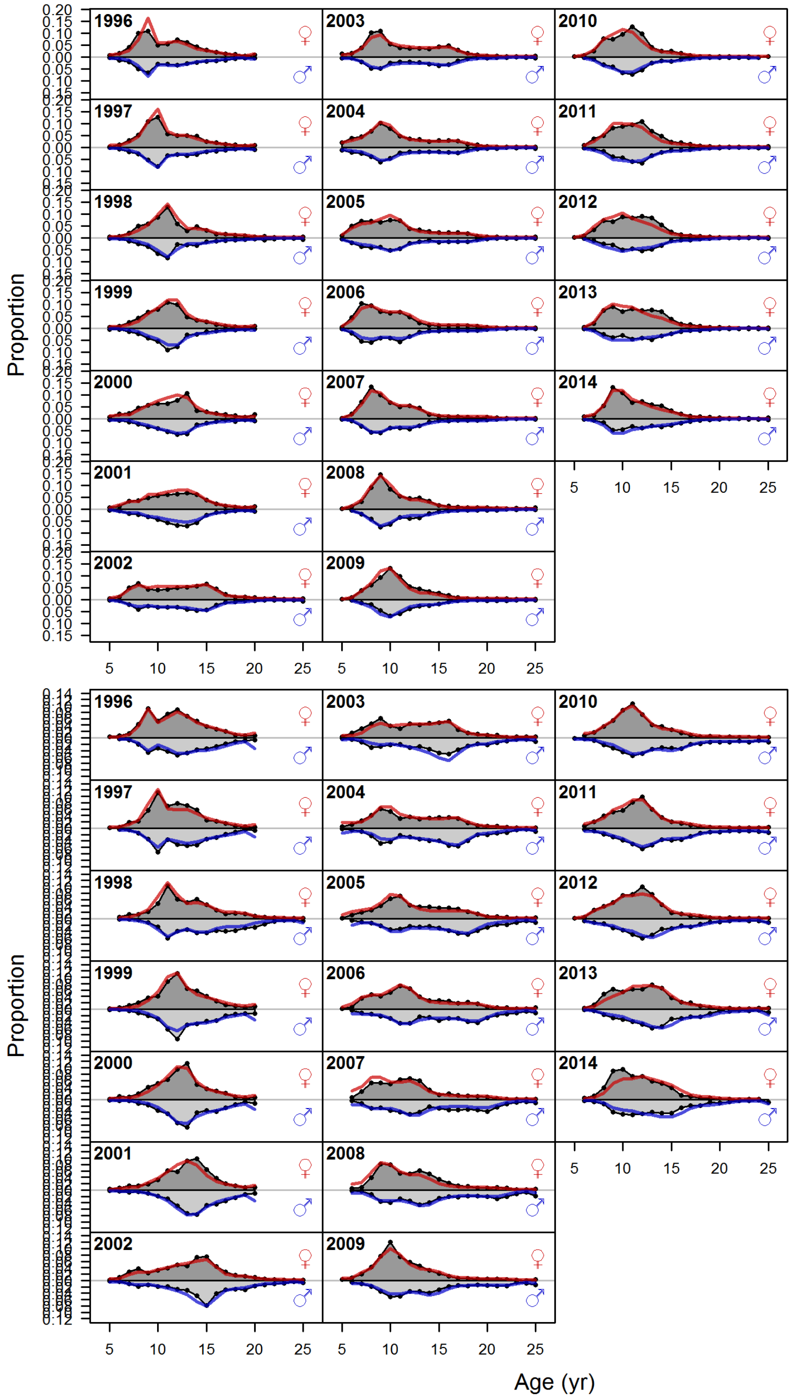


Figure 43. Fit to the age data for the survey in Area 2 (upper panel) and Area 3 (lower panel) in the AAF short model.

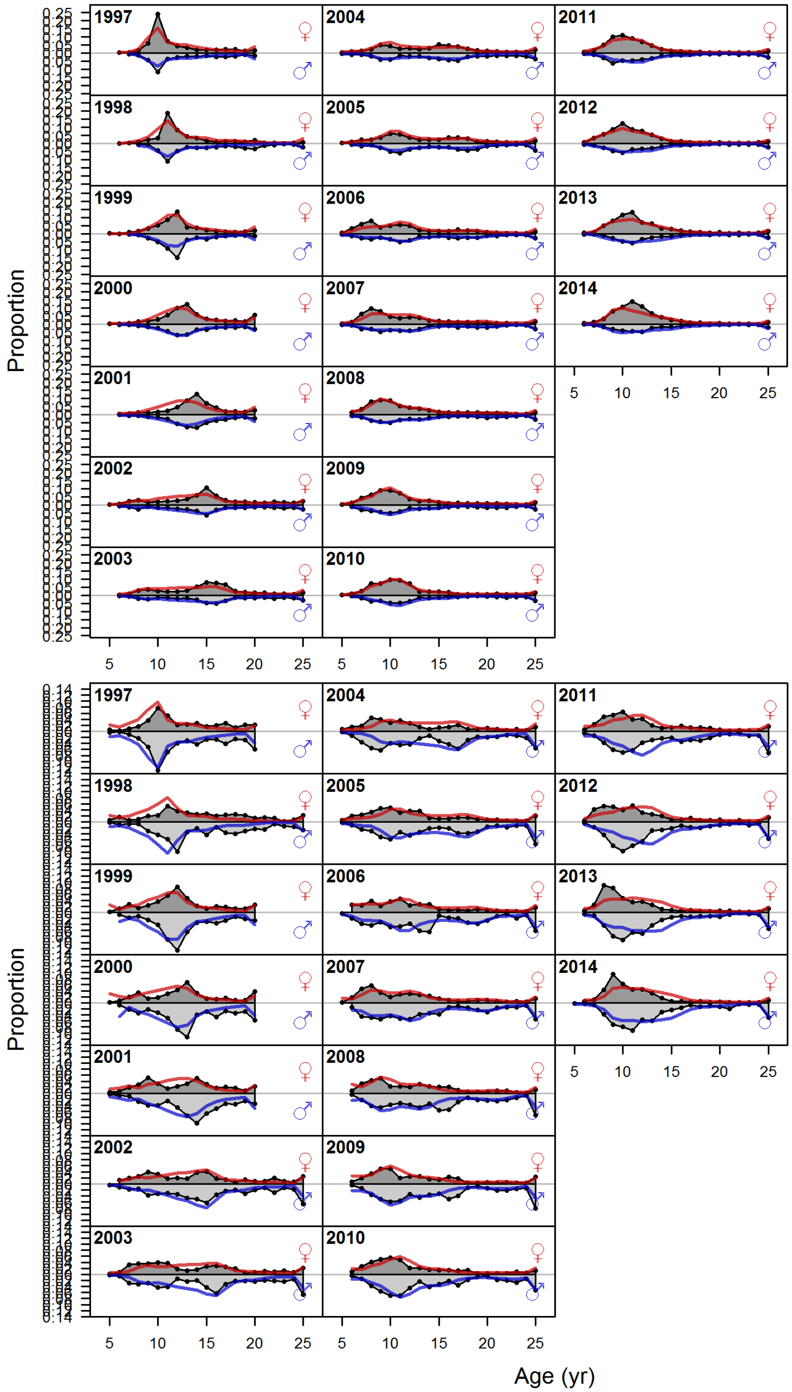


Figure 44. Fit to the age data for the survey in Area 4 (upper panel) and Area 4B (lower panel) in the AAF short model.

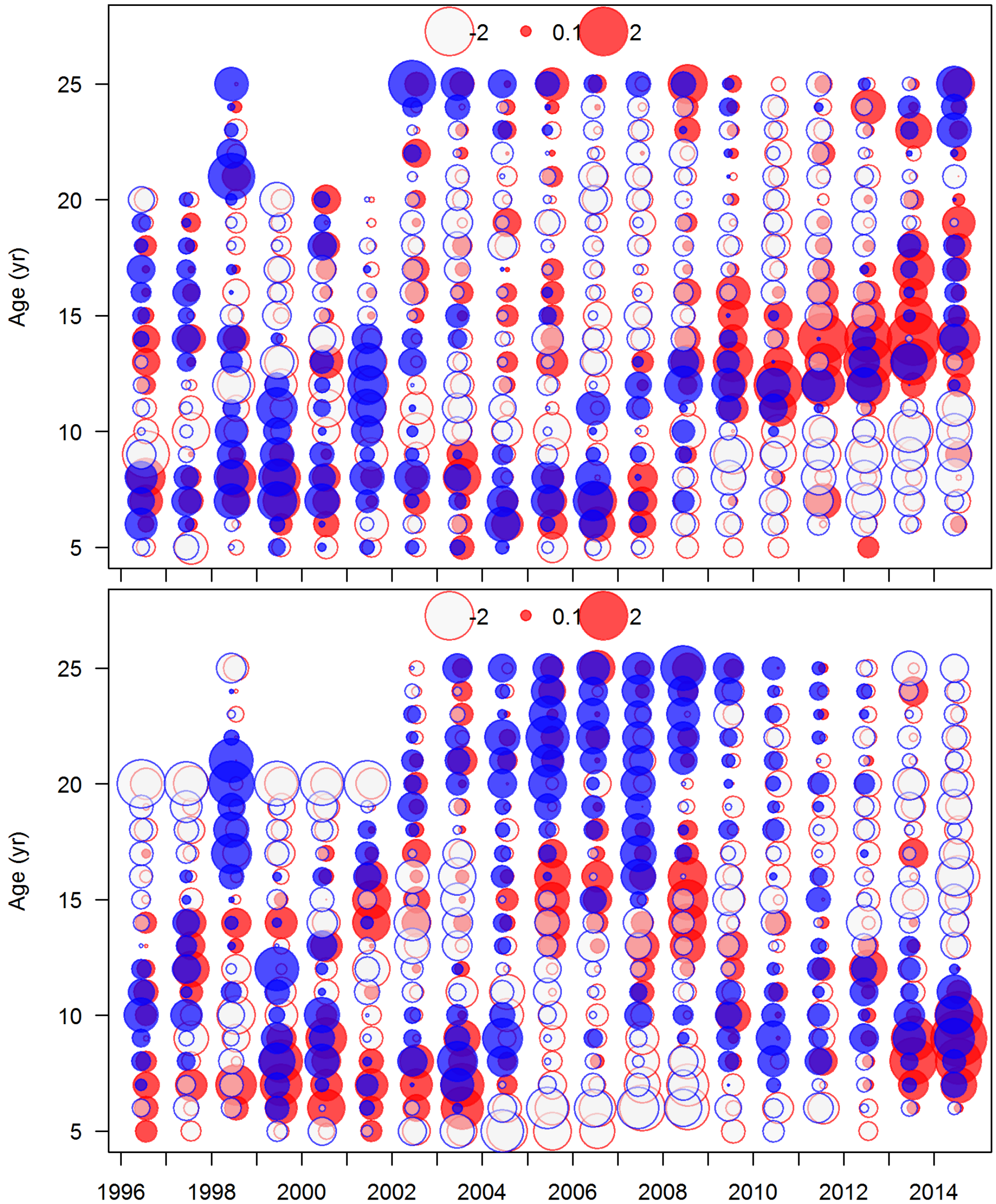


Figure 45. Standardized residuals to the age data for the survey in Area 2 (upper panel) and Area 3 (lower panel) in the AAF short model. Red and blue denote female and male residuals respectively.

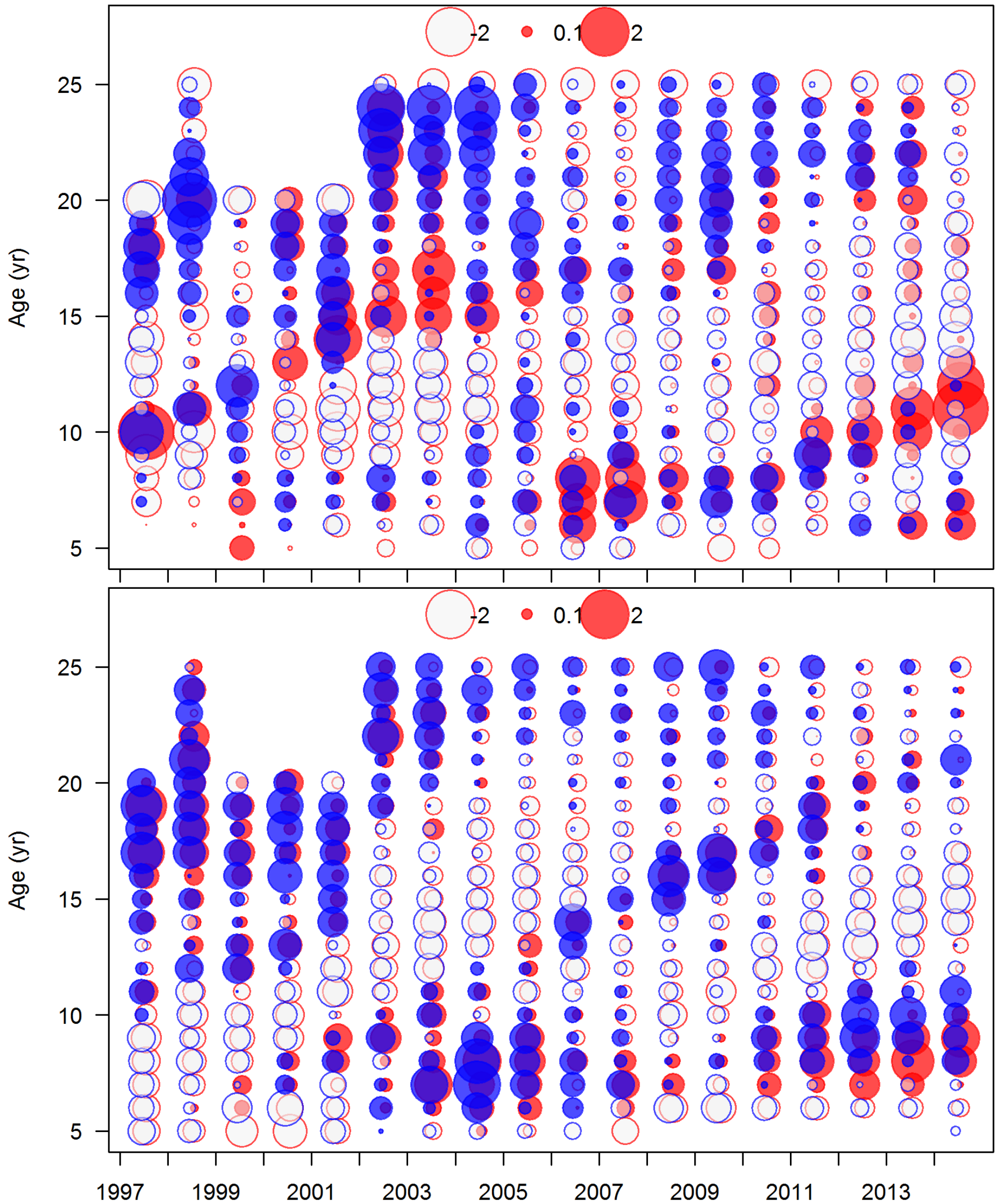


Figure 46. Standardized residuals to the age data for the survey in Area 4 (upper panel) and Area 4B (lower panel) in the AAF short model. Red and blue denote female and male residuals respectively.

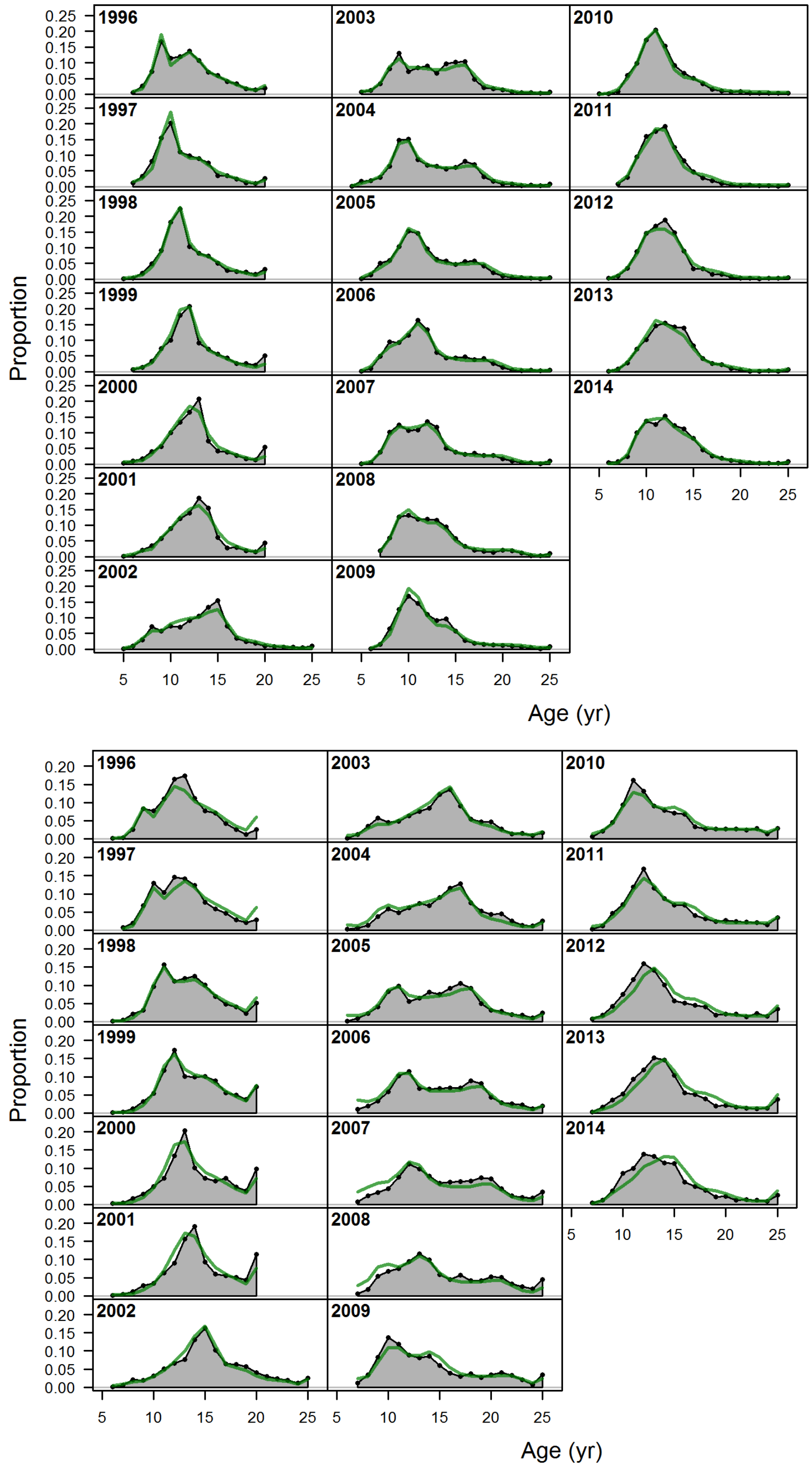


Figure 47. Fit to the age data for the commercial fishery in Area 2 (upper panel) and Area 3 (lower panel) in the AAF short model.

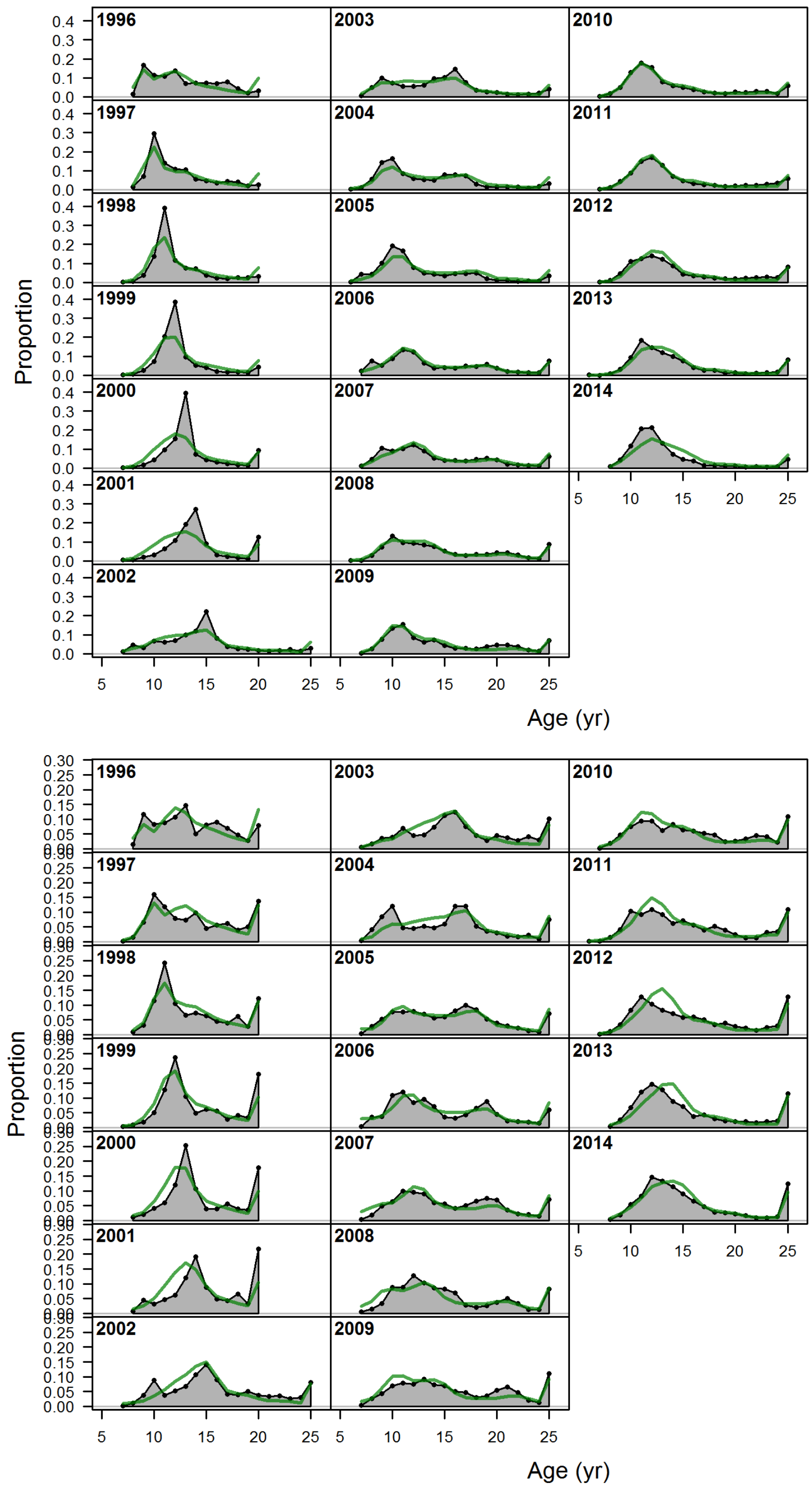


Figure 48. Fit to the age data for the commercial fishery in Area 4 (upper panel) and Area 4B (lower panel) in the AAF short model.

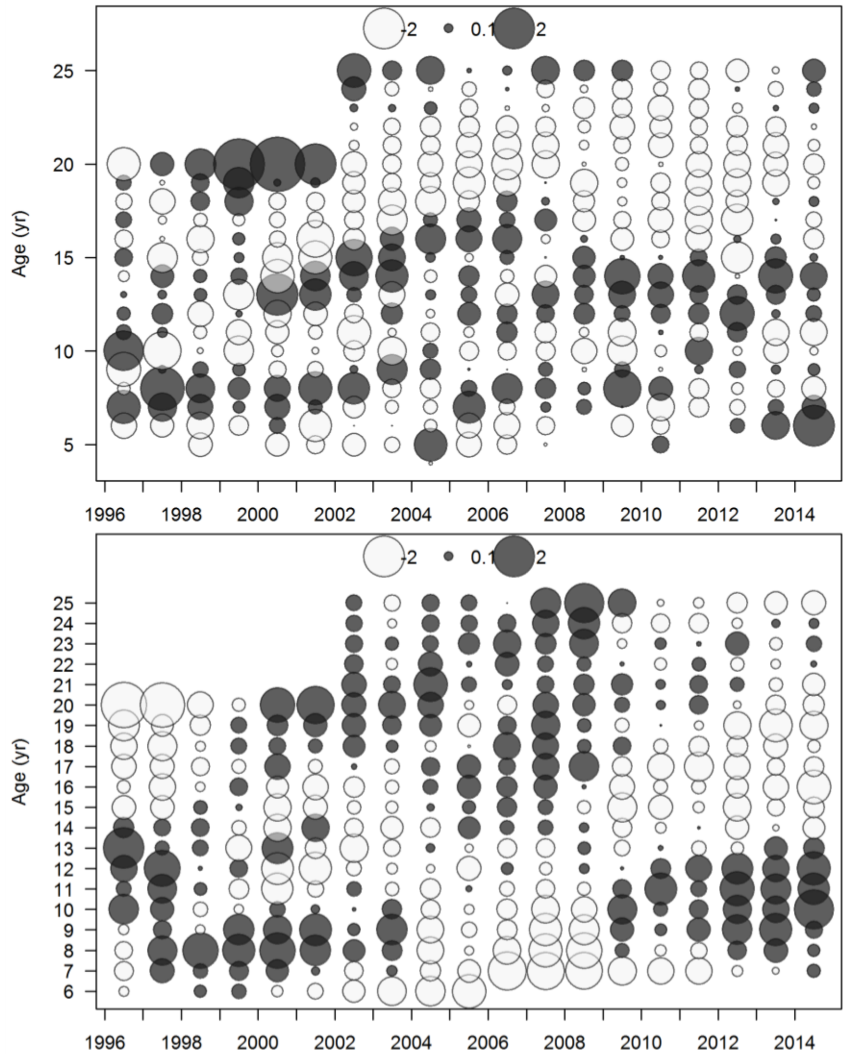


Figure 49. Standardized residuals to the age data for the commercial fishery in Area 2 (upper panel) and Area 3 (lower panel) in the AAF short model.

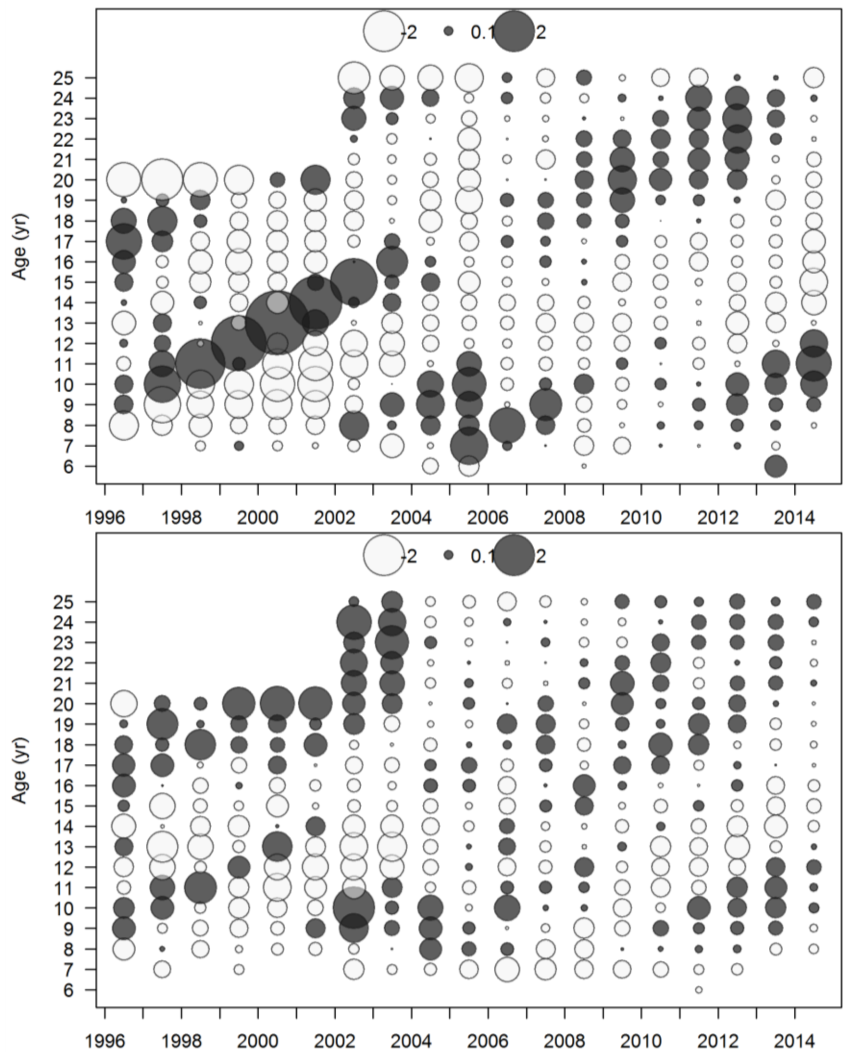
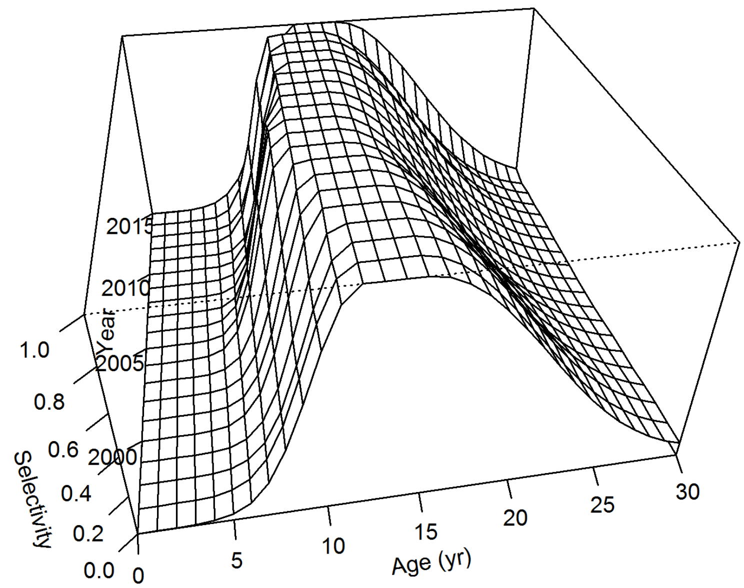


Figure 50. Standardized residuals to the age data for the commercial fishery in Area 4 (upper panel) and Area 4B (lower panel) in the AAF short model.



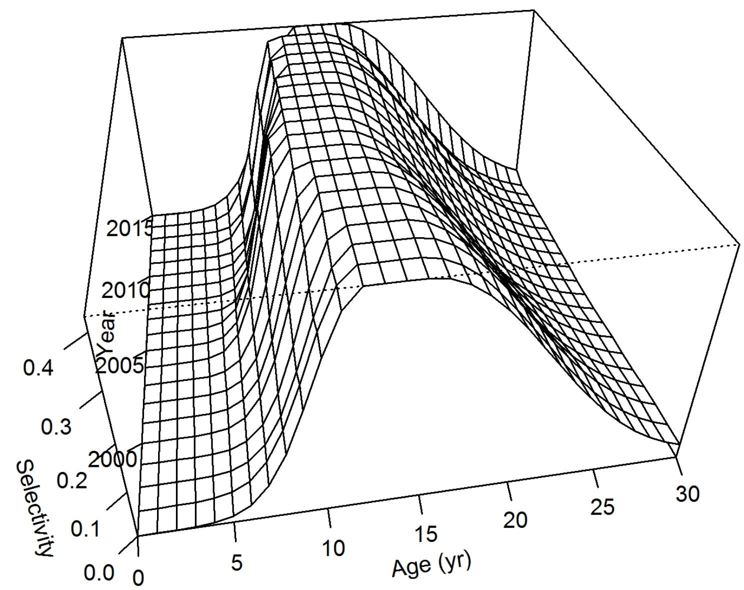


Figure 51. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 2 survey in the AAF short model.

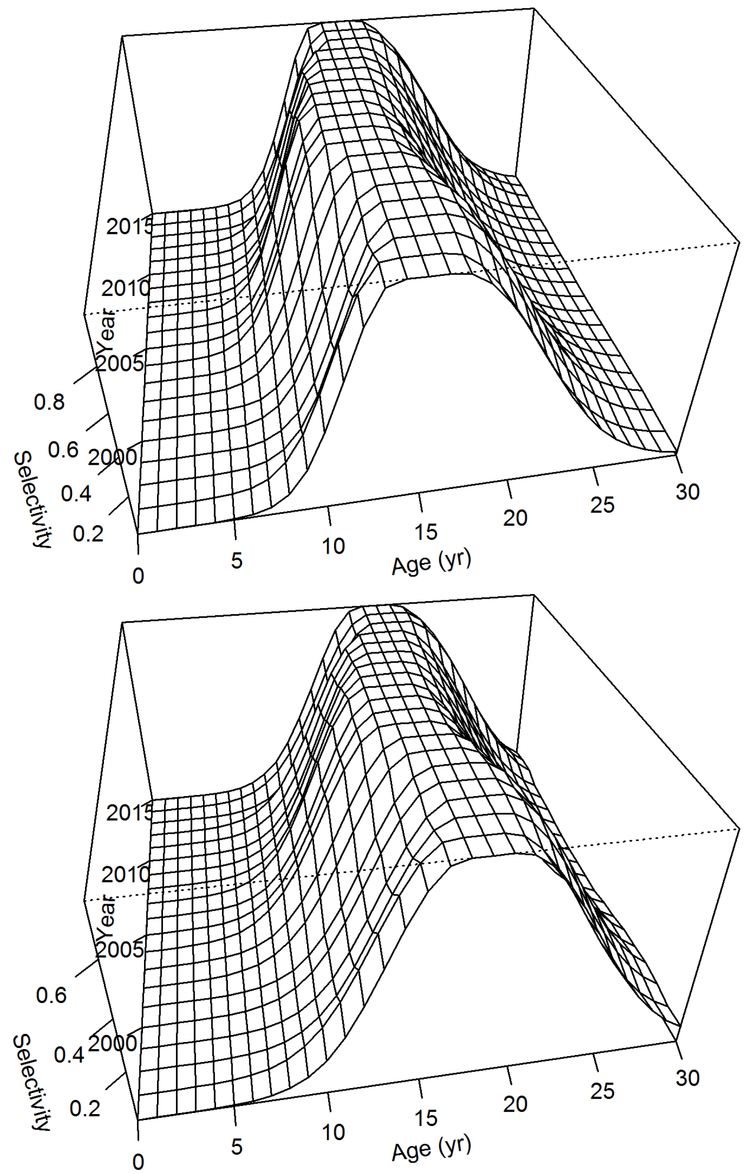


Figure 52. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 3 survey in the AAF short model.

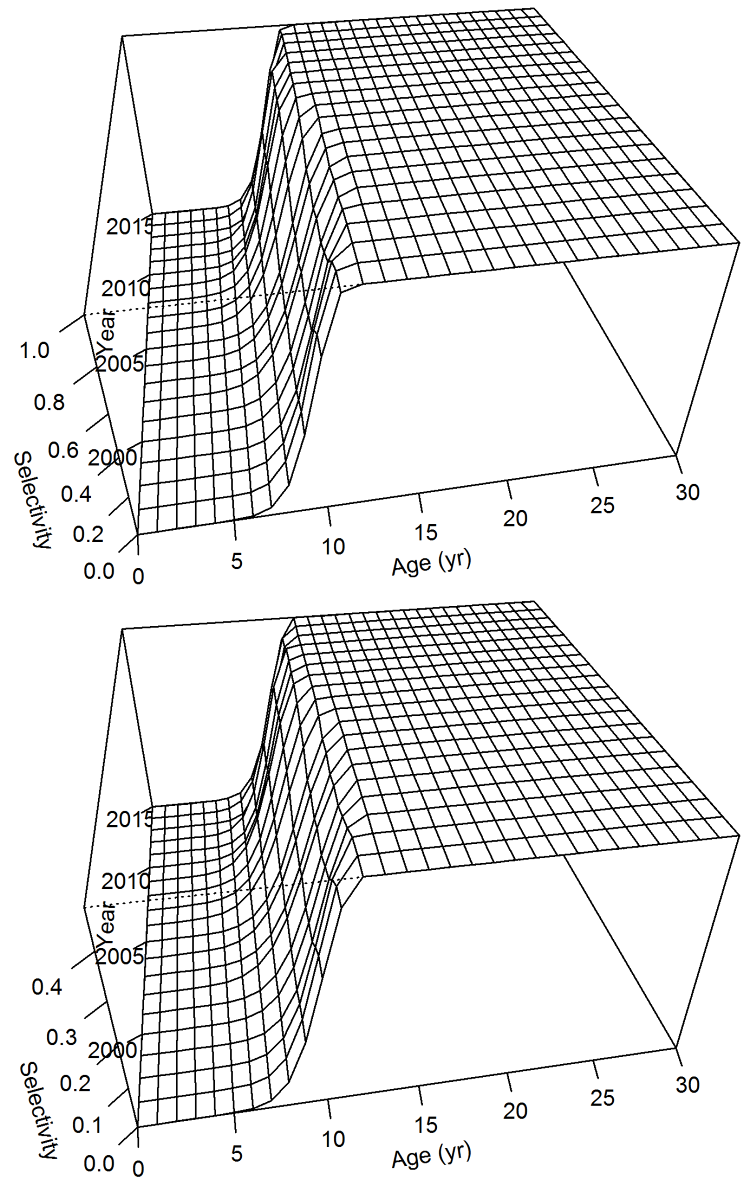


Figure 53. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 survey in the AAF short model.

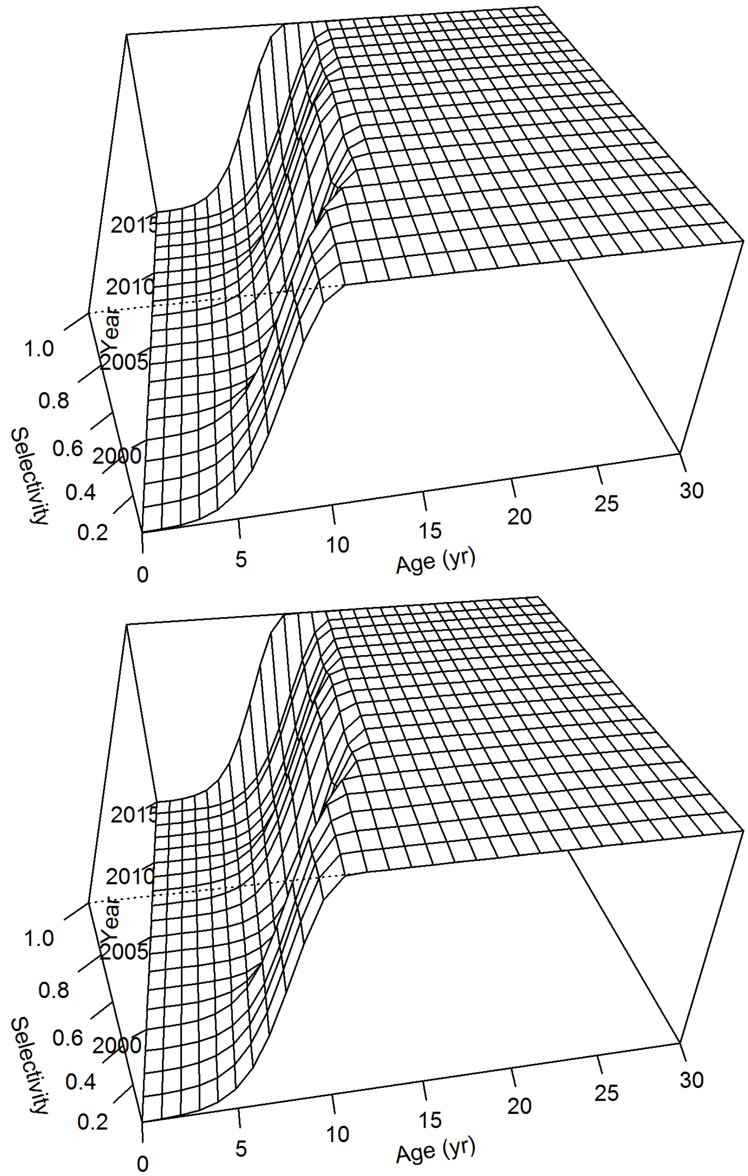


Figure 54. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B survey in the AAF short model.

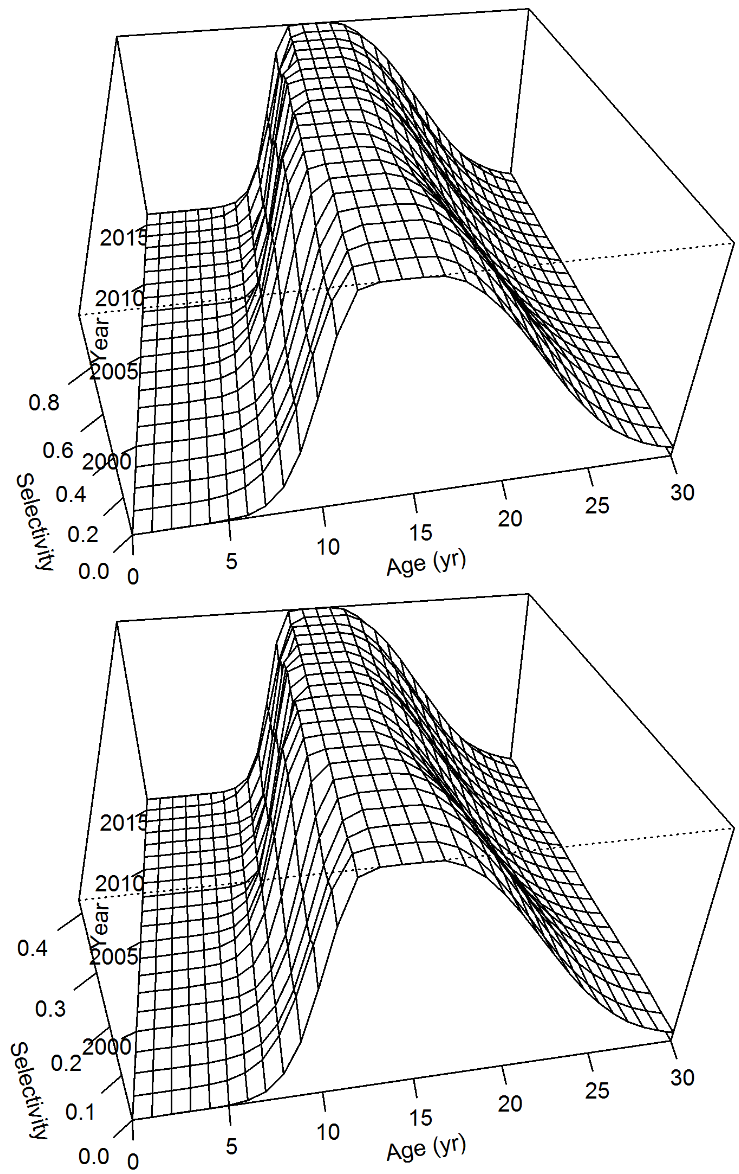


Figure 55. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 2 fishery in the AAF short model.

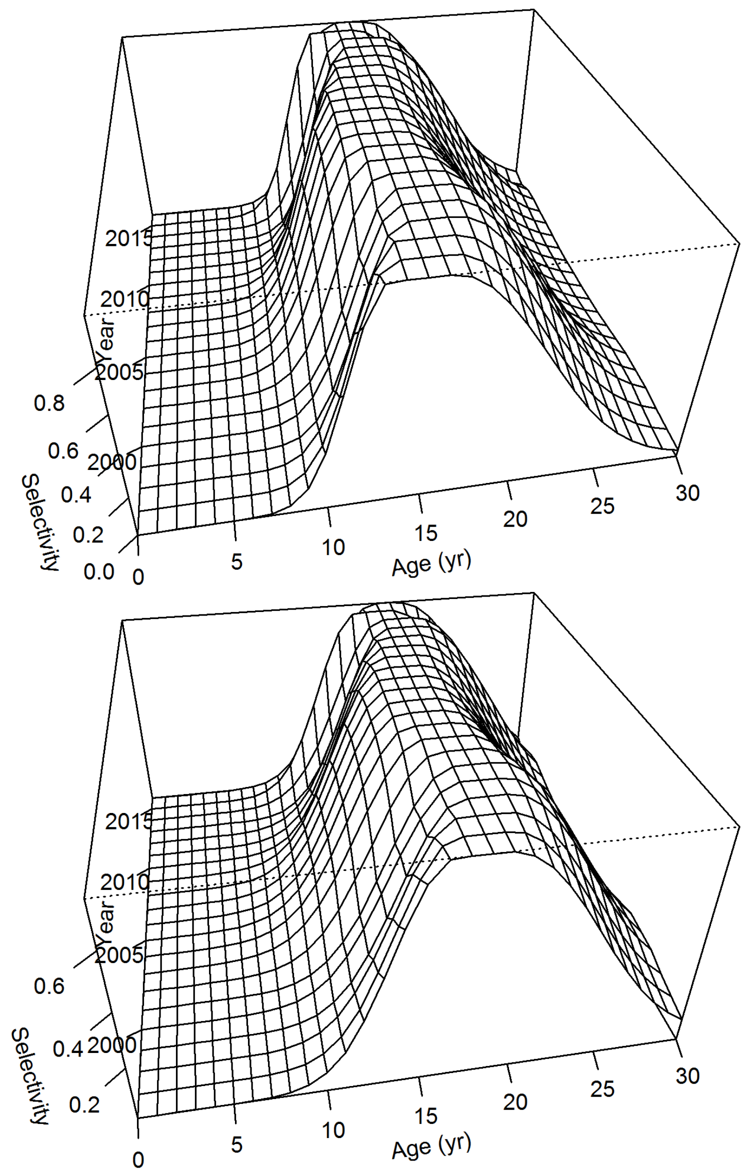


Figure 56. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 3 fishery in the AAF short model.

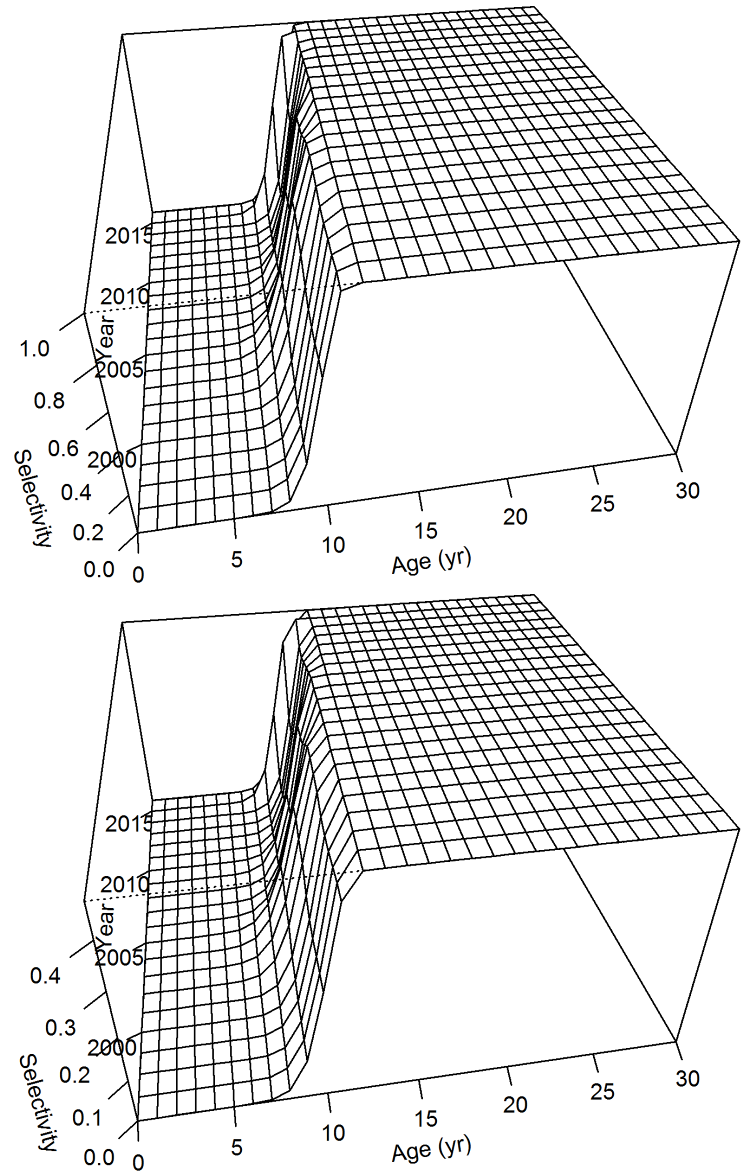
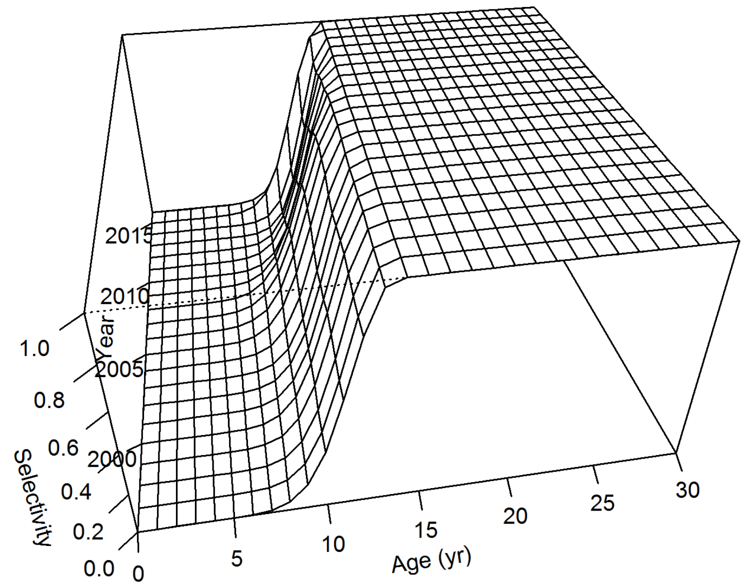


Figure 57. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 fishery in the AAF short model.



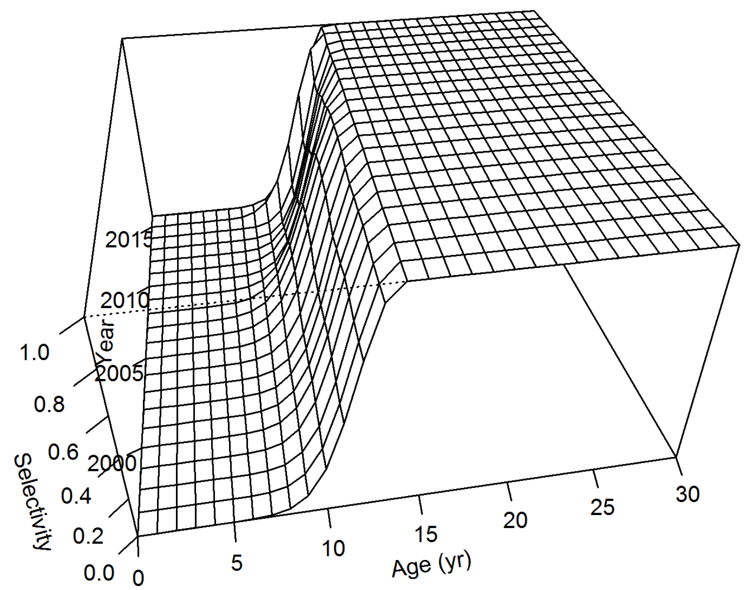


Figure 58. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B fishery in the AAF short model.

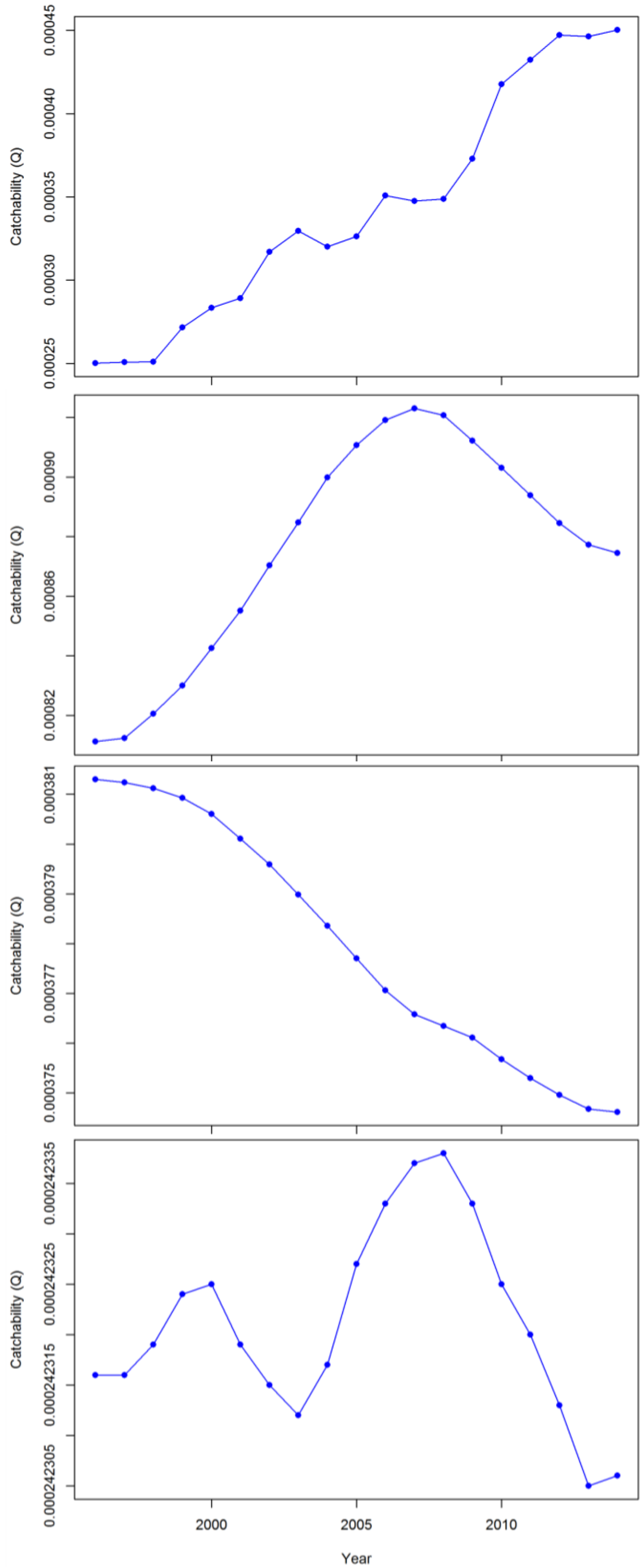


Figure 59. Estimated trends in fishery catchability for Area 2 (upper panel), Area 3 (second panel), Area 4 (third panel) and Area 4B (lower panel) in the AAF short model.

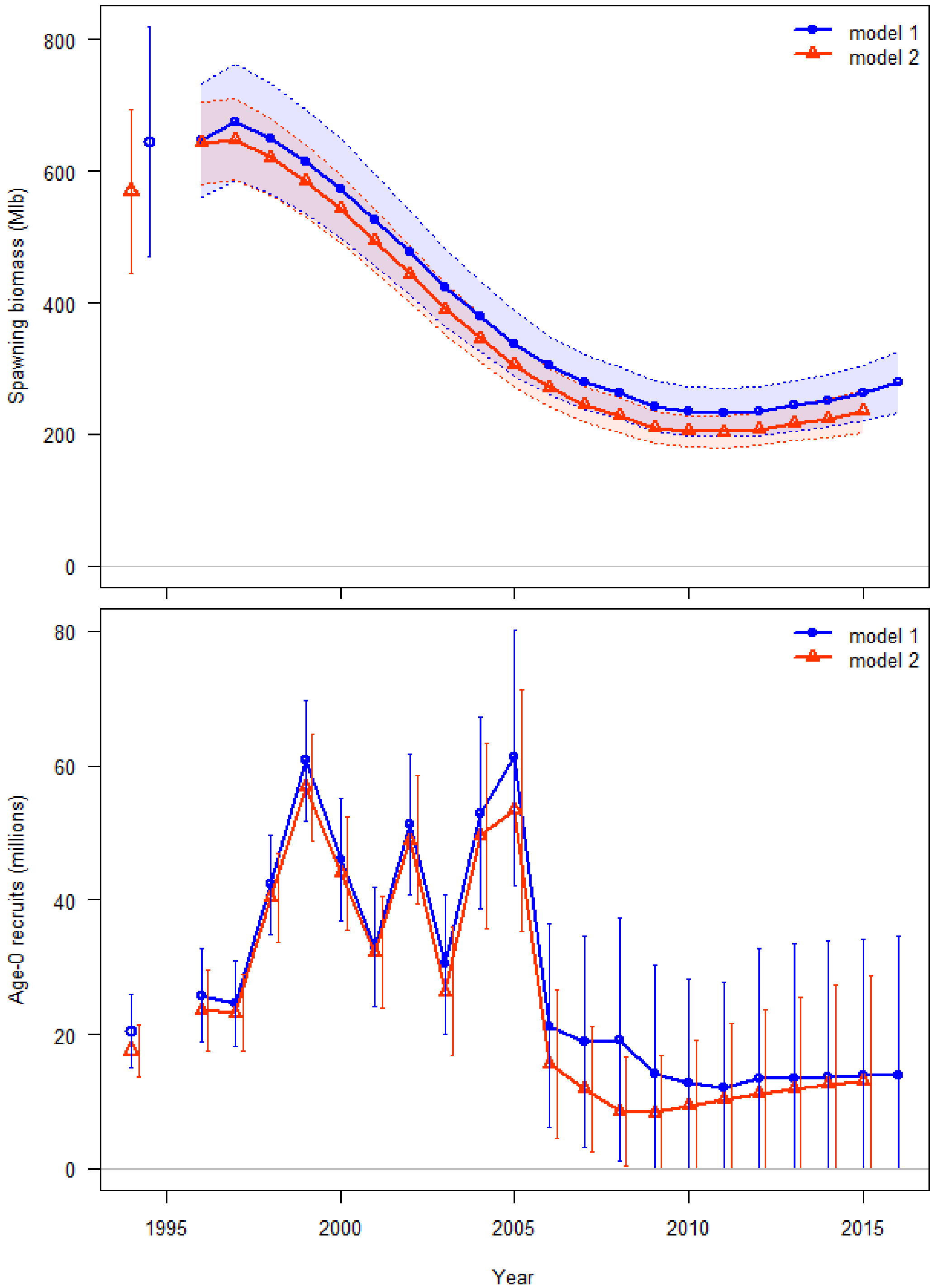


Figure 60. Comparison of the spawning biomass (upper panel) and recruitment (lower panel) from the 2014, and preliminary 2015 AAF short models including all updates to data and model structure.

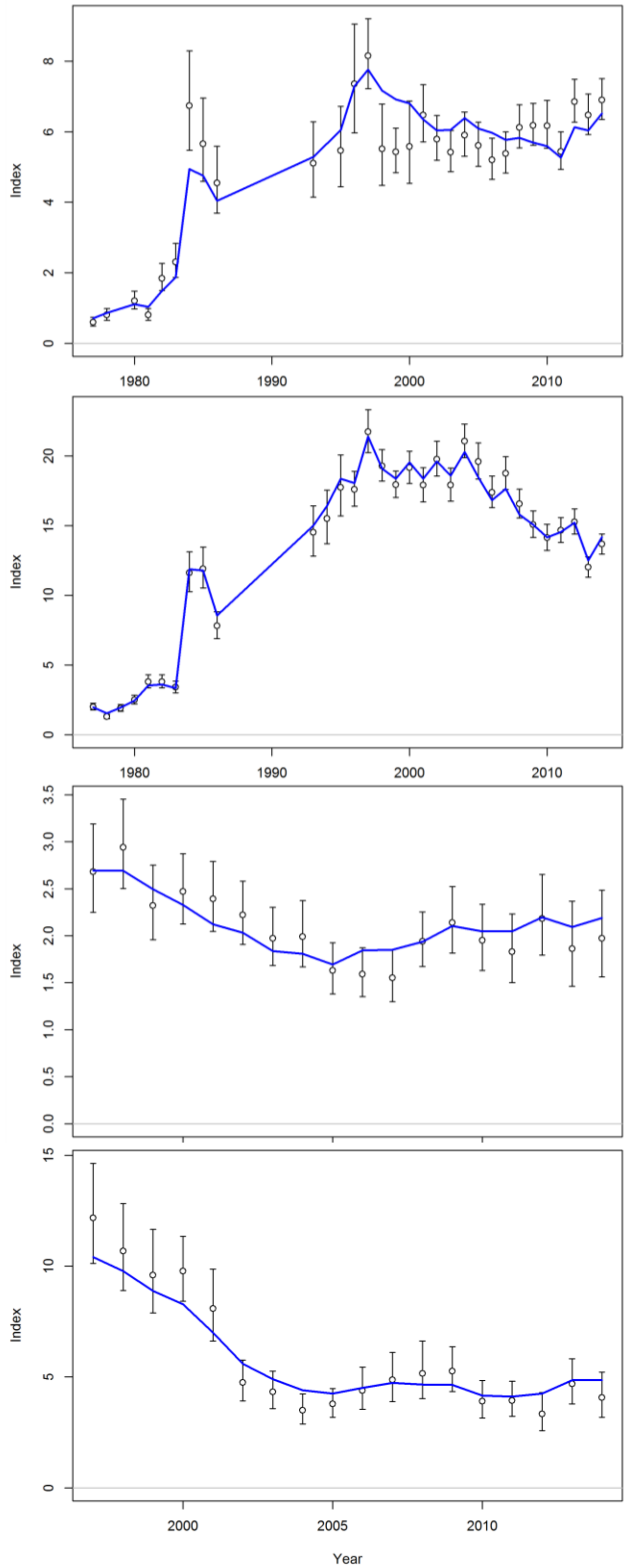


Figure 61. Fit to the index of abundance from the survey (numbers-per-unit-effort) for Area 2 (upper panel), Area 3(second panel), Area 4 (third panel) and Area 4B (bottom panel) in the AAF long model.

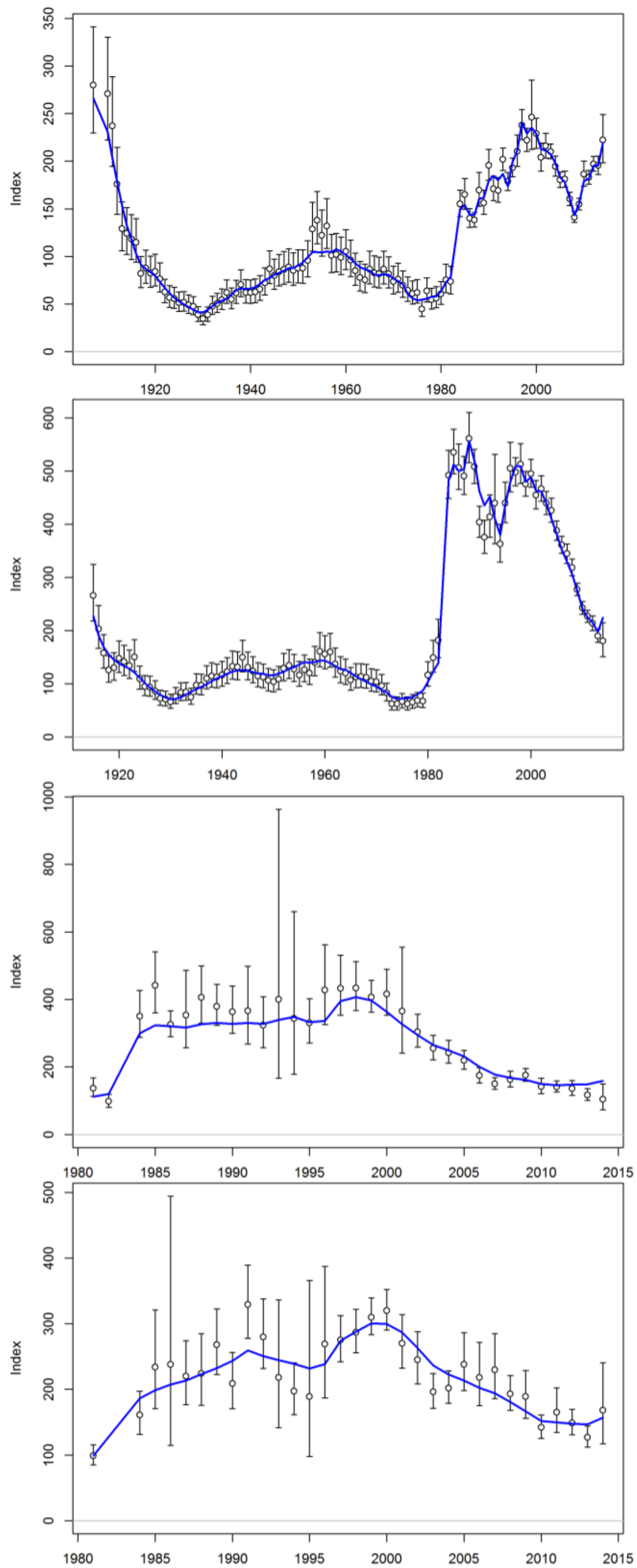


Figure 62. Fit to the index of abundance from the commercial fishery (weight-per-unit-effort) for Area 2 (upper panel), Area 3(second panel), Area 4 (third panel) and Area 4B (bottom panel) in the AAF long model.

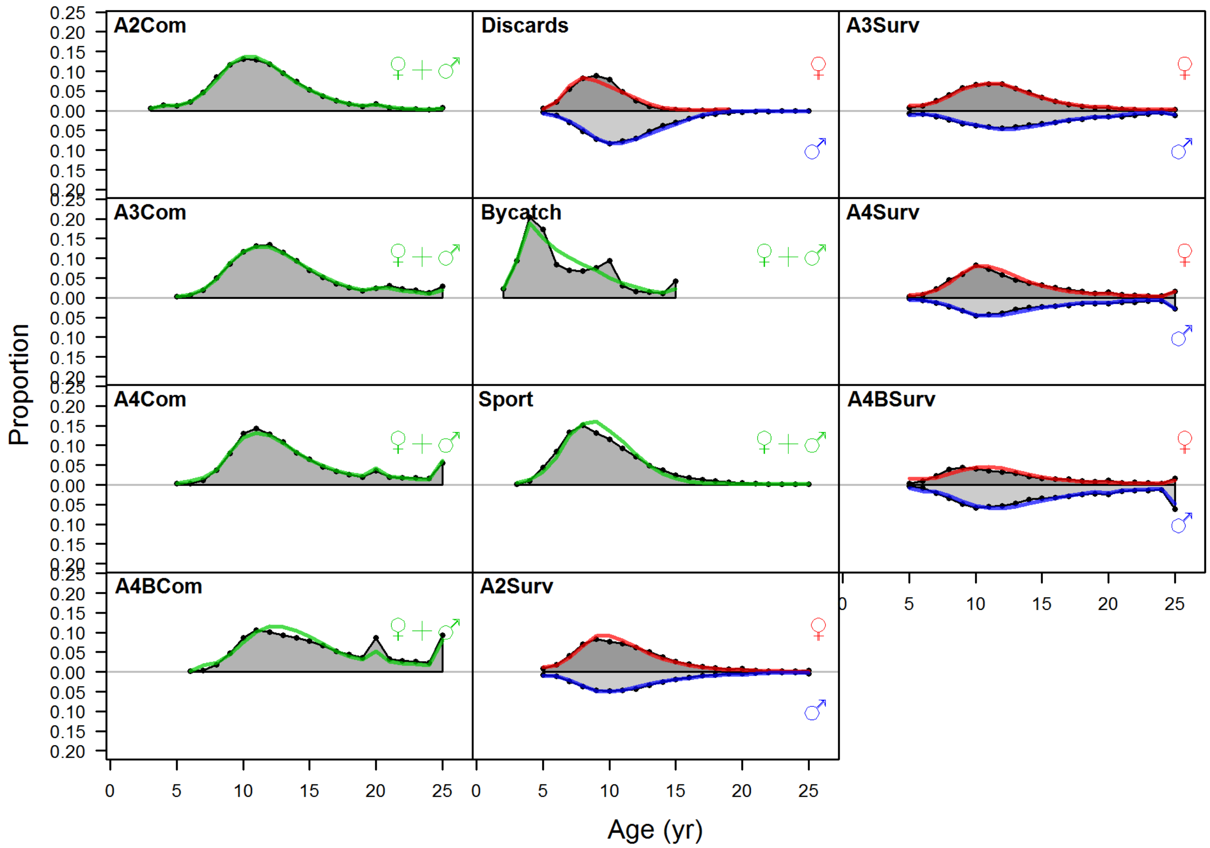


Figure 63. Aggregate fit (observed and expected values summed across all years) to the age data for each fleet in the AAF long model.

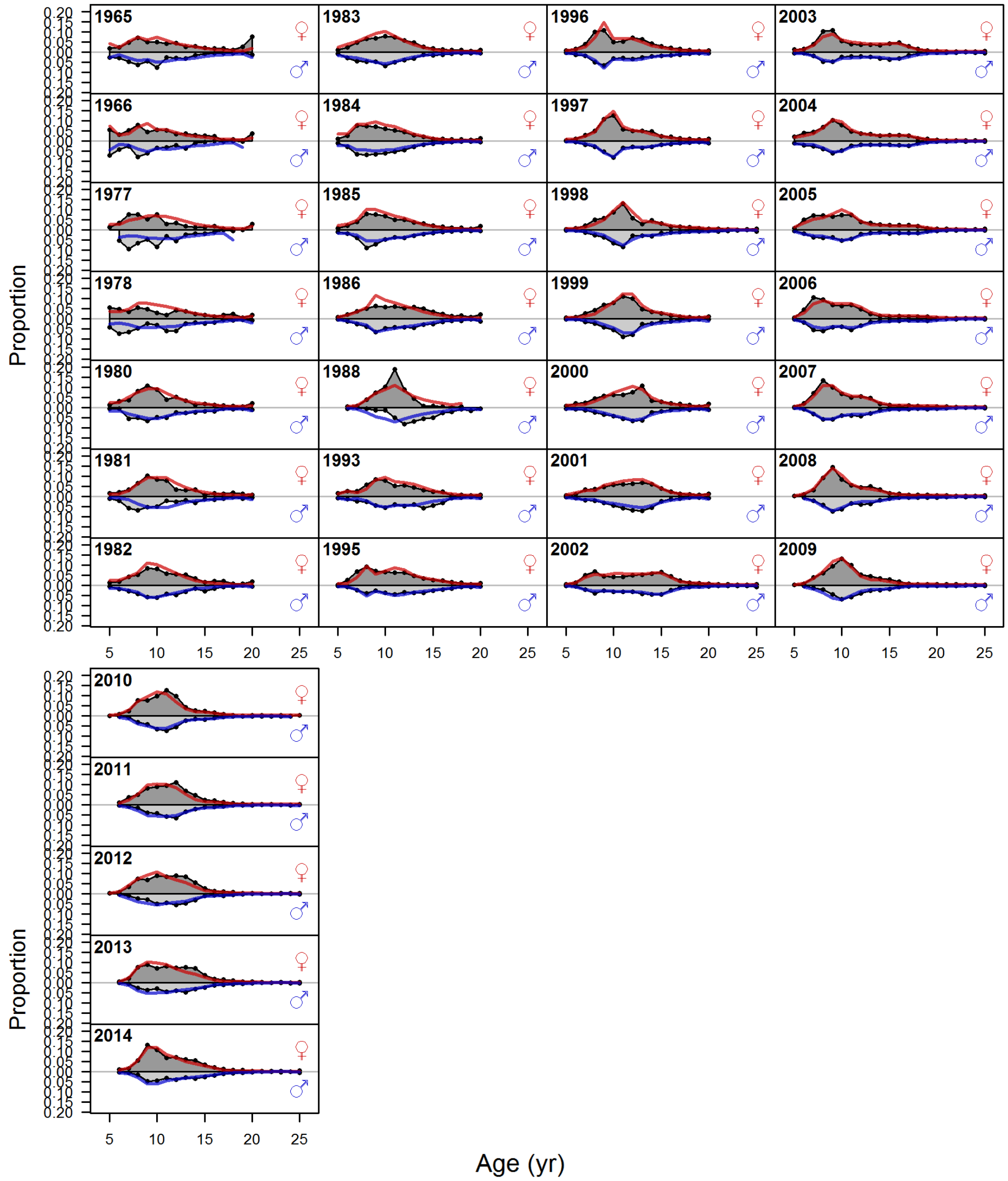


Figure 64. Fit to the age data for the survey in Area 2 in the AAF long model.

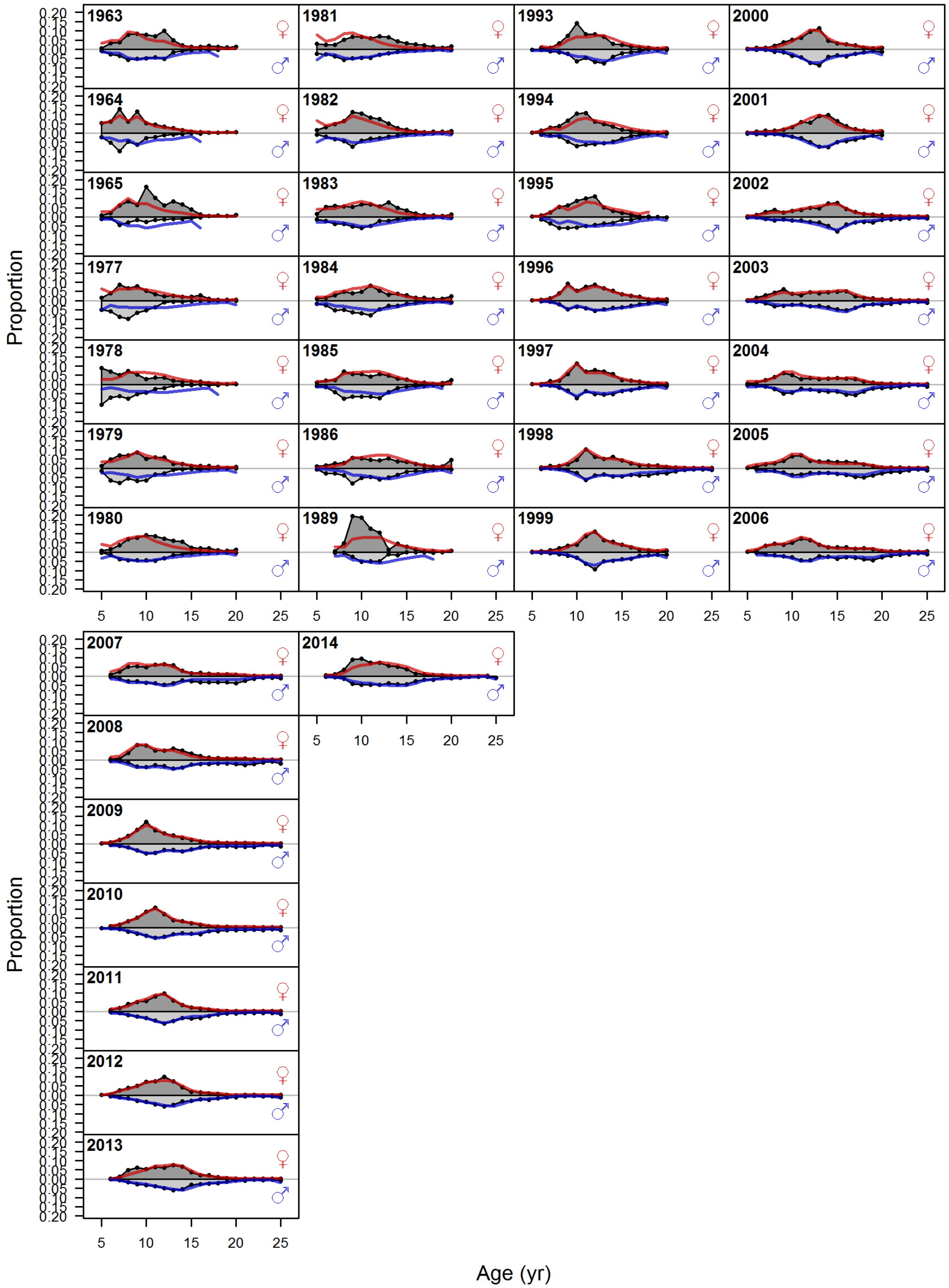


Figure 65. Fit to the age data for the survey in Area 3 in the AAF long model.

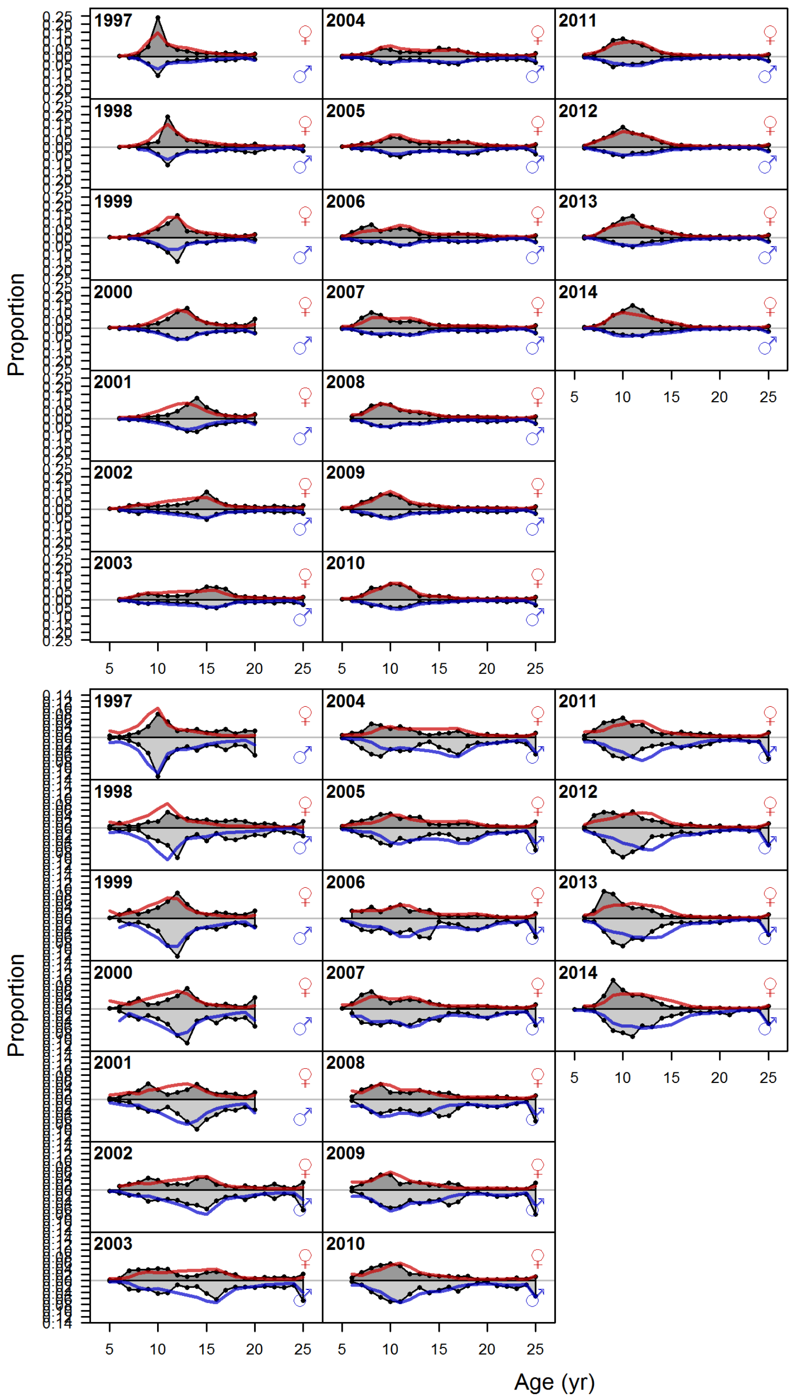


Figure 66. Fit to the age data for the survey in Area 4 (upper panel) and Area 4B (lower panel) in the AAF long model.

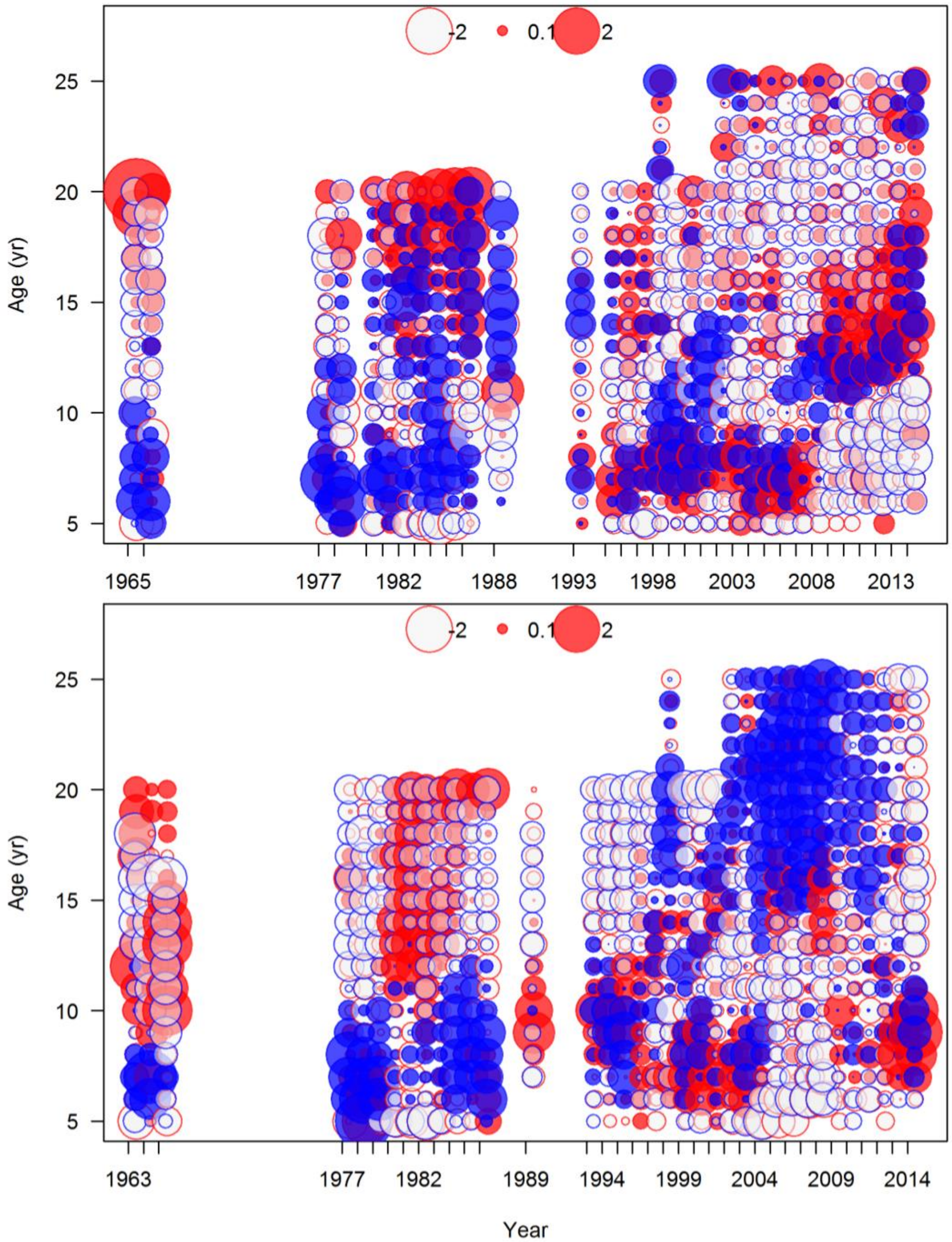


Figure 67. Standardized residuals to the age data for the survey in Area 2 (upper panel) and Area 3 (lower panel) in the AAF long model. Red and blue denote female and male residuals respectively.

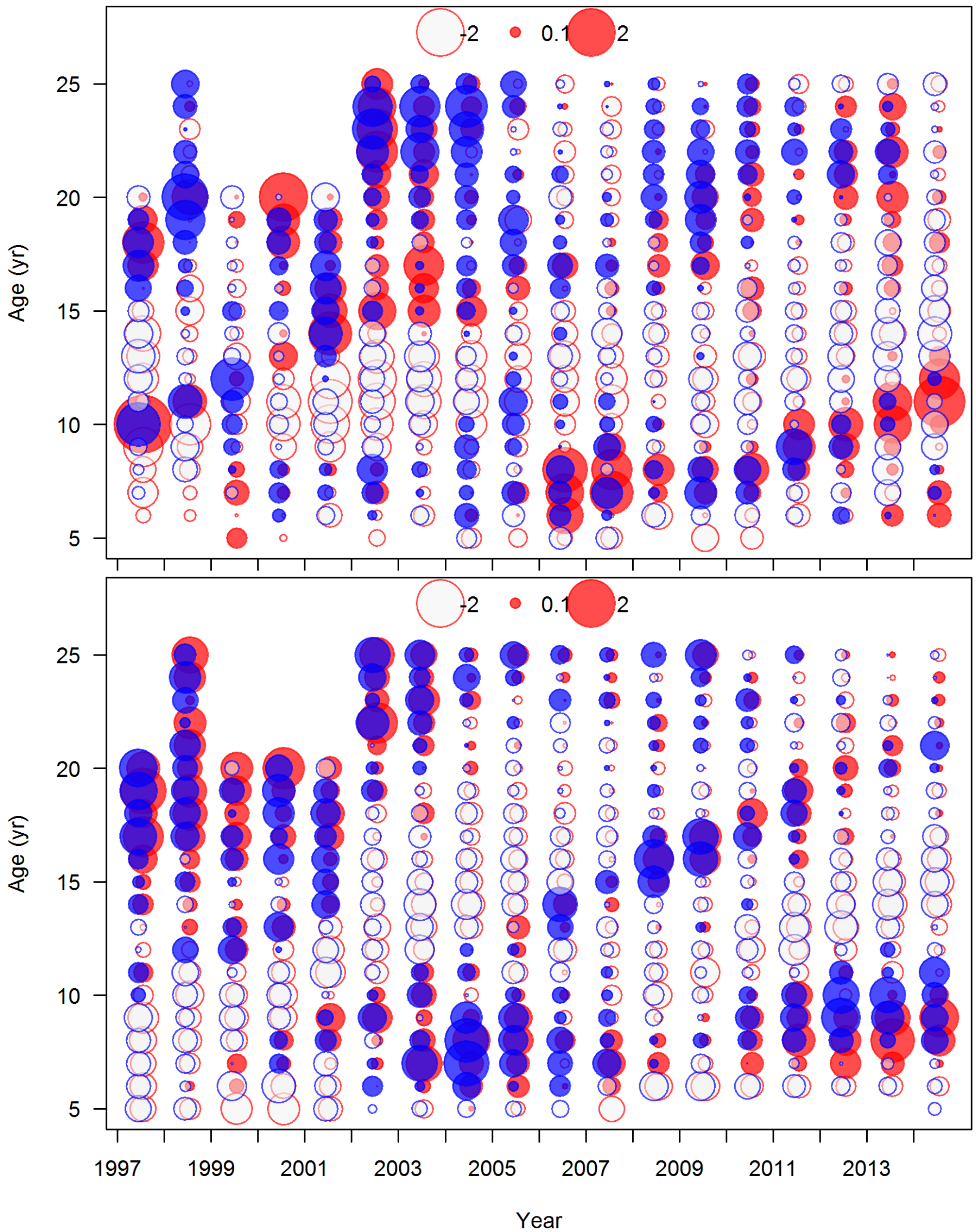


Figure 68. Standardized residuals to the age data for the survey in Area 4 (upper panel) and Area 4B (lower panel) in the AAF long model. Red and blue denote female and male residuals respectively.

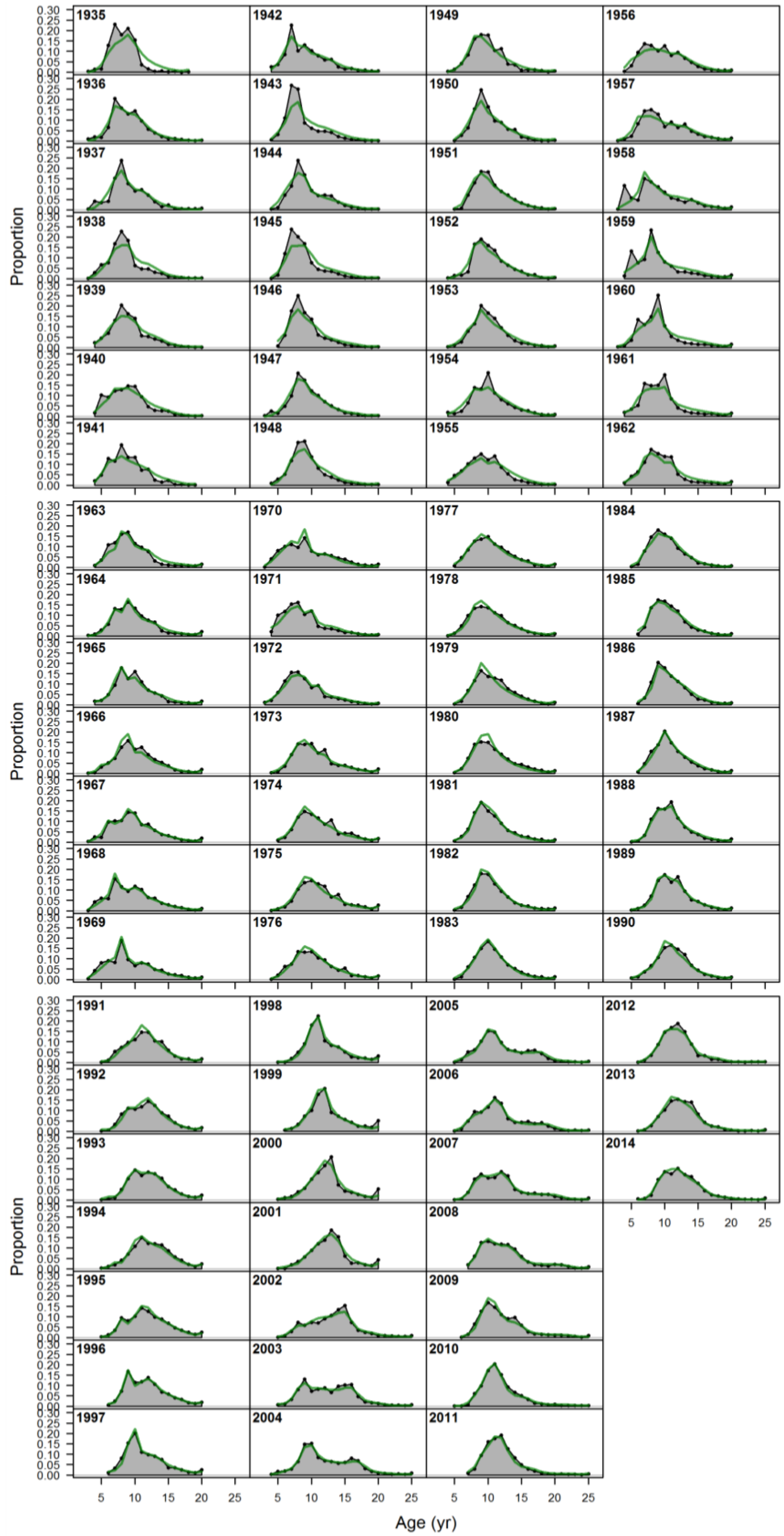


Figure 69. Fit to the age data for the commercial fishery in Area 2 in the AAF long model.

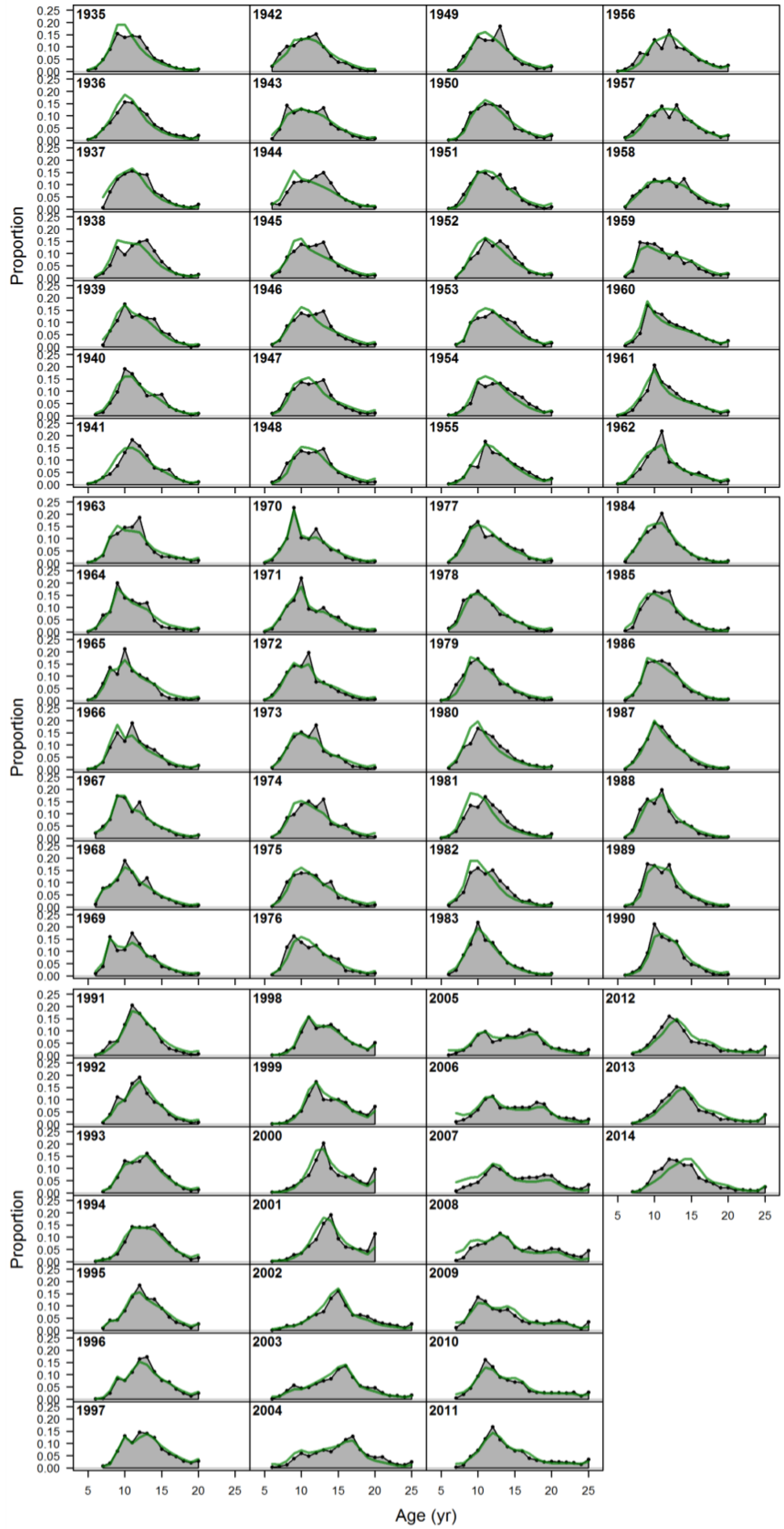


Figure 70. Fit to the age data for the commercial fishery in Area 3 in the AAF long model.

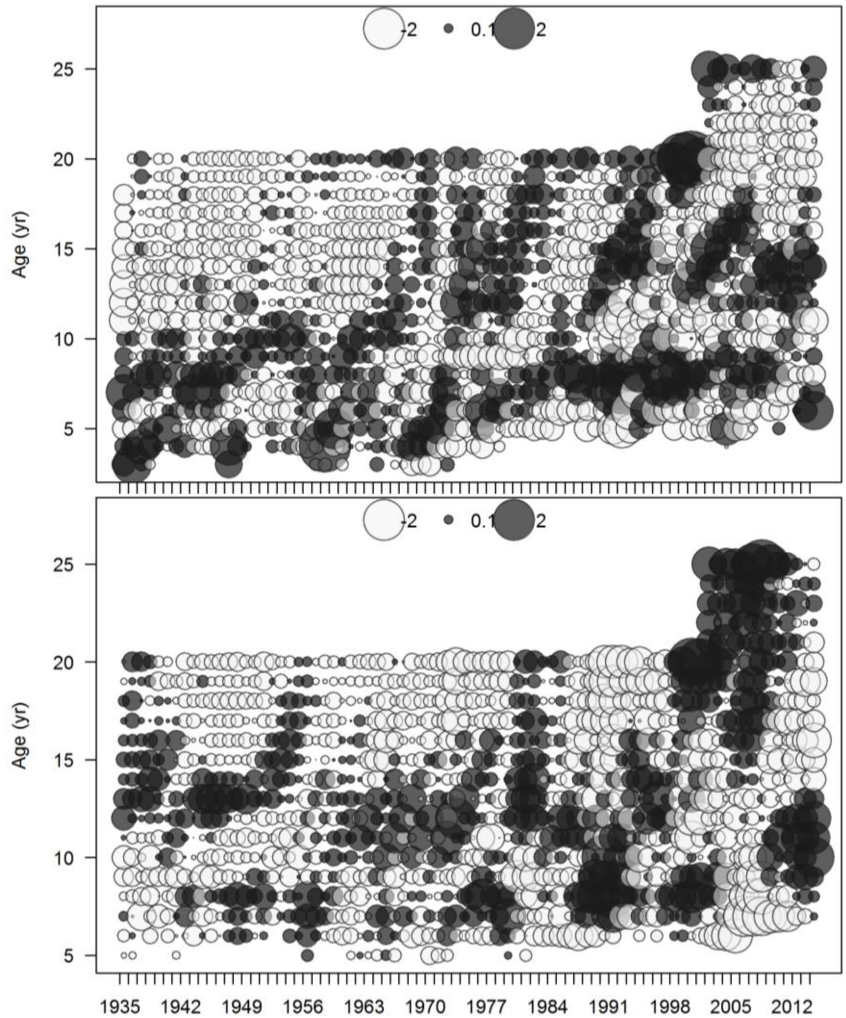


Figure 71. Standardized residuals to the age data for the commercial fishery in Area 2 (upper panel) and Area 3 (lower panel) in the AAF long model.

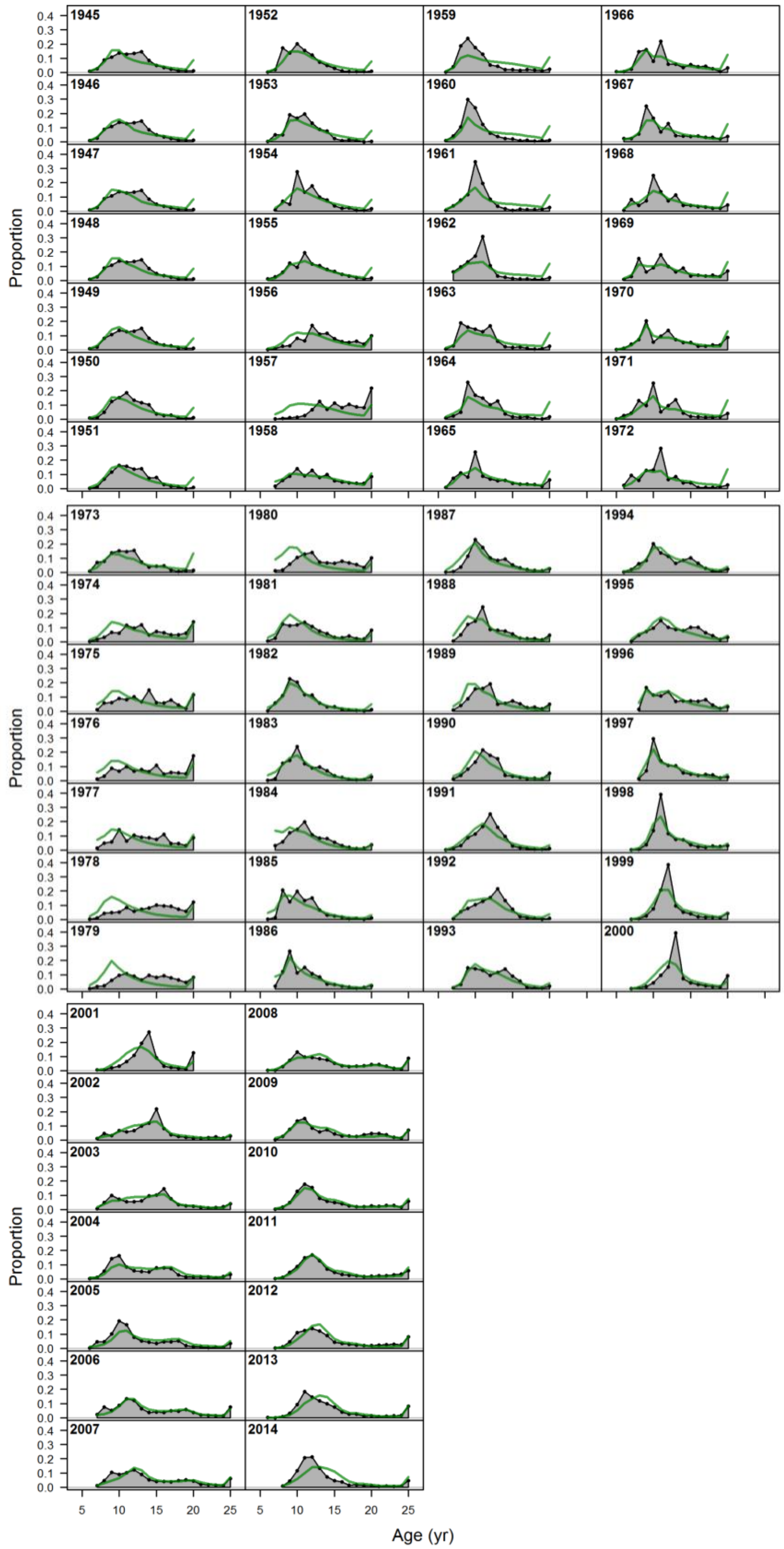


Figure 71. Fit to the age data for the commercial fishery in Area 4 in the AAF long model.

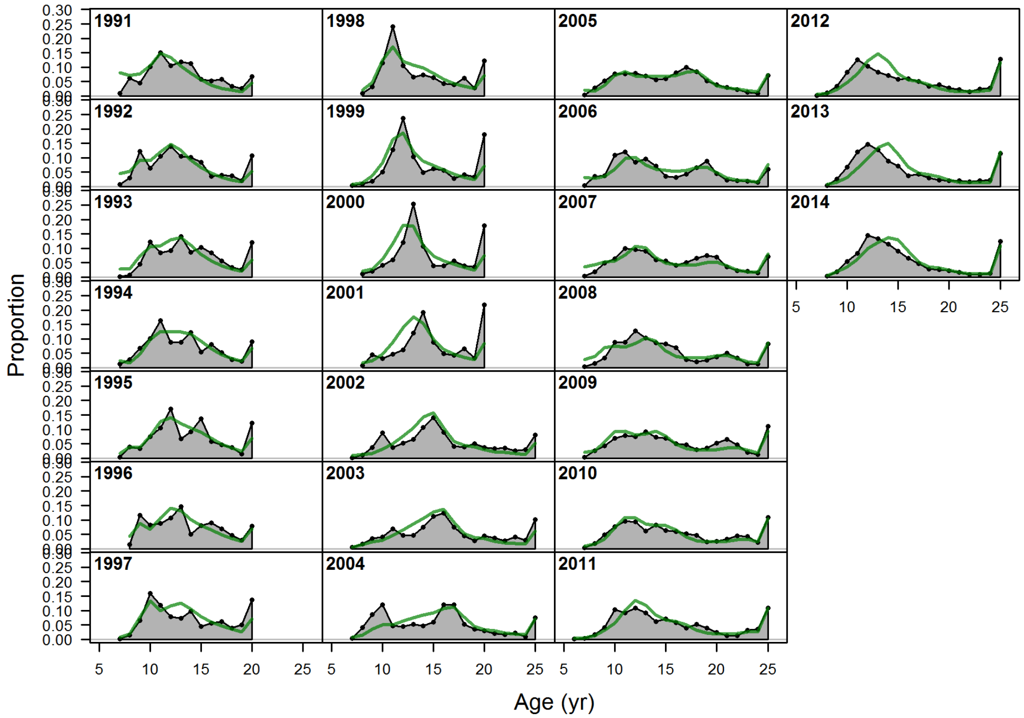


Figure 72. Fit to the age data for the commercial fishery in Area 4B in the AAF long model.

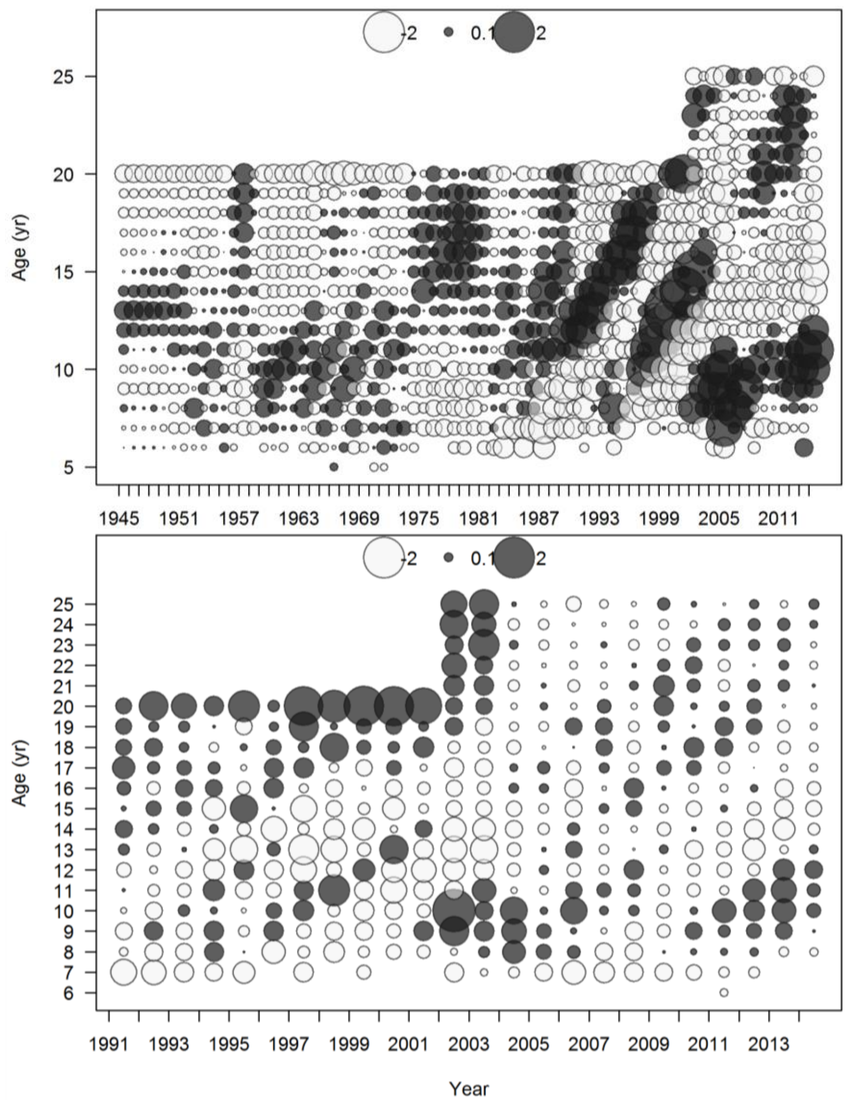


Figure 73. Standardized residuals to the age data for the commercial fishery in Area 4 (upper panel) and Area 4B (lower panel) in the AAF long model.

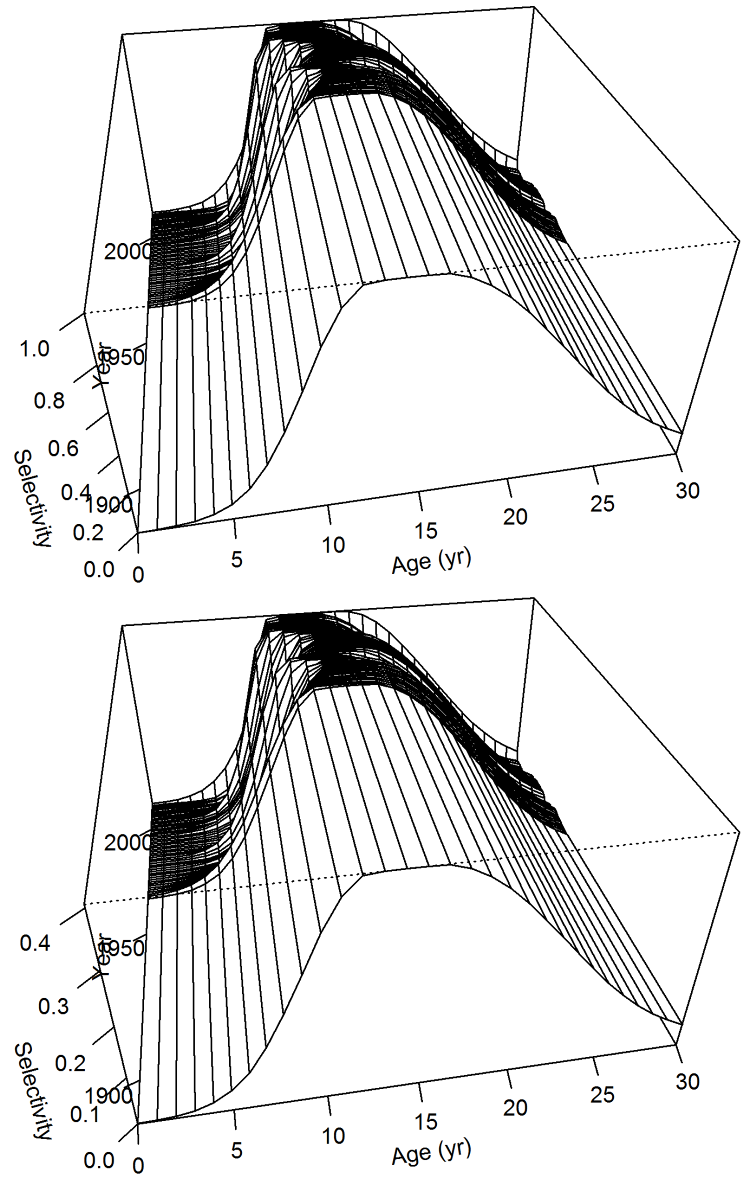


Figure 74. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 2 survey in the AAF long model.

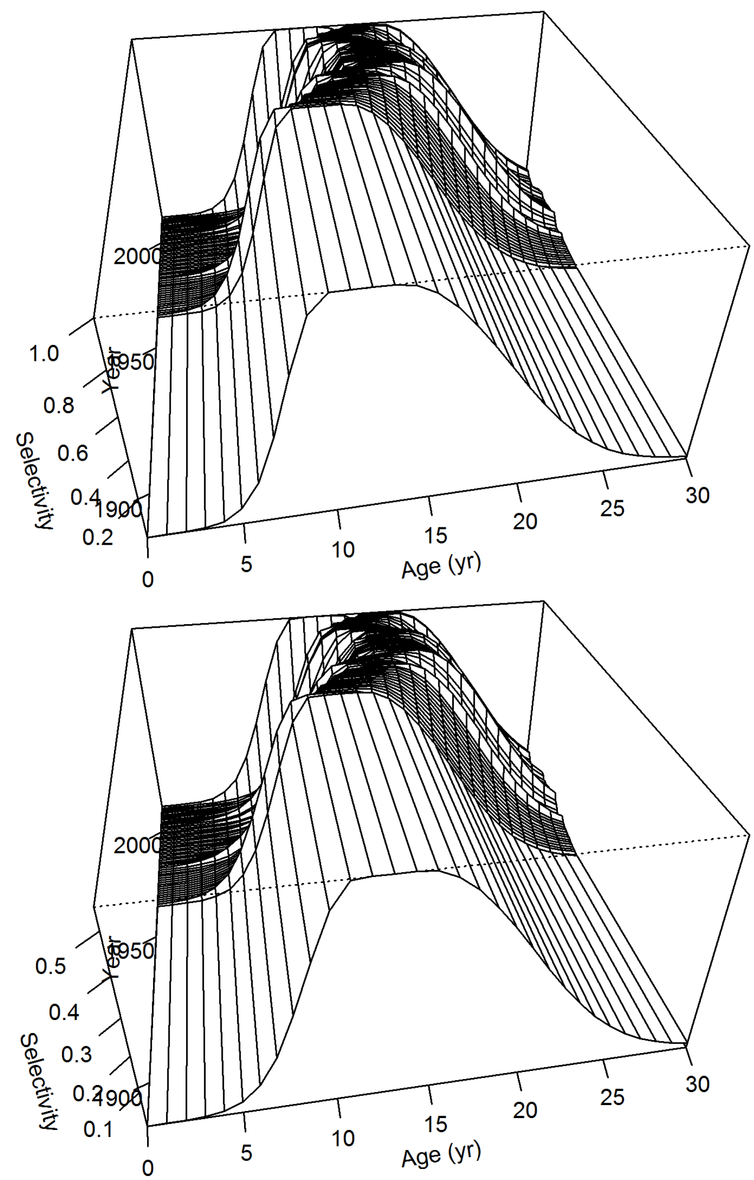


Figure 75. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 3 survey in the AAF long model.

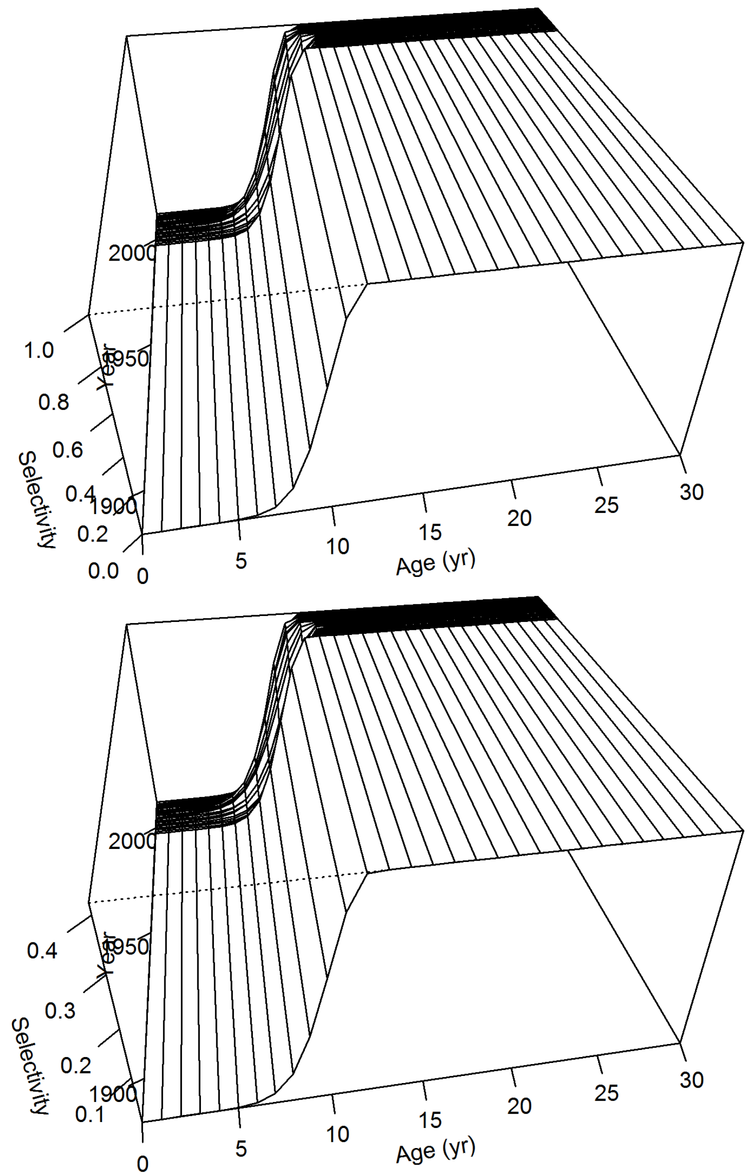


Figure 76. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 survey in the AAF long model.

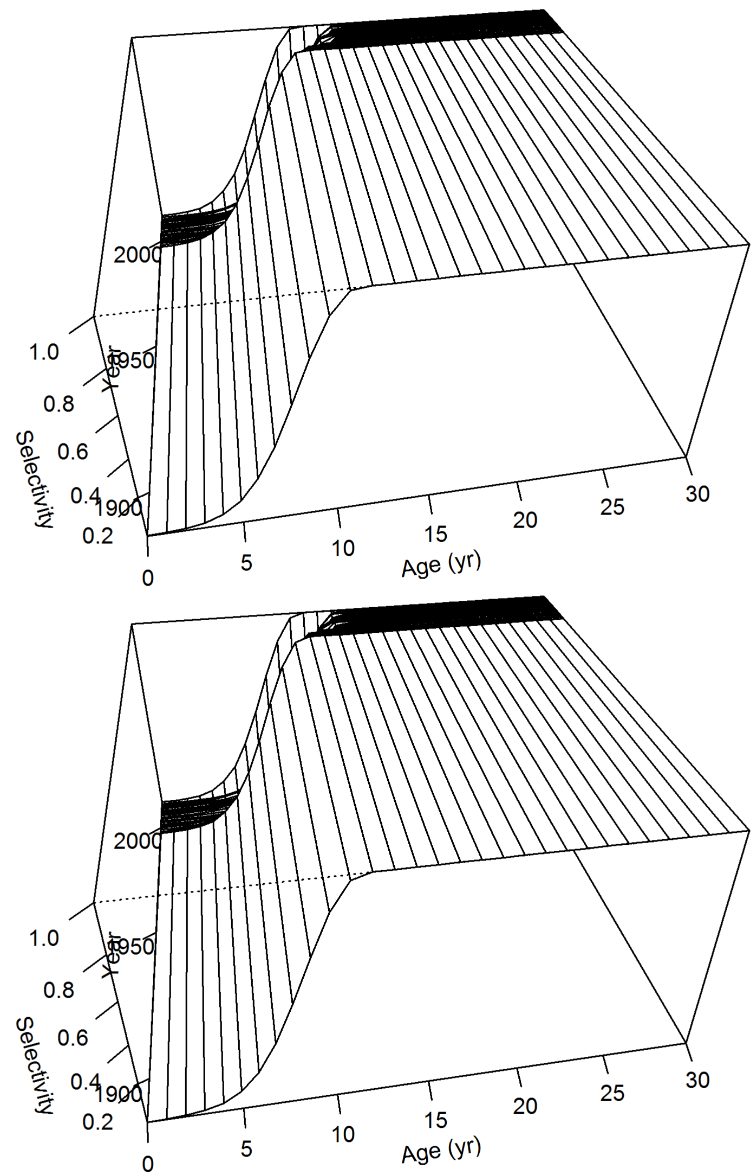


Figure 77. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B survey in the AAF long model.

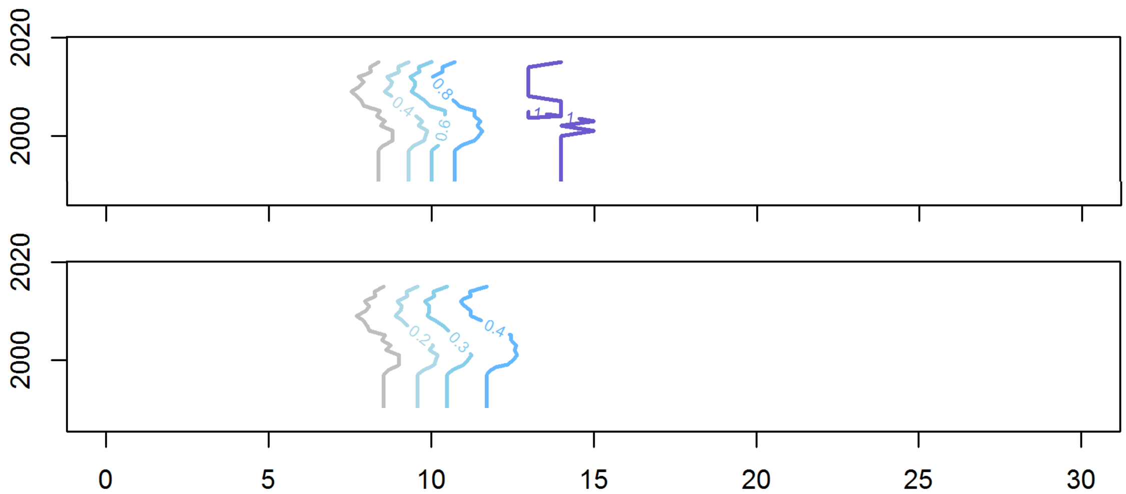


Figure 78. Recent selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 survey in the AAF long model.

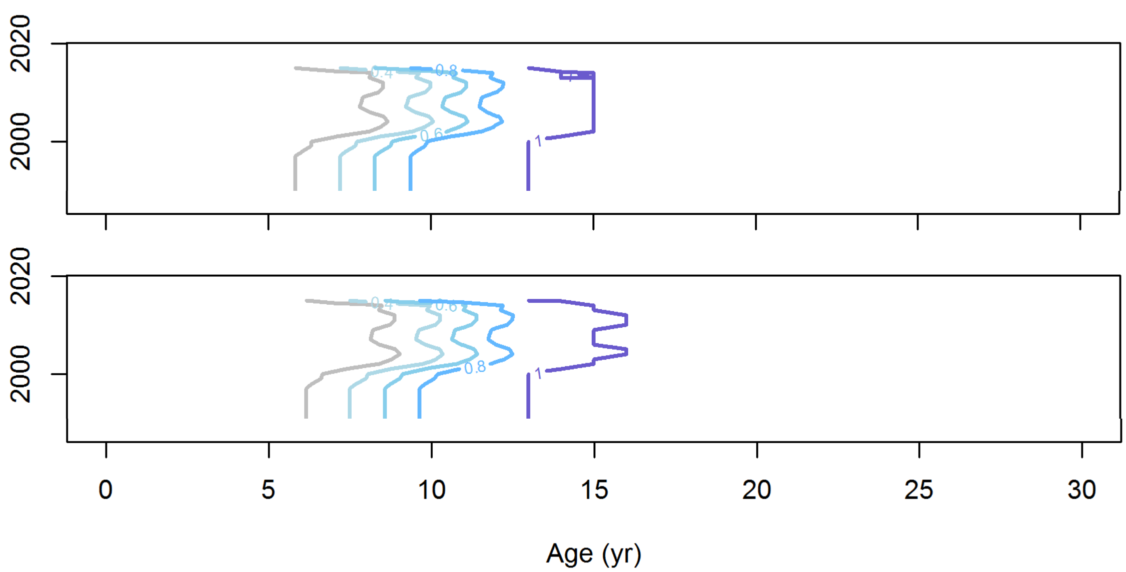


Figure 79. Recent selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B survey in the AAF long model.

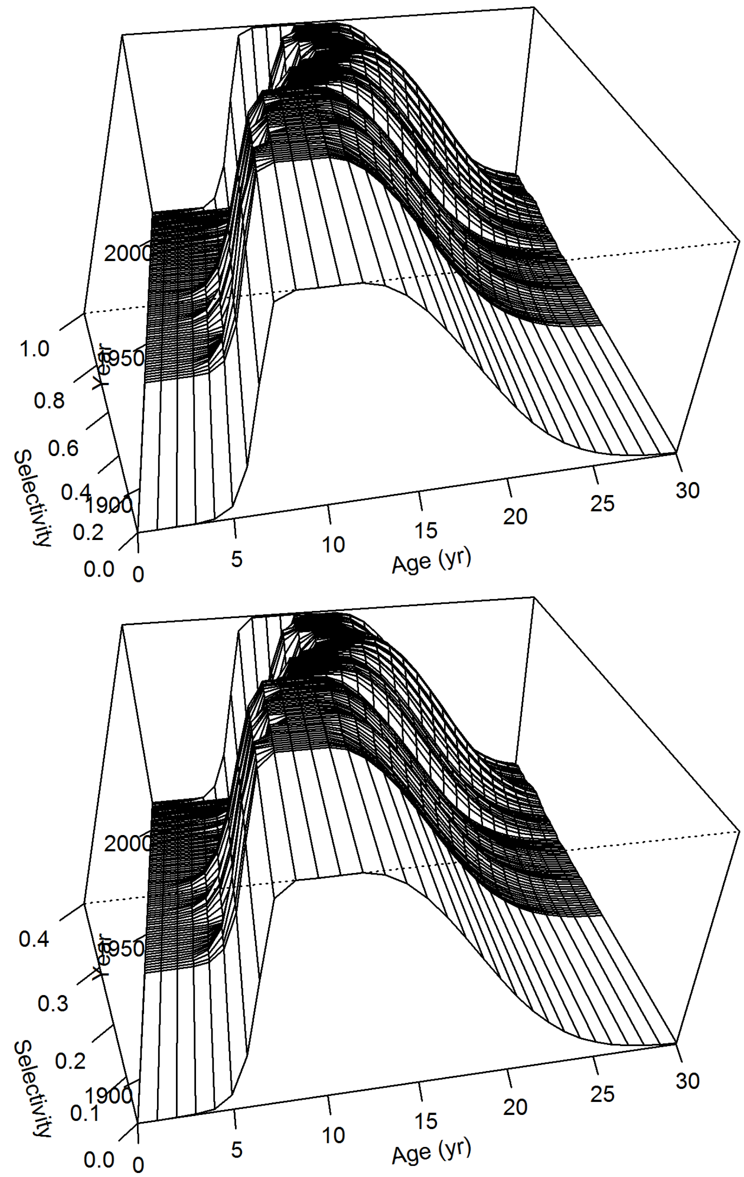


Figure 80. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 2 fishery in the AAF long model.

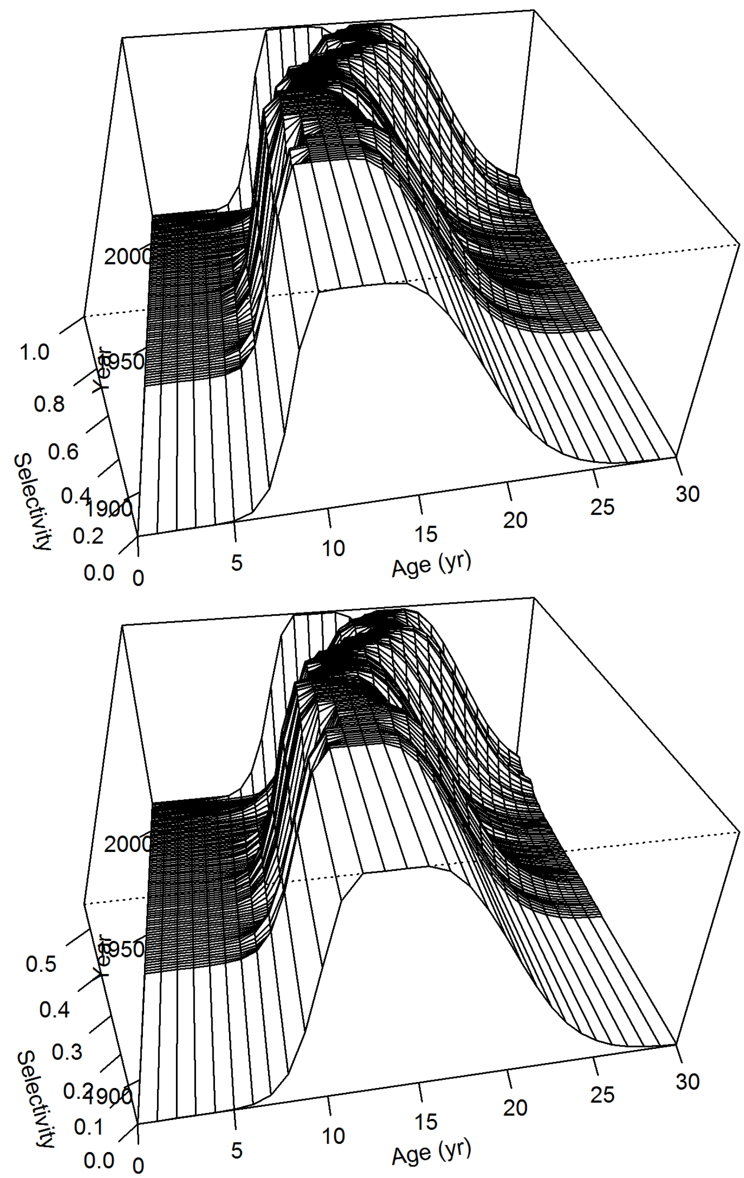


Figure 81. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 3 fishery in the AAF long model.

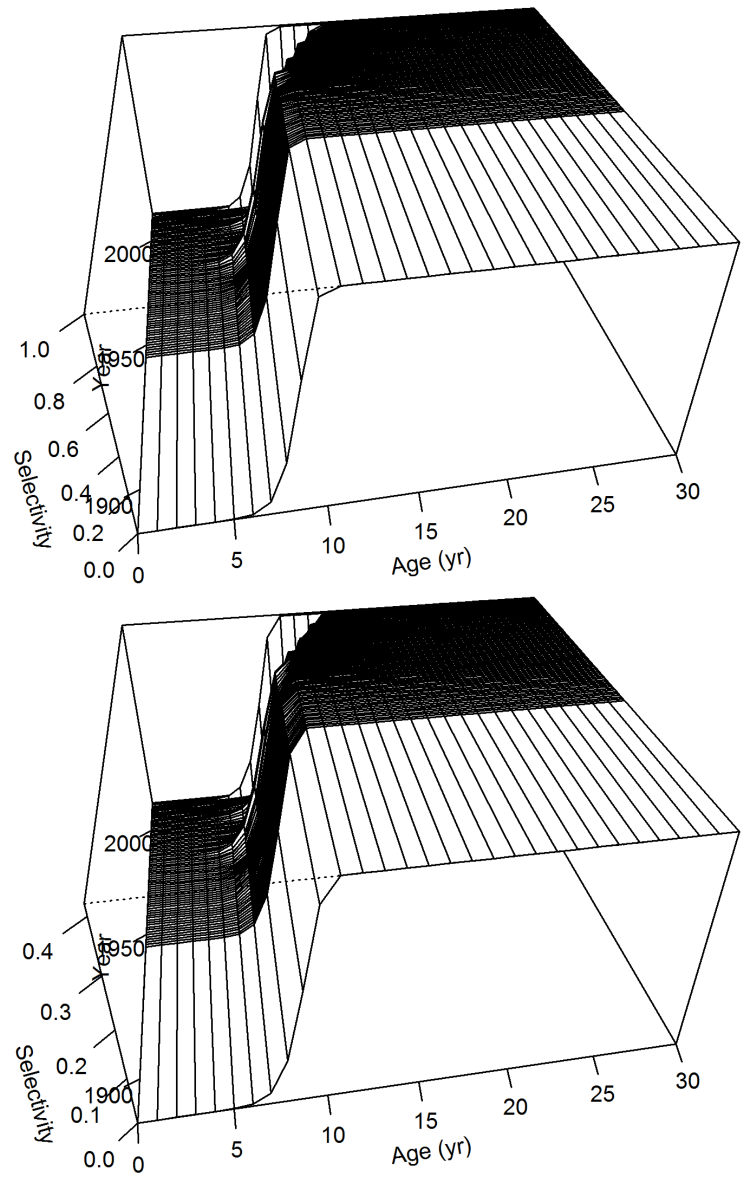


Figure 82. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 fishery in the AAF long model.

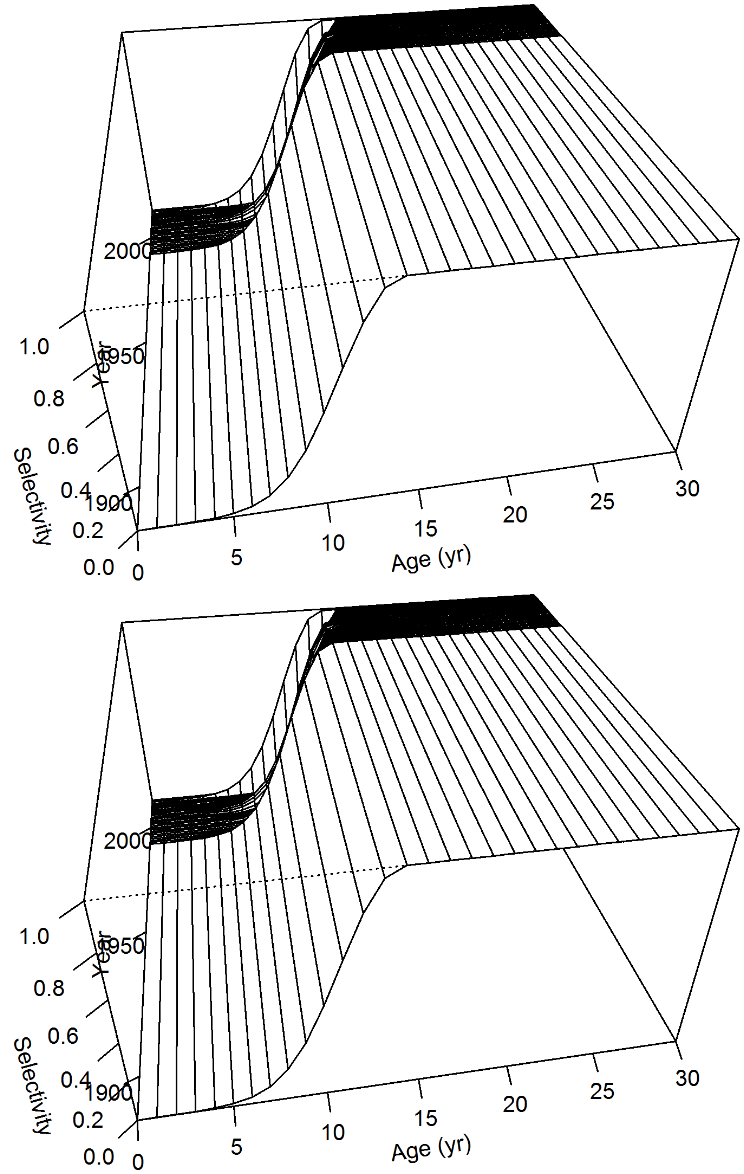


Figure 83. Selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B fishery in the AAF long model.

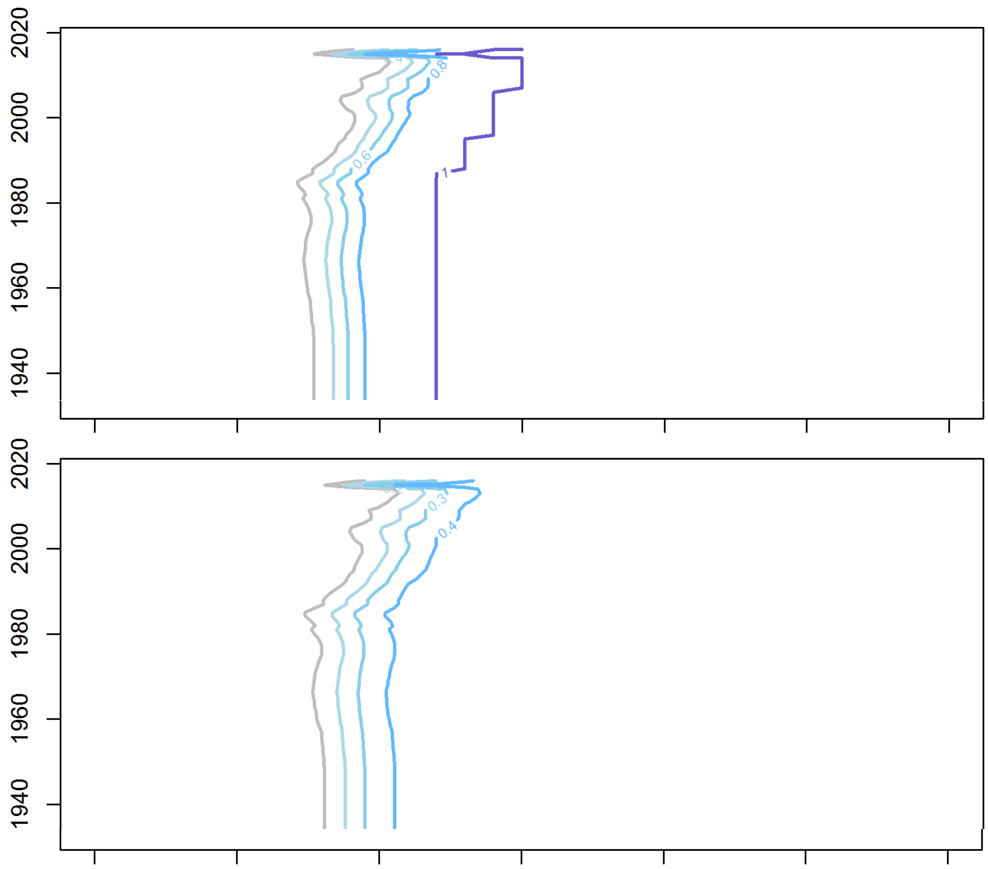


Figure 84. Recent selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4 fishery in the AAF long model.

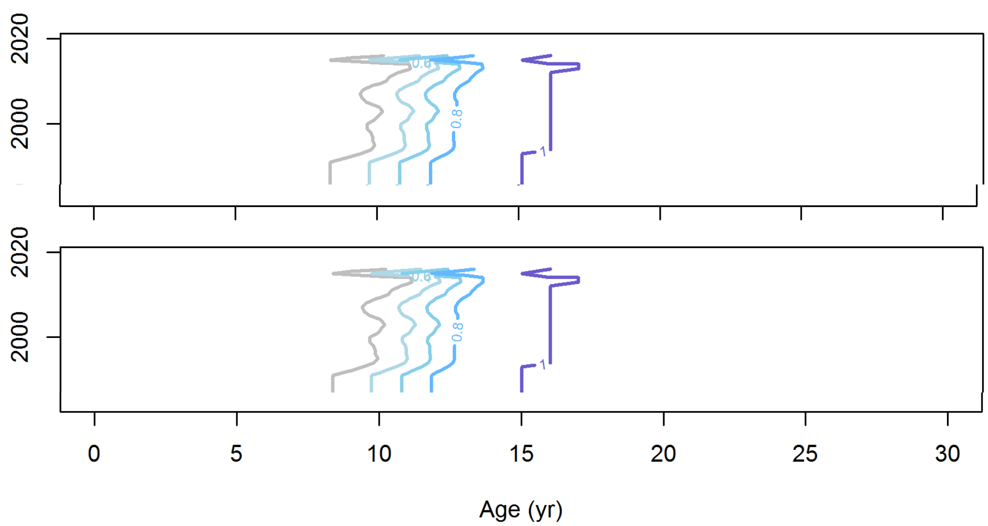


Figure 85. Recent selectivity patterns estimated for female (upper panel) and male (lower panel) halibut captured by the Area 4B fishery in the AAF long model.

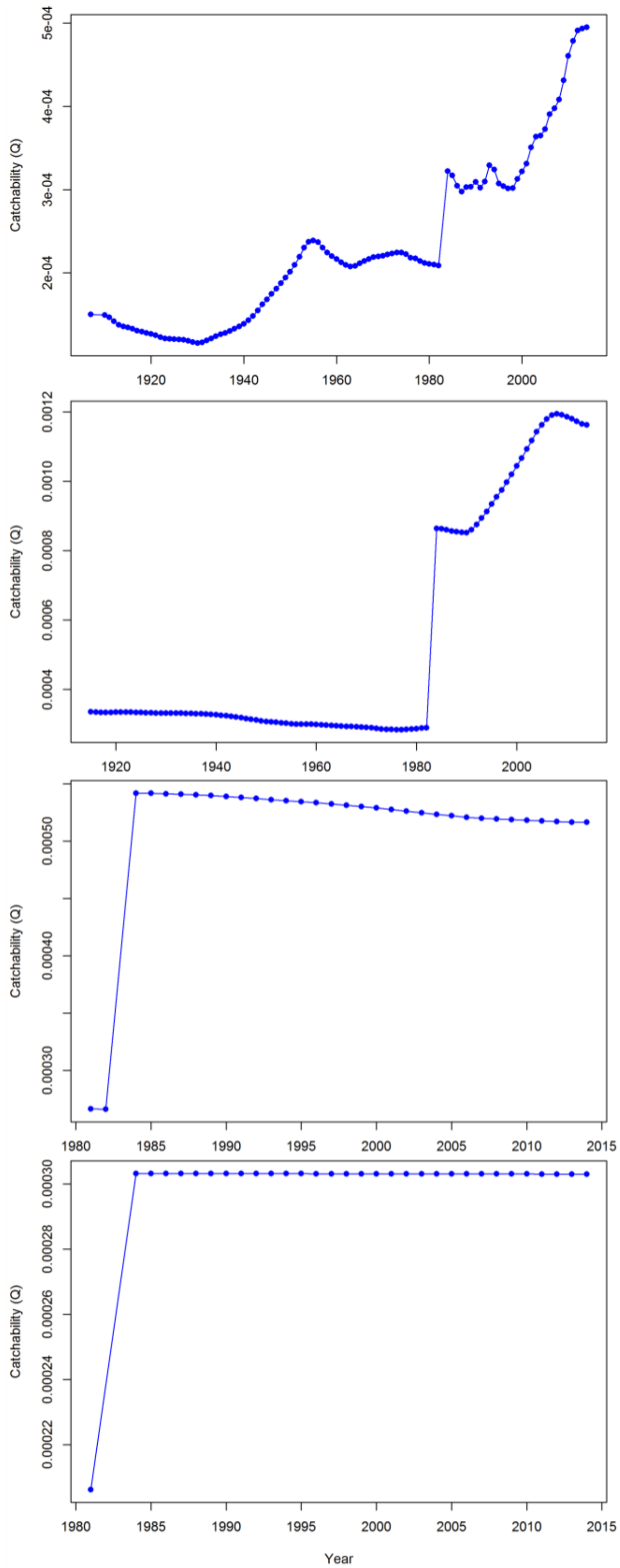


Figure 86. Estimated trends in fishery catchability for Area 2 (upper panel), Area 3 (second panel), Area 4 (third panel) and Area 4B (lower panel) in the AAF long model.

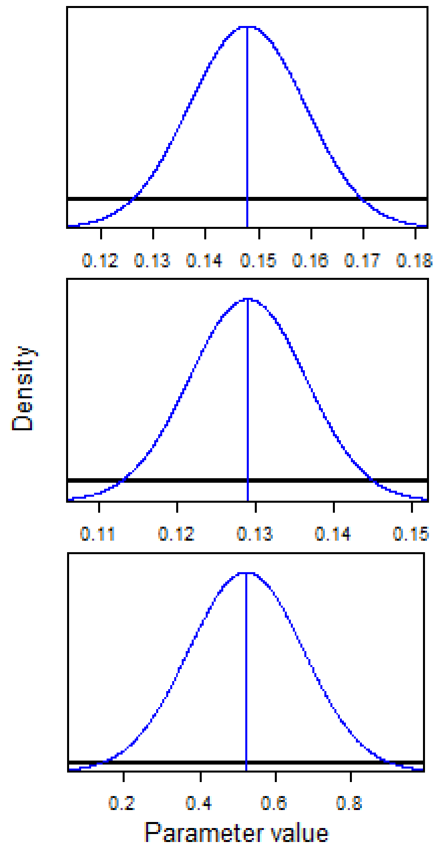


Figure 87. Prior (thick line) and MLE-based distributions for female natural mortality (top panel), male natural mortality (middle panel) and the environmental link coefficient (bottom panel) from the AAF long model.

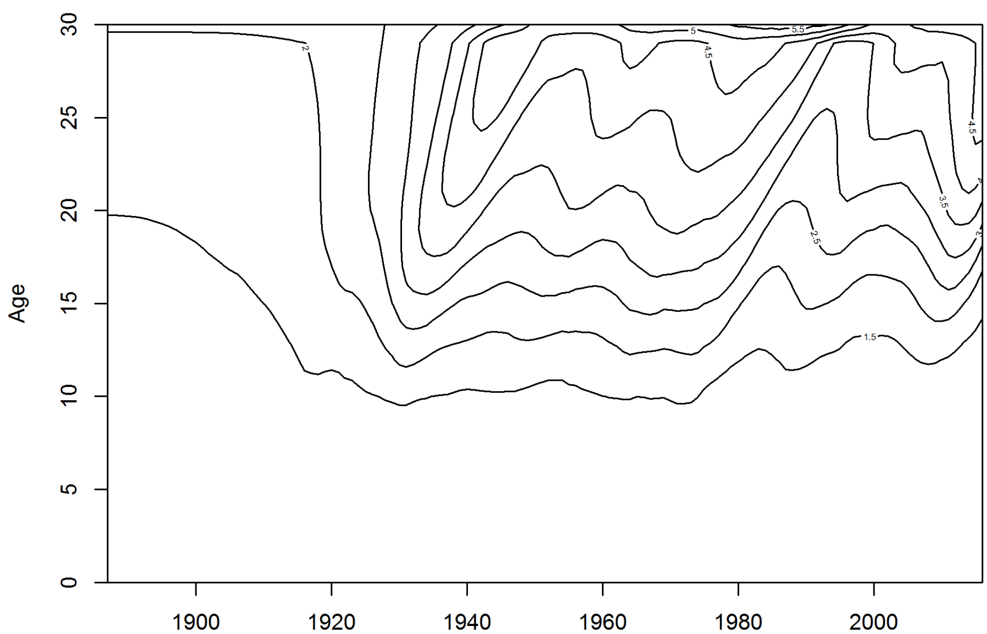


Figure 88. Population sex ratio (numbers of males/females) contours estimated by the AAF long model.

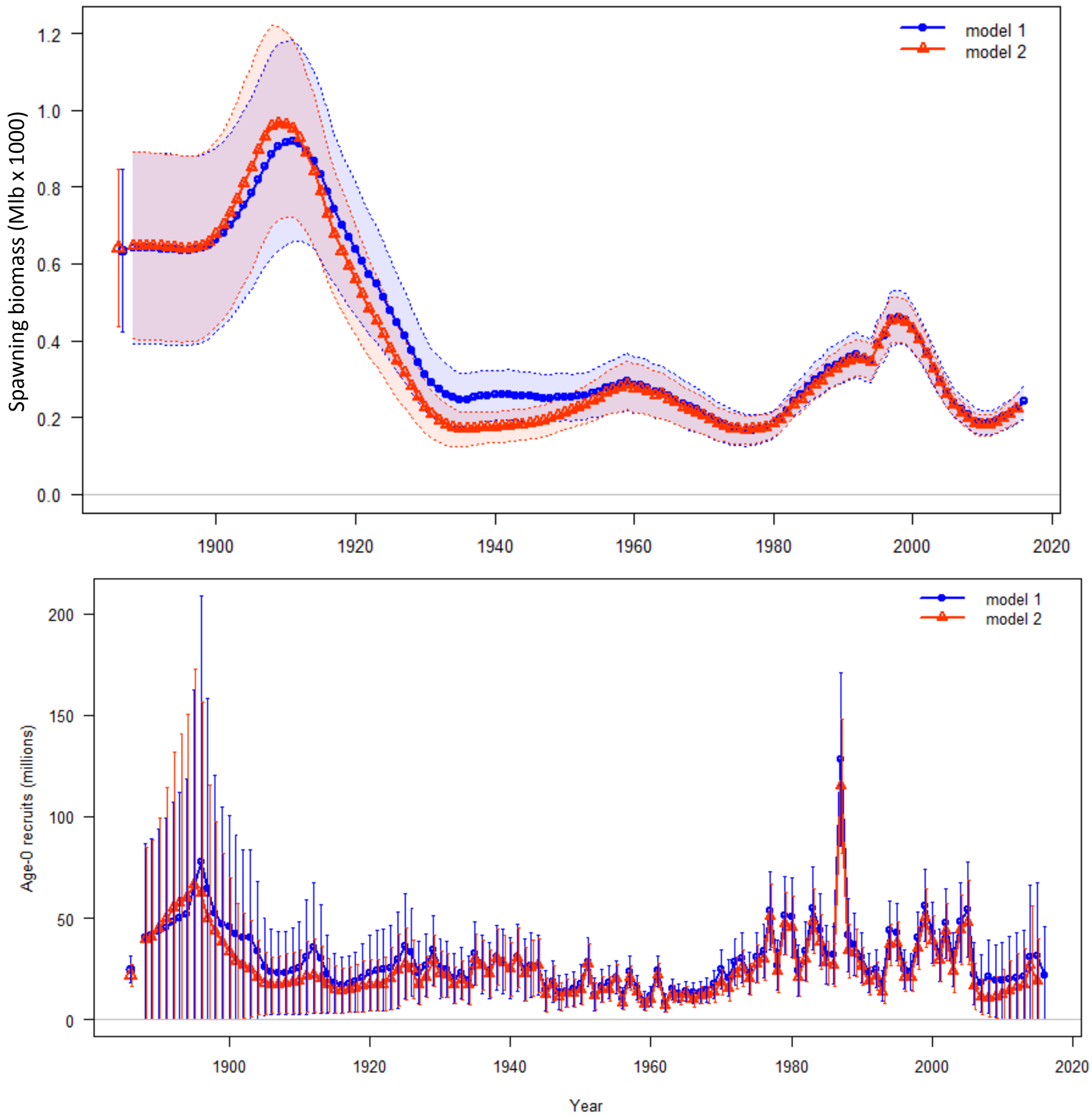


Figure 89. Comparison of the spawning biomass (upper panel) and recruitment (lower panel) estimated from the 2014, and preliminary 2015 AAF long model including all updates to data and model structure.

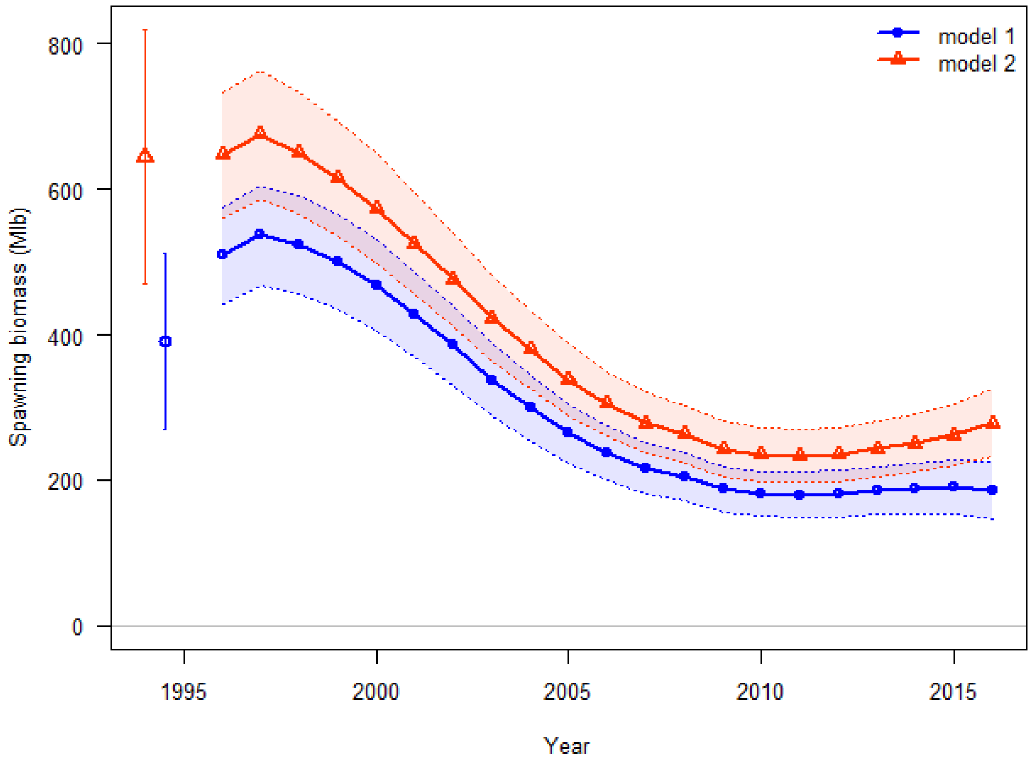


Figure 90. Comparison of the AAF short model (upper series) and the CW short model (lower series).

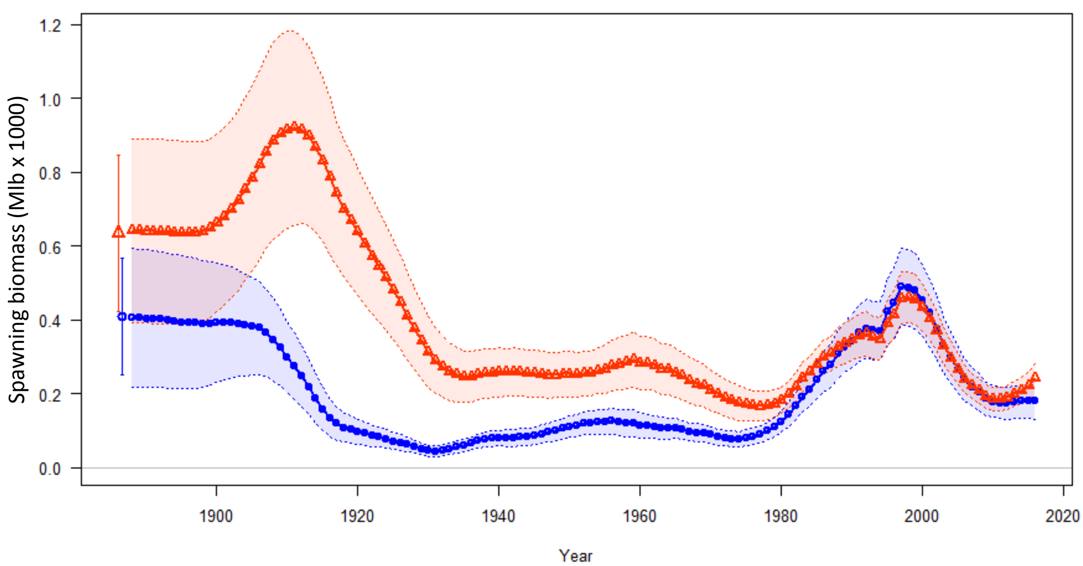


Figure 91. Comparison of the long AAF model (upper series in the early years), and the CW long model.

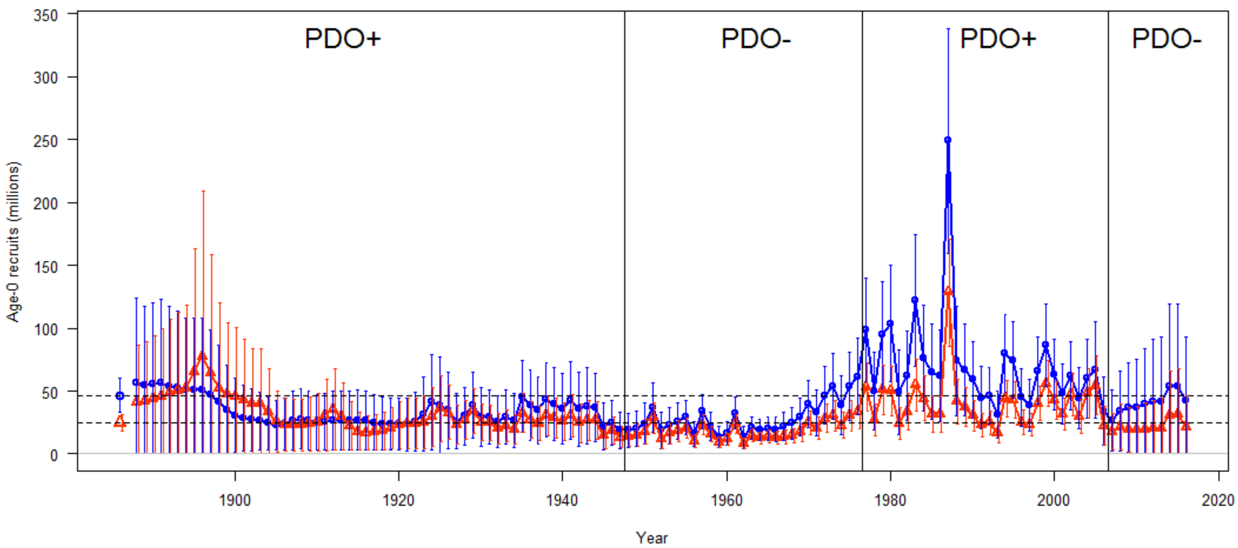


Figure 92. Comparison of recent recruitment estimates from the coastwide (blue series) and AAF long (red series) time-series models. Periods of the PDO are indicated by the vertical lines; mean recruitment for each model is indicated by the dashed lines.

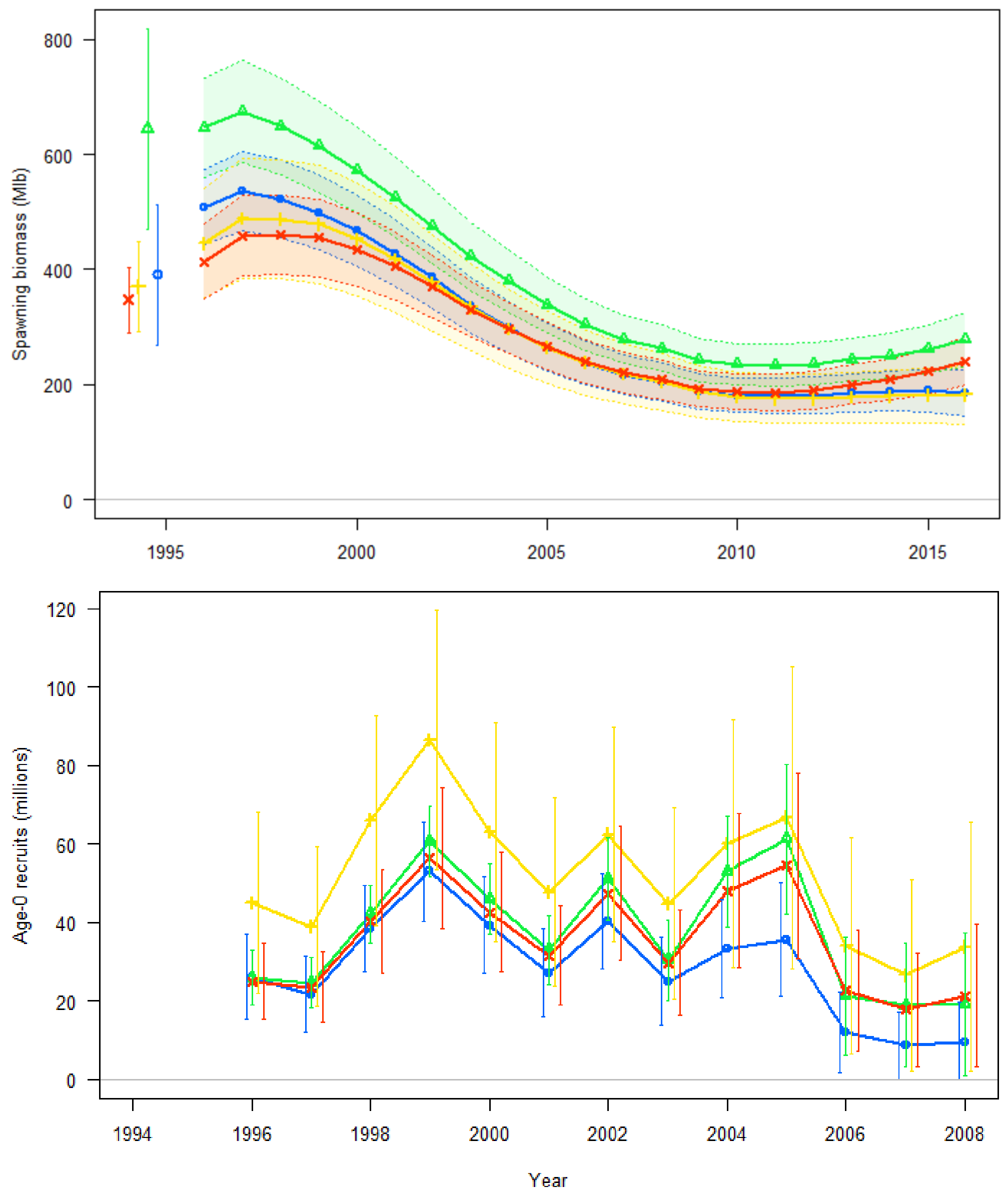


Figure 93. Comparison of recent spawning biomass (upper panel) and recruitment (lower panel) estimates from all four updated models (blue series = coastwide short, red = AAF short, yellow = coastwide long, green = AAF long).

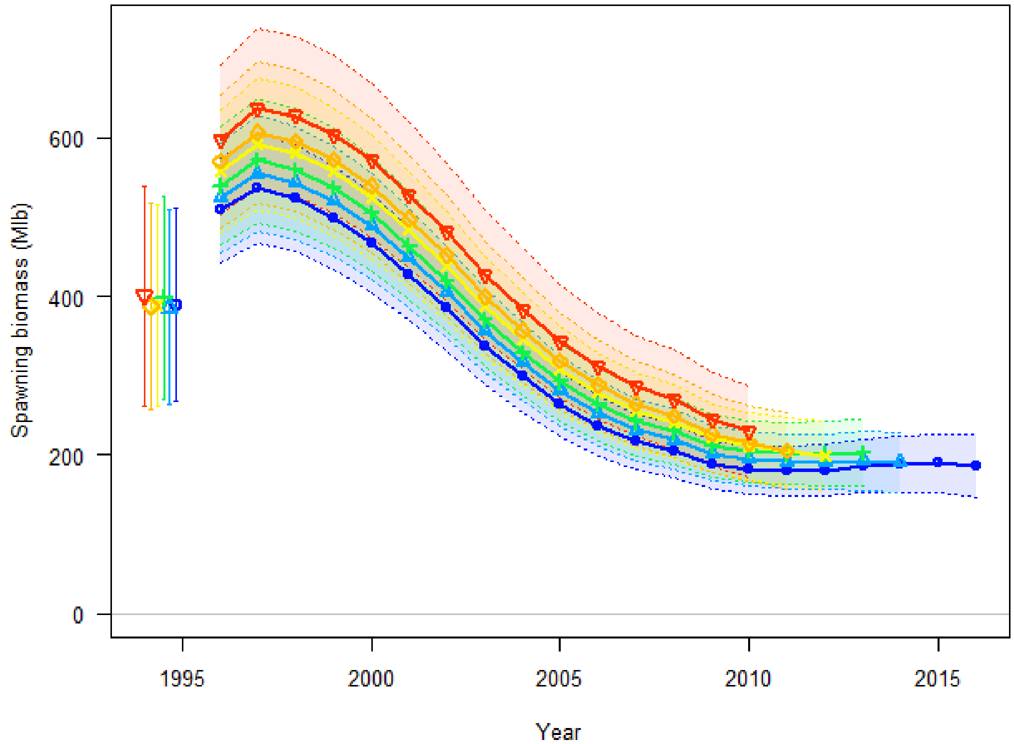


Figure 94. Five-year retrospective analysis (skipping 2015, which currently has no data) for the CW short model.

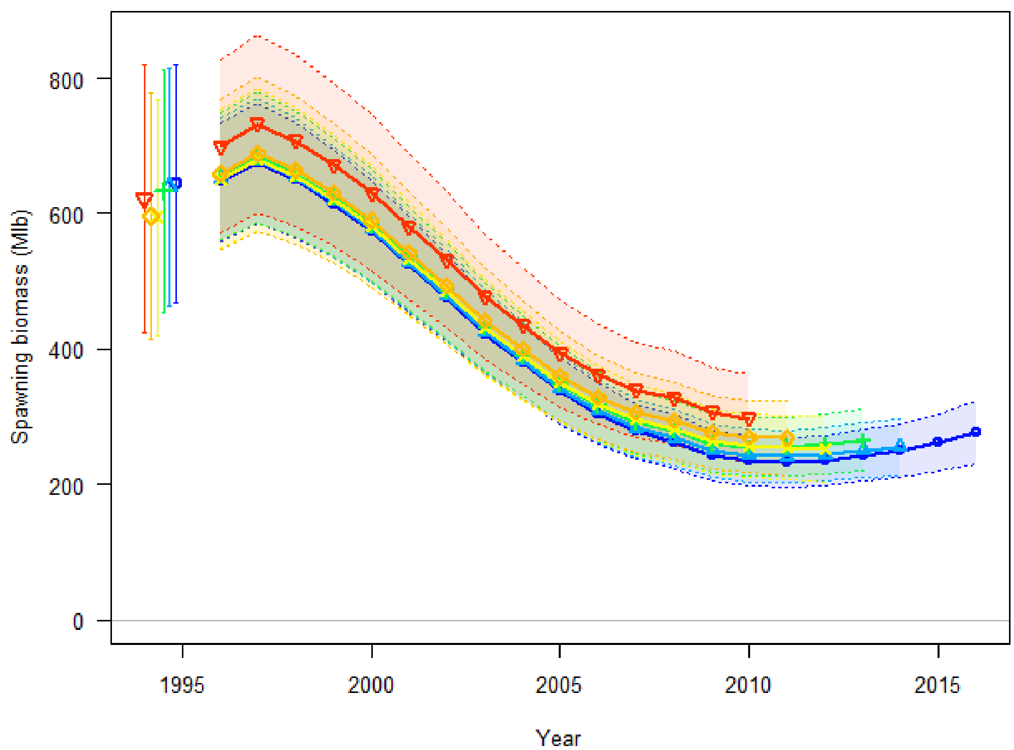


Figure 95. Five-year retrospective analysis (skipping 2015, which currently has no data) for the AAF short model.

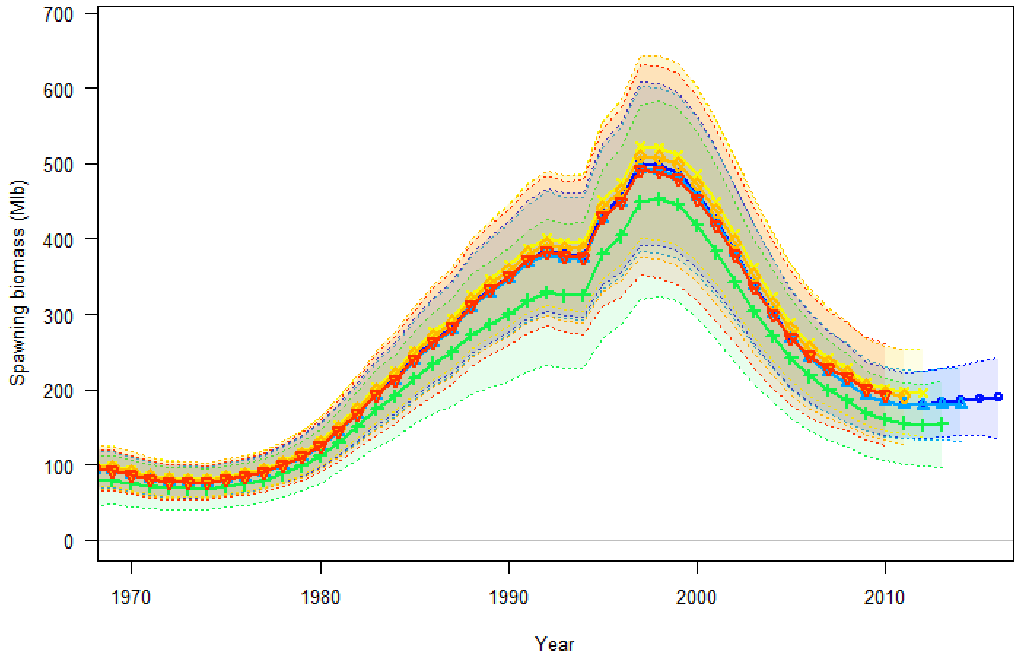


Figure 96. Five-year retrospective analysis (skipping 2015, which currently has no data) for the CW long model.

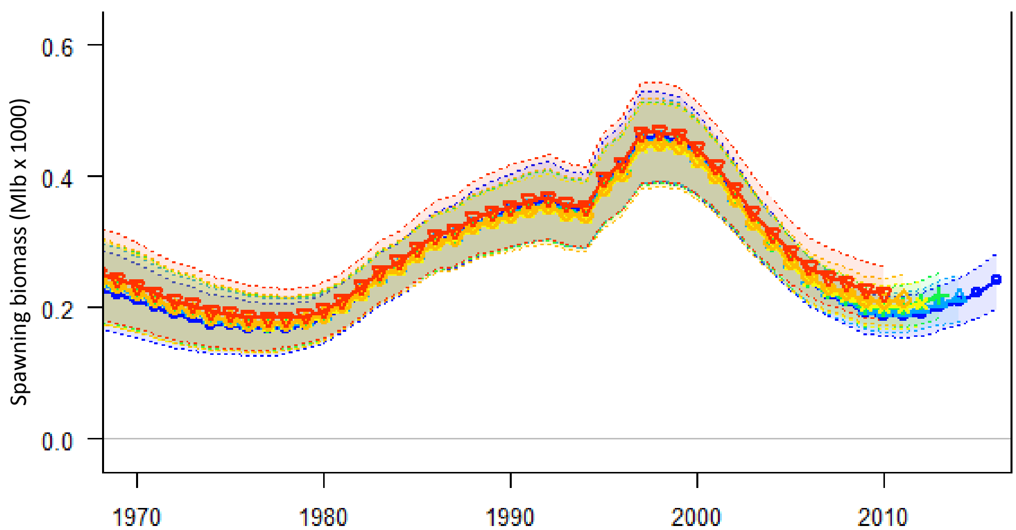


Figure 97. Five-year retrospective analysis (skipping 2015, which currently has no data) for the AAF long model.

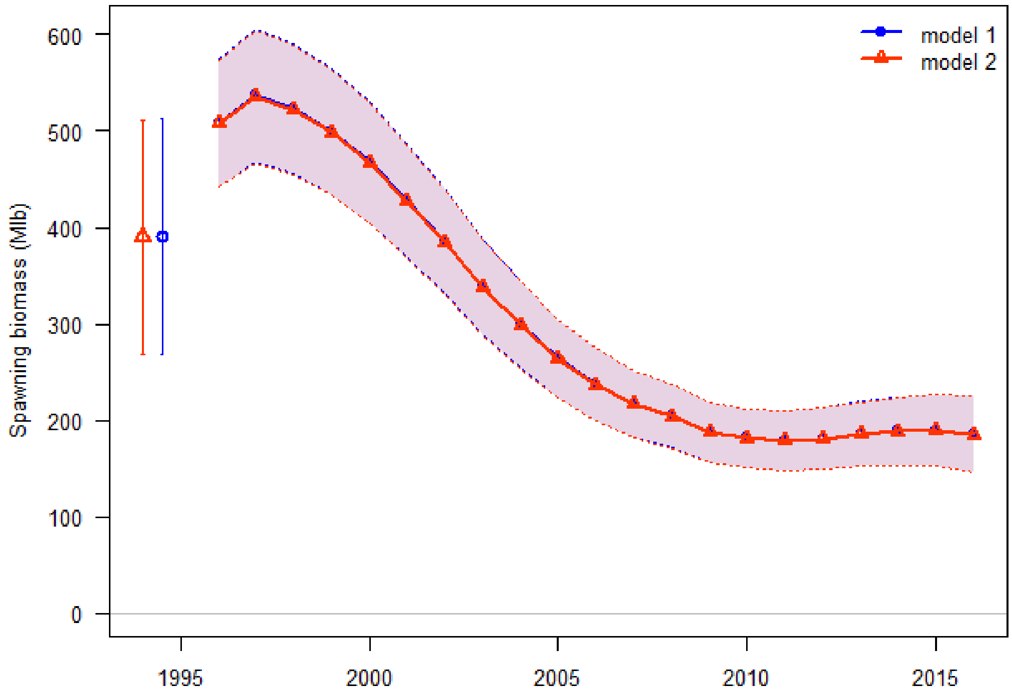


Figure 98. Comparison of the CW short model (model 1) and an alternative model that did not allow any temporal variation in fishery catchability (model 2).

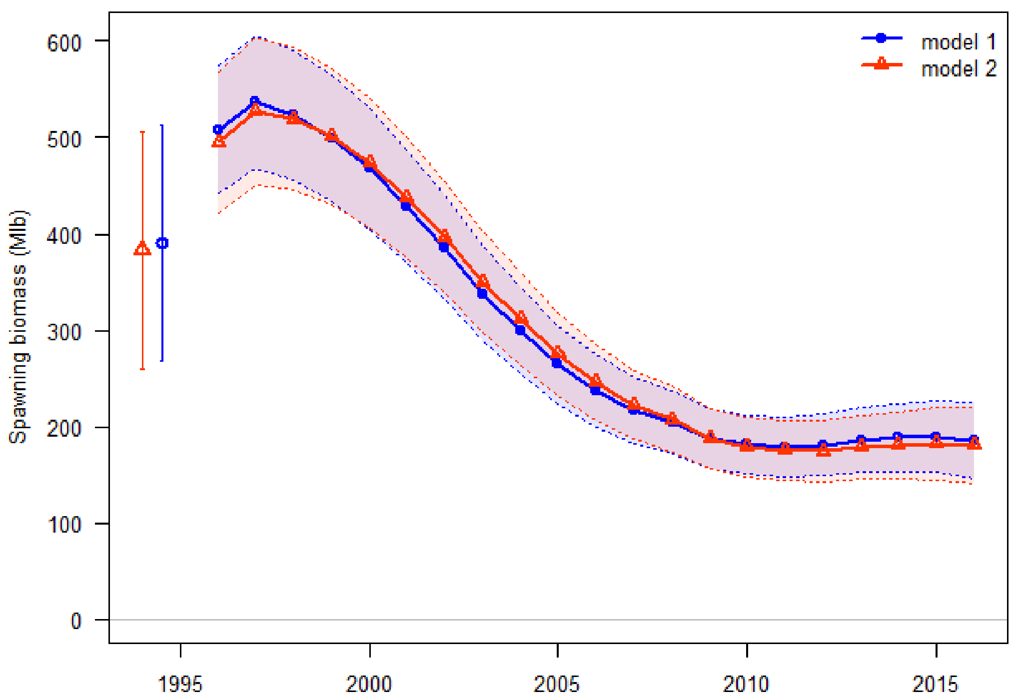


Figure 99. Comparison of the CW short model (model 1) and an alternative model that did not allow the ratio of male to female selectivity to vary over time (model 2).

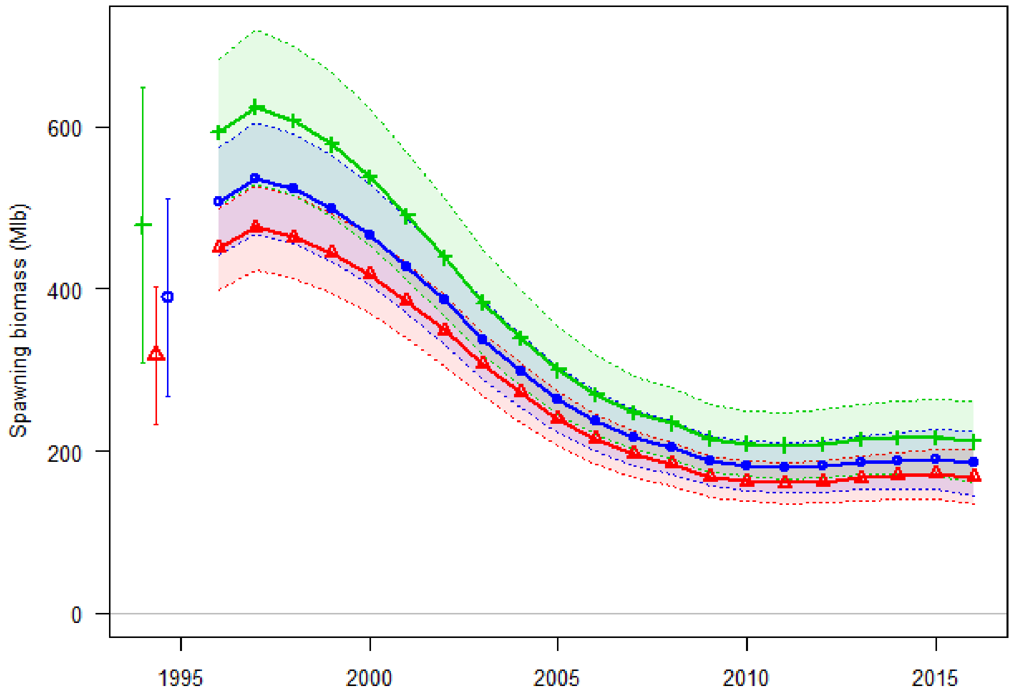


Figure 100. Comparison of the CW short model (middle series) with two alternative models forcing the natural mortality female halibut to be 0.18 (upper series) and 0.12 (lower series).

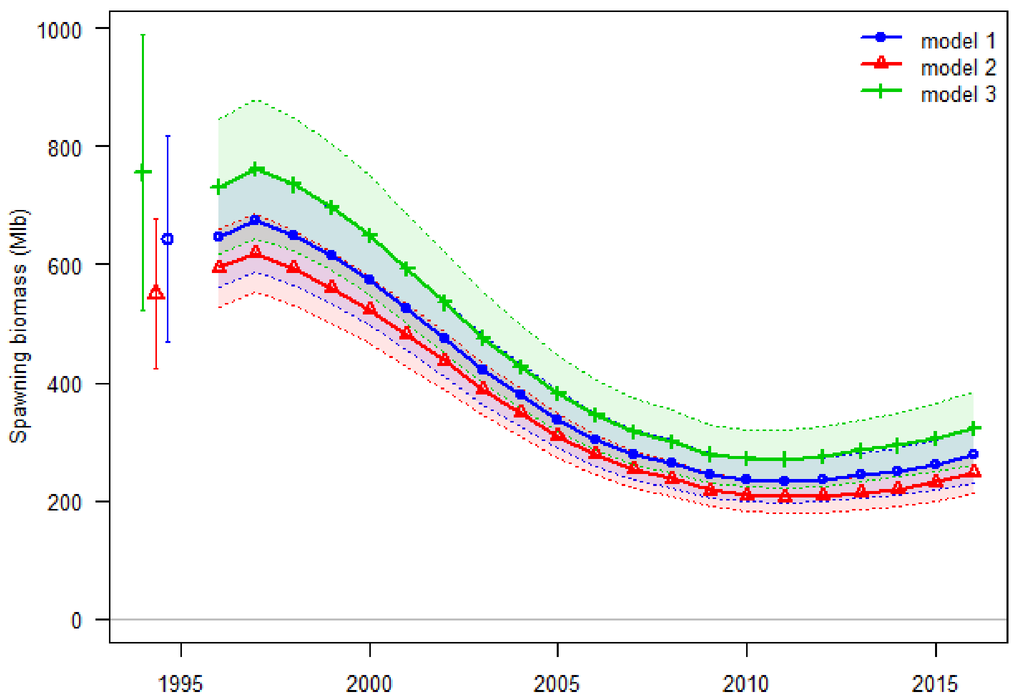


Figure 101. Comparison of the AAF short model (middle series) with two alternative models forcing the natural mortality female halibut to be 0.18 (upper series) and 0.12 (lower series).

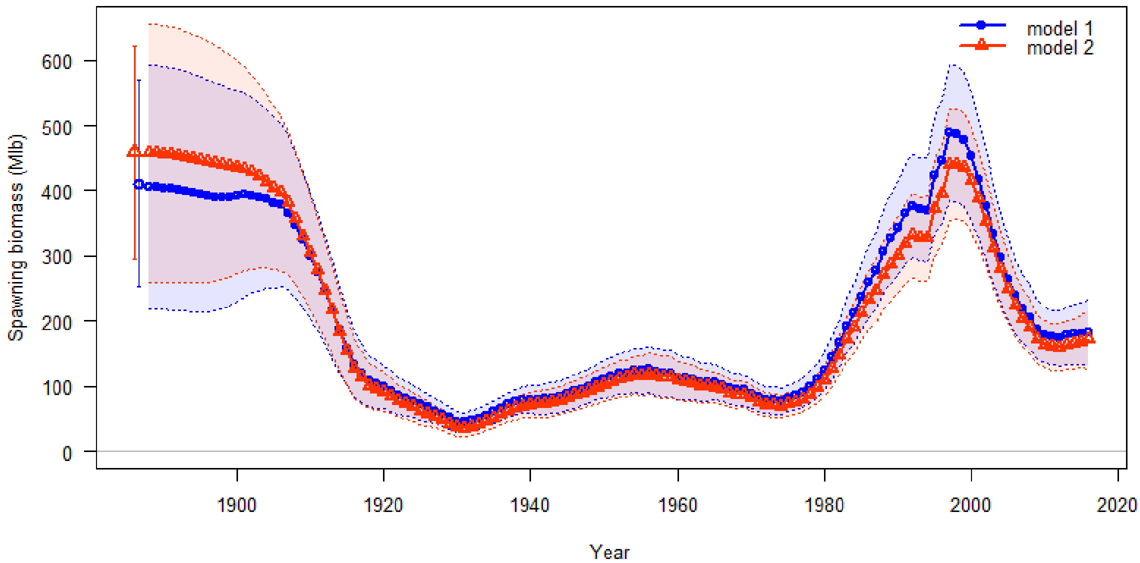


Figure 102. Comparison of the CW long model (model 1) with an alternative model allowing steepness to be estimated at 1.0 (model 2).

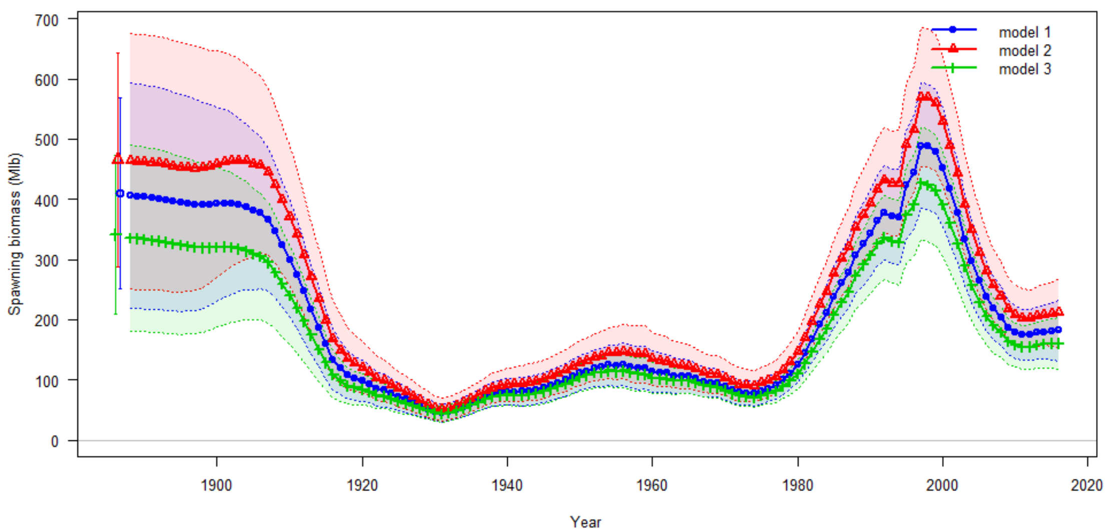


Figure 103. Comparison of the CW long model (middle series) with two alternative models making the relative selectivity of male halibut 10% lower (upper series) and 10% higher (lower series).

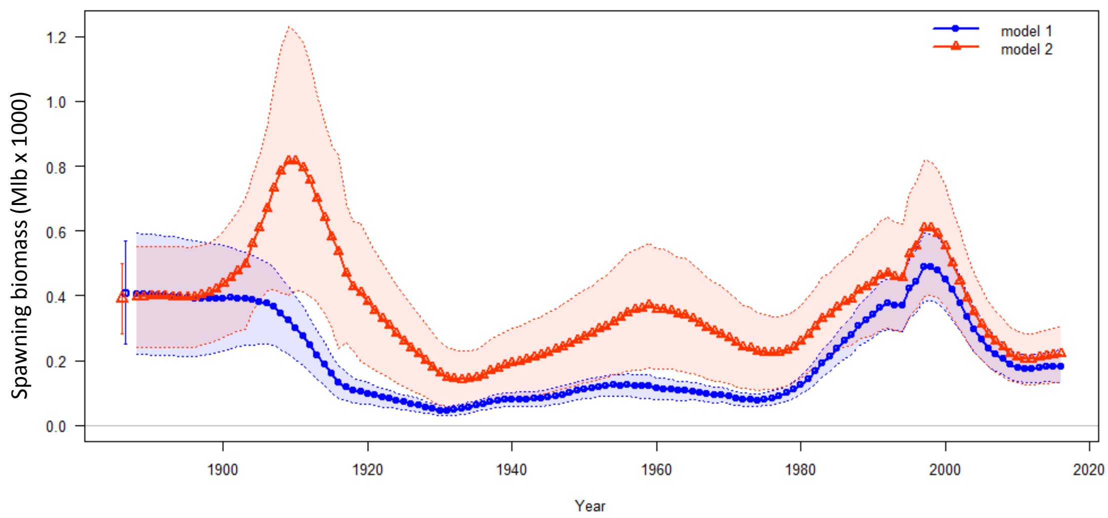


Figure 104. Comparison of the CW long model (lower series) with the CW long model allowing dome-shaped selectivity during the time-periods <1958, 1959-1980, 1981-1996.

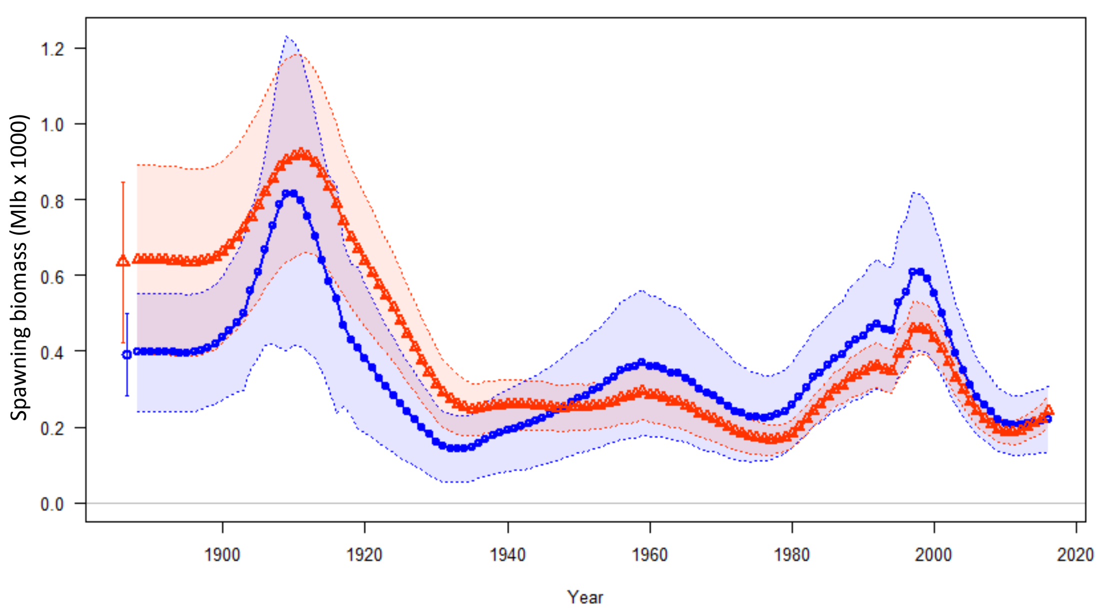


Figure 105. Comparison of spawning biomass estimates from the long AAF model (red series) and a long CW model (blue series) allowing dome-shaped selectivity during the time-periods <1958, 1959-1980, 1981-1996 (blue series; lower at the start of the time-period).

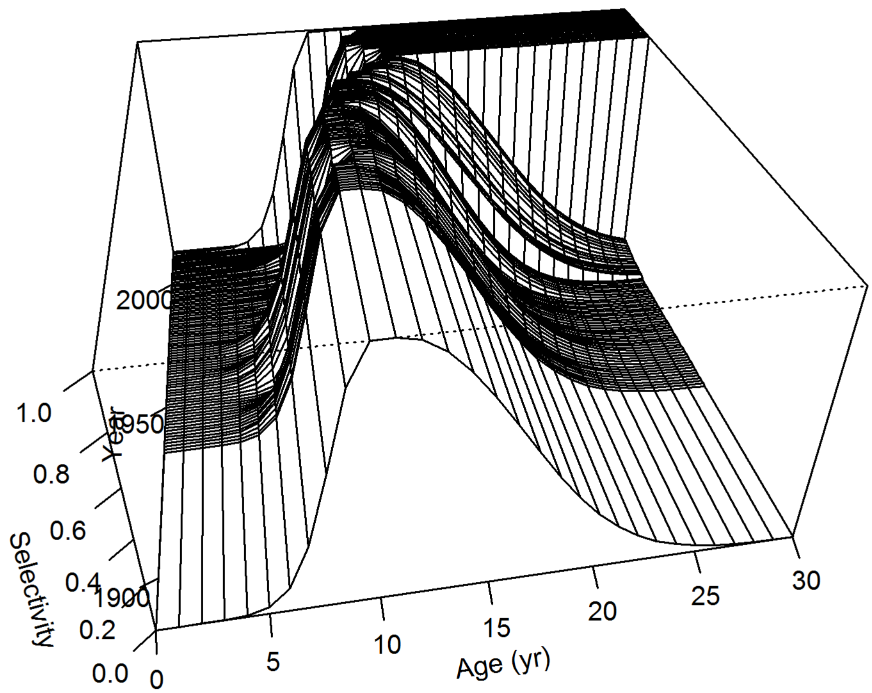


Figure 106. Estimated female selectivity pattern for the long CW model allowing dome-shaped selectivity during the time-periods <1958, 1959-1980, 1981-1996.

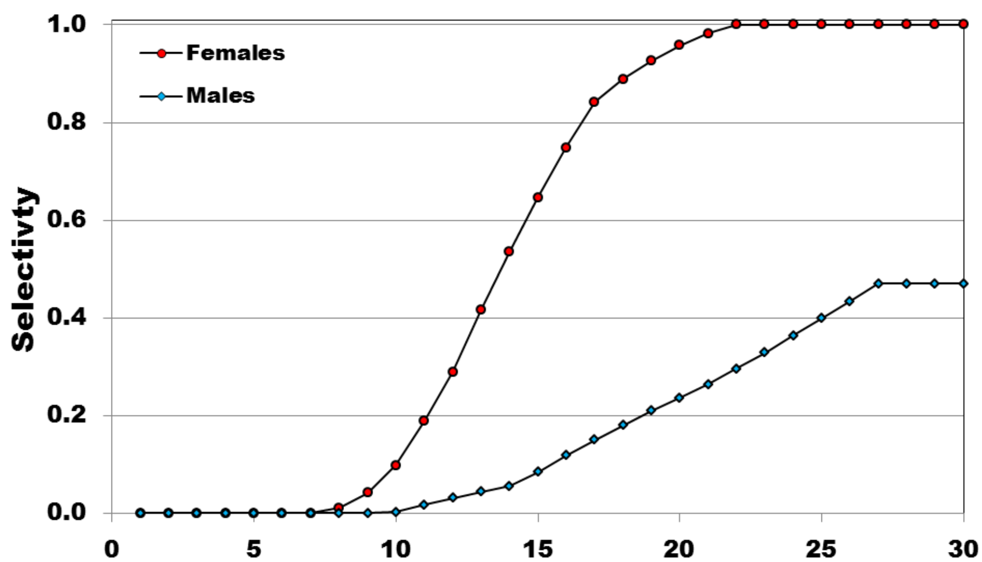


Figure 107. Terminal year selectivity pattern assigned to the exploitable biomass calculation, consistent with the existing IPHC harvest policy.

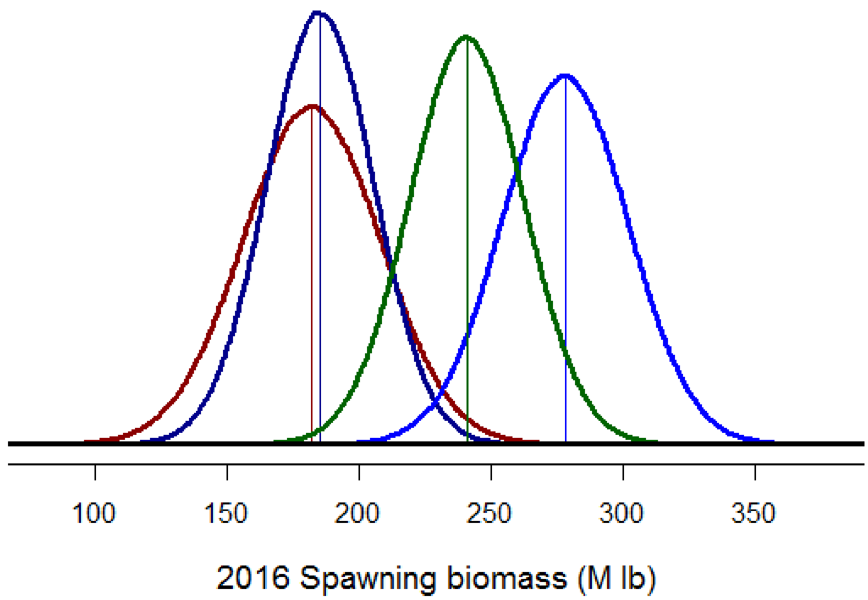


Figure 108. Estimated distributions from each of the four preliminary models for 2016 spawning biomass.

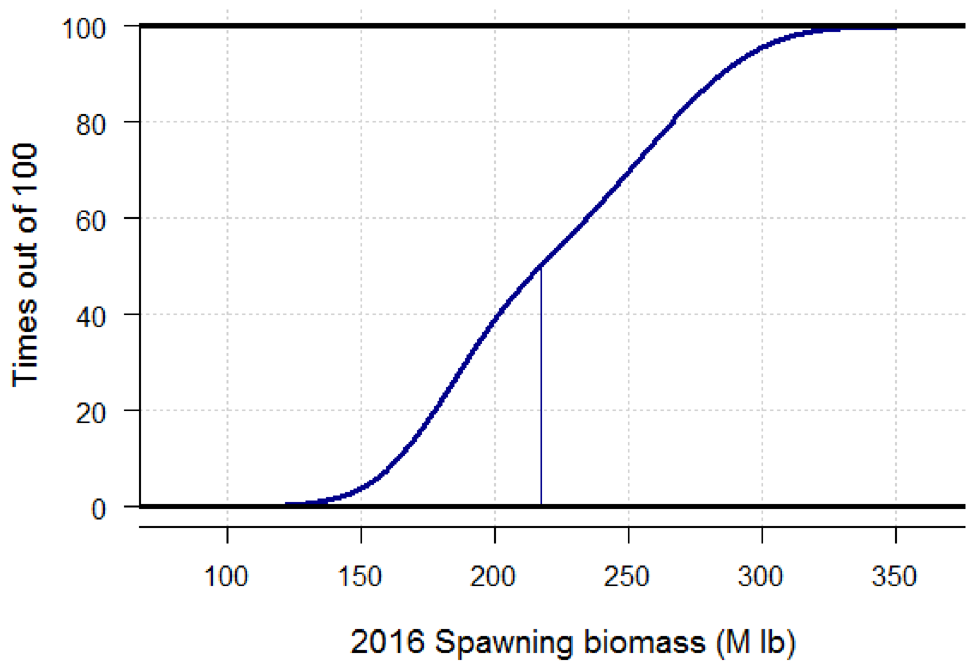


Figure 109. Cumulative distribution function for the integrated estimate of 2016 spawning biomass from the four preliminary models.

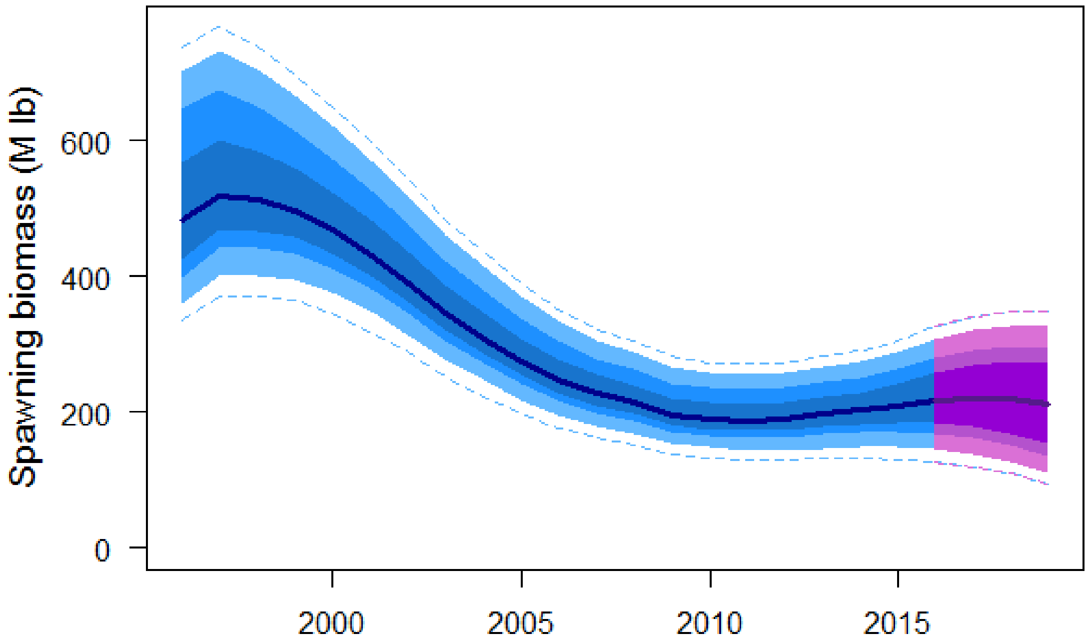


Figure 110. Integrated model results, with three year projection in lighter color. Line represents the median, color bands the interquartile (50/100), 75/100, 95/100, and the dashed lines represent the 99/100 range.

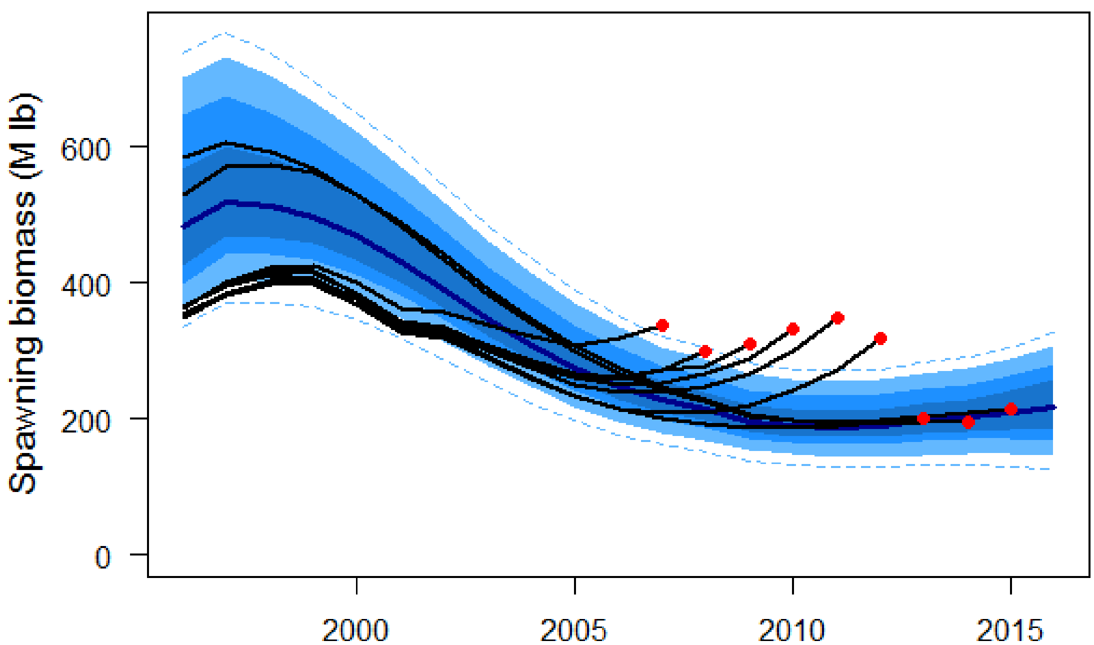


Figure 111. Retrospective comparison of stock assessment results since 2006. Red circles represent the terminal point estimate (MLEs from 2006-11, ensemble medians from 2012) from each historical assessment.

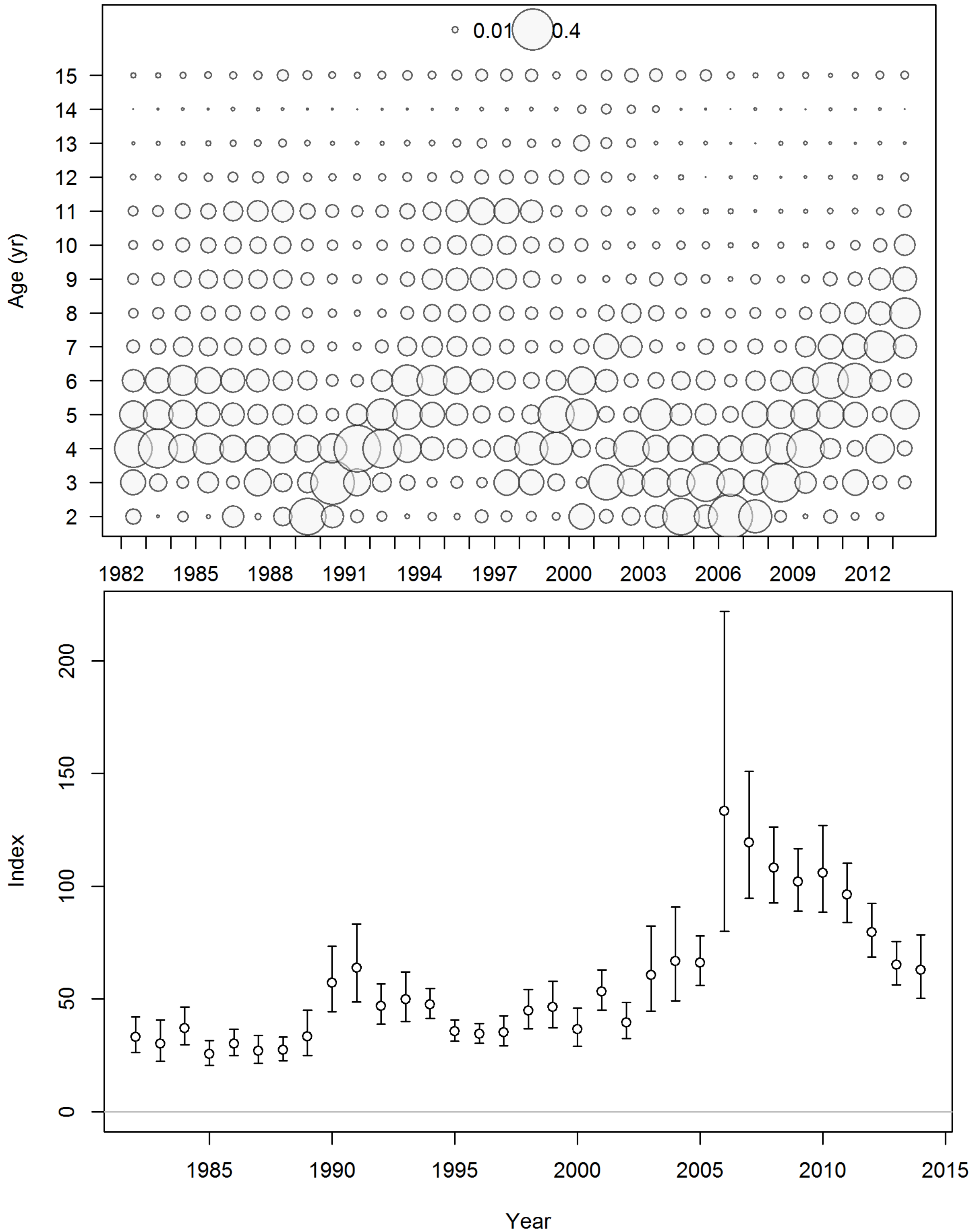


Figure 112. Estimated proportions-at-age (upper panel, years prior to 1998 from age-length key in 1997, 2014 not yet available) and index data (in numbers; lower panel) from recent NMFS Bering Sea trawl surveys.

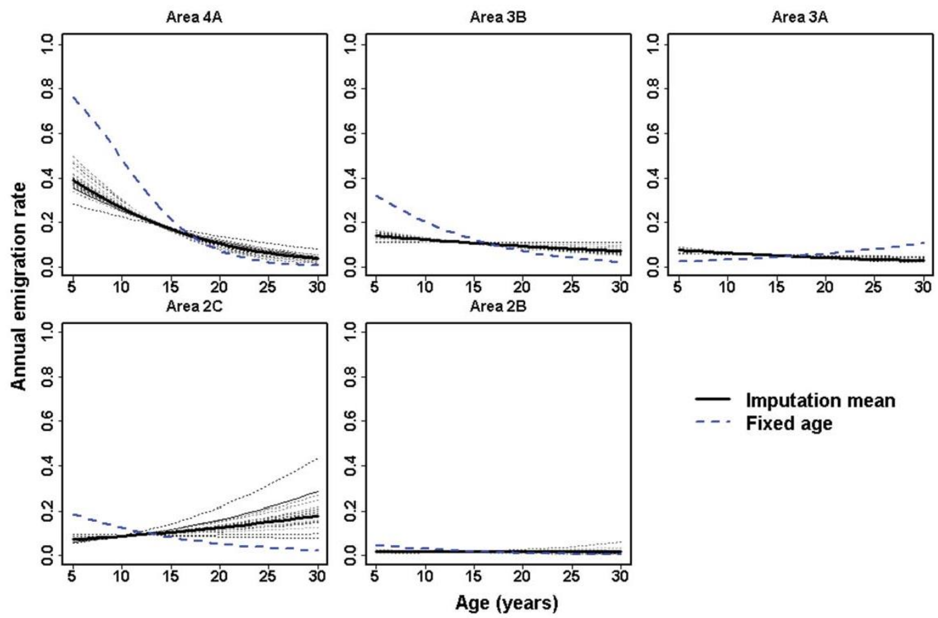


Figure 113. Emigration rates by age based on imputed age (solid line denotes the mean, dotted lines represent example estimates) and fixed age models (dashed line); reproduced from Figure 5 in Webster (2015).

Appendix A: Background material

Included with this document for SRB review is more extensive material supporting each assessment model, as well as several key references. These materials include:

1) Input files for each of the four assessment models (one directory for each of: coastwide short, coastwide long, AAF short, and AAF long): input data file, weight-at-age file, control file with model configuration, starter and forecast files with additional settings. Each of these has been annotated to aid in locating the various sections, as well as identifying which options and features were implemented or irrelevant.

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 the HTML files), and the full report file from each model. The report file has not been annotated beyond the standard output from Stock Synthesis.

3) A summary of the modelling approach implemented in Stock Synthesis and the detailed model equations. This overview is supplemented with detail regarding specific features in the user manual. This is provided to allow for a cross-check of specific options used in each stock assessment model with the descriptions and technical details. *References*: Methot Jr, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142: 86-99. Methot Jr, R.D., and Wetzel, C.R. 2013b. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Appendix A: Technical description of the Stock Synthesis assessment program. Fish. Res. 142: 26 p. Methot Jr, R.D. 2015. User manual for Stock Synthesis. Model version 3.24s. NOAA Fisheries. Seattle, WA. 152 p.

4) The summary of data sources and the stock assessment results from 2014. *References*: Stewart, I.J. 2015. Overview of data sources for the Pacific halibut stock assessment and related analyses, IPHC Report of Assessment and Research Activities 2014. p. 87-160. Stewart, I.J., and Martell, S. 2015. Assessment of the Pacific halibut stock at the end of 2014. IPHC Report of Assessment and Research Activities 2014. p. 161-180.

5) Two recent manuscripts describing the history of the halibut stock assessment and the general rationale for the ensemble approach. *References*: Stewart, I.J., and Martell, S.J.D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. 158: 40-49. Stewart, I.J., and Martell, S.J.D. 2015b. Reconciling stock assessment paradigms to better inform fisheries management. ICES J. Mar. Sci. *Advance online publication*.