Preliminary results of the BSAI Greenland turbot stock assessment

By

Meaghan Bryan and Alberto Rovellini

SSC and Plan Team comments (2024)

Comment/Recommendation	Authors' response
The SSC appreciates the authors' development of a linear approximation approach to support the AFSC longline relative population numbers and recommends using this method in future model development where appropriate.	This has been incorporated in the model this year
The SSC recommends the following efforts focused on the input data for the next stock assessment: ● Include bootstrapped input sample sizes for all compositional data. ● Along with the length-at-age data, consider using the age data from the AFSC trawl survey rather than the lengths to provide more direct information on recruitment. ● As requested in 2022, include the sex-specific length data from the recent longline surveys. ● Investigate and provide an explanation for the apparent binning in the length/age data for larger fish and the lack of small males in the shelf survey composition data.	We have model runs using the bootstrapped ISS for the bottom trawl surveys, which this approach is available for. The approach is not currently available for the other fleets in this assessment A sensitivity run using the ENS shelf BTS age data is presented A sensitivity including the sex-specific length data is presented The lack of small males was due to combining females and males below 21cm. This feature has been removed from the model runs, as it had no impact on the model Exploration of different length bins will be done for the next full assessment
Specific recommendations from the Team for future assessments included: ● Using interpolated AFSC longline survey RPN data for all future models, following established best practices. ● Displaying survey mean length-at-age across all model runs to examine interactions with selectivity time blocks and Francis reweighting. ● Exploring later start years, closer to the 1977 regime shift, for potential insights into recruitment dynamics. ● Likelihood profiling over M and von Bertalanffy parameters to address retrospective bias in survey catchability (Q). ● Developing a Tier 5 REMA model to compare with Tier 3 models, given the data losses.	The interpolated AFSC longline survey RPN have been included in the assessment model Fits to size at age are presented for all model runs A later start year has been explored Likelihood profiles over von Bertalanffy parameters are presented. Profiles over natural mortality were not completed because it is not an estimated parameter. A tier 5 REMA model will be presented and compared to the tier 3 model in November

The SSC recommends that the authors continue to explore the use of VAST for the slope survey as requested in December 2020, consider excluding pre-1977 data, and consider new methods for deriving apportionments given the lack of recent EBS slope data.	The authors will explore the use of VAST for the slope survey for the next full assessment. We have model runs exploring the exclusion of the pre-1977 data. We present a new method for apportionment using the REMA model
The Team emphasized the importance of updating the maturity schedule, reiterating prior comments made in September. The authors' efforts to collect fishery samples this year were commended, but the Team suggested increased collaboration with the Fisheries Monitoring and Analysis (FMA) Division to enhance sample sizes. Greenland turbot is not predominant in the catch and has been a lower priority for biological collections, further complicating sample acquisition.	Meaghan Bryan and Todd ten Brink (REFM/A&G) have a special project with the observer program. Currently, we have a small sample size and are waiting for the remainder of the samples from 2025 to analyze the data.

Introduction

During the 2024 assessment process, the authors brought forward a concern about the model start year, given that the accepted assessment model used to provide management advice estimates a small initial population size and estimates large recruitment peaks in the early 1960s and early 1970s when there are limited size composition data to inform these large recruitments (Figure 1). The recruitment peaks are orders of magnitude larger than any recruitment event evidenced in the size composition data time series after 1980 when these data were more routinely collected. In 2025, our main focus has been to determine whether the available data provides sufficient information for the model to discriminate between two hypotheses with regards to Greenland turbot stock dynamics: a smaller, highly productive stock and a larger, less productive stock.

The following table summarizes the data sources, data types, and years included in the model:

Source	Data	Years
NMFS Groundfish survey:	EBS shelf BTS biomass	1987-2024
	EBS shelf BTS length composition	1987-2024
	EBS shelf BTS mean length at age	1982, 1998-2019, 2021-2023
	EBS slope BTS biomass	2002, 2004, 2008, 2010, 2012, 2016
	EBS slope BTS length composition	Same as above and 1979, 1981-1982, 1985, 1988, 1991
	EBS slope BTS mean length at age	2002, 2004, 2008, 2010, 2012, 2016
AFSC longline survey	Relative population numbers	1996-2023
	Length composition	1979-1987, 1993-1994, 1996-2023
U.S. fisheries	Catch	1960-2024
	Trawl length composition	1978-1991, 1994-1996, 1998-2024
	Fixed Gear length composition	1979-1985, 1993-2021

Length information used in the model included length composition data from all fleets: EBS shelf and slope bottom trawl surveys, AFSC longline survey, trawl fishery, and fixed gear fishery. Length-at-age data from EBS shelf and slope bottom trawl surveys were also used. Age data are available for these two surveys, but to avoid using age information twice, they are not included in the model likelihood. Therefore, recruitment deviates are informed by the length data. Early recruitment deviations are estimated for years 1945-1969, main recruitment deviations are estimated 1970 – 2020, and late deviations are estimated for 2021-2024.

We present several model updates following an iterative process. We first present data updates that led to minor differences in assessment outcomes. We then present model changes to model 16.4c that follow stock assessment best practices. Lastly, we present models with different assumptions about the initial model conditions (e.g., later model start year with average historical catch and estimation of an initial F) to address whether the data and models can sufficiently discriminate between the two hypotheses about initial population size and productivity.

Model Summary

The 2024 accepted model (model 16.4c) was updated to capture some changes in the data and requests from the SSC and Plan Team in 2024. The data updates included: using a new interpolation method for the AFSC longline survey relative population numbers; using pre-2002 EBS slope bottom trawl survey length data which was obtained using an updated query; updating pre-1997 AFSC longline survey length data to use raw length compositions as per recent recommendations; removing mintail compression on length data; and separating males and females under 21 cm in the lengths. These data updates were made that had a minor effect on the assessment results.

Model reweighting and likelihood profiling pointed to a clear conflict between the slope trawl survey and longline survey data. The three surveys sample different components of the Greenland turbot population by virtue of targeting different areas. The EBS shelf survey is a good index of recruitment because small fish are found predominantly on the shelf. The slope and longline surveys, in contrast, sample the adult and spawning population that moves off the shelf and deeper as the fish grow larger. While the slope bottom trawl survey and the AFSC longline survey are both conducted on the continental slope, the only year of overlap in the last 10 years was 2016 (the last year the slope survey was conducted). Therefore, comparing how data from these two surveys inform population status in recent years is difficult. The trade-off between fitting the slope and AFSC longline indices is a long-standing issue that should be further explored in a future assessment.

These changes led to model 25.3 as the most comparable, improved, model compared to those previously accepted for management advice. Data and model updates were conducted following current best practices. However. this model inherited the long-standing concern around large estimates of early recruitment that are an order of magnitude larger than any recruitment estimated where we have more composition data (after 1980). The 1963-1964 recruitment event is likely an artifact of the model estimating large productivity to support the historical catch. A second recruitment peak in 1975 is possibly better informed by length data that start in 1979, but still predates the more consistently sampled lengths. To address this concern, two models starting in 1977 were developed, in response to research priorities identified in the 2022 and 2024 assessments. A crucial difference between the models presented is the hypotheses they represent. More specifically, models 16.4c and 25.3 represent the hypothesis that the Greenland turbot population was relatively small initially and highly productive, whereas models 25.4 and 25.5 represent the hypothesis that the population was initially large and less productive. The year

1977 was chosen because this is the year recognized as the start the Bering Sea regime shift and many BSAI models start in this year. Also given the relative weights placed on the early length composition data and the results from the analysis removing the length data from the assessment model, the model should have information to estimate recruitment in 1977. The two models starting in 1977 carried different assumptions around the productivity of the population. The first model, 25.4, included equilibrium catch as the average removals between 1960-1977 to account for the early catch data and estimate initial fishing mortality. The second model, 25.5, only included catch data starting 1977, disregarding previous removals. The model 25.5 assumes that, even though we know that large removals were supported by this stock in the 1960-1977 period, the stock moved to a new state of reduced productivity after 1977 and should be managed based on this new state moving forward. Diagnostic analyses were conducted on these two models. The two 1977 models were more stable than M25.3. However, they both presented similar retrospective patterns, and the same trade-offs in fitting the survey index data as for model 25.3. Importantly, the data could not clearly discriminate which of these two 1977 models was best.

Summary of model 16.4c (2024 accepted assessment model)

Stock Synthesis software (Methot 1990) has been used to model the eastern Bering Sea component of Greenland turbot since 1994. The software and assessment model configuration have changed over time, as newer versions have become available. This assessment uses Stock Synthesis version 3.30.21.

Total catch estimates used in the model were from 1960 to 2024. Model parameters were estimated by maximizing the log posterior distribution of the predicted observations given the data. The model included two fisheries, those using fixed gear (longline and pots) and those using trawls, and up to three surveys (EBS shelf and EBS slope bottom trawl survey and the AFSC longline survey) covering various years.

The assessment model also uses the Beverton-Holt stock-recruitment curve, with early recruitment deviations beginning in 1945. Early recruitment deviations are estimated for 1945 – 1970, main recruitment deviations, where there are length composition data available to inform them, are estimate for 1971-2020, and late recruitment deviations are estimated for 2021-2024. Age data from the bottom trawl survey fleets are not included in the likelihood to avoid double counting of age information. Unfished recruitment is estimated, steepness is fixed at 0.79 (Myers et al.1999), and sigmaR is fixed at 0.6, and an autocorrelation parameter is estimated to account for evidence of autocorrelated recruitment, although this is no longer considered best practice (Johnson et al. 2016).

Catchability for the EBS shelf and EBS slope surveys were fixed in the last accepted assessment model. Since the 2015 assessment, the catchability values used in the model were $\log(q_{shelf}) = -0.485$ and $\log(q_{slope}) = -0.556$. These values were estimates from the 2015 model run that excluded 2007-2015 data to avoid effects of the large 2007 through 2010 year classes. Selectivity for the AFSC longline survey was modeled using the logistic selectivity pattern, where the length at 50% selectivity and the slope parameter were estimated. Selectivity for the AFSC longline survey is not sex-specific because prior to 2021 sex-specific lengths were not collected. Sex-specific size-based selectivity functions were estimated for the two trawl surveys and the two fisheries and modeled using a double normal pattern.

Time blocks were used to estimate time varying selectivity for the fishery and the shelf and slope bottom trawl surveys. This is a longstanding feature of the Greenland turbot model and in doing so, it is assumed that selectivity changes over time due to changing availability of certain size classes (e.g., when there is a recruitment event). The time blocks used in Model 16.4c are as follows:

Fleet/Survey			
EBS shelf bottom trawl survey	1945 – 1991	1992 – 1995	1996-2000, 2001 - 2022
EBS slope bottom trawl survey	1945 – 2001	2002 – 2010	2011 - 2022
Trawl gear fishery	1945 – 1988	1989 – 2005	2006 - 2022
Fixed gear fishery	1945 – 1990	1991 – 2007	2008 - 2022

Data updates

Several data updates were made that had a minor effect on the assessment results (Figure 4-Figure 5). The cumulative updates include:

d1. Using the AFSC longline survey relative population numbers (RPNs) using an interpolation method that was approved by the Plan Team and NPFMC SSC.

d2. Updated pre-2002 EBS slope bottom trawl survey length data (Figure 2). The authors compared the previously used data with data using a new data query provided by the AFSC RACE Groundfish Assessment Program (GAP).

d3. Updated pre-1997 AFSC longline survey length data. The authors reviewed the data and determined the pre-1997 data were raw length composition data from the cooperative survey with Japan. Any mismatch in the pre-1997 data is due to the use of RPN weighted values in 1987, 1993, 1994 in previous assessments (Figure 3). The authors decided to use the raw length composition data for these years moving forward to use a consistent approach for the pre-1997 data, which has been recommended by ABL/MESA (Chris Lunsford, personal communication).

d4. Removing mintail compression on the length data (this is necessary for OSA residual calculation)

d5. Male and female lengths were combined below 21cm in model 16.4c. The SSC expressed concern that there were no small males in the data; however, this is not the case. The data are sufficient for both sexes and male and female lengths below 21cm are no longer combined in the model.

The results from the model runs d1-d5 are included in the report to track changes, but will not be elaborated on as the fits to the data were similar and the differences in assessment outcomes were negligible.

Model changes following best practices

Several model changes following best practices were explored and implemented. The model changes were made starting with model d5, which represents the cumulative data updates, and include:

Change	Model number
Input sample size from the surveyISS R package for the bottom trawl surveys (Williams and Hulson 2024)	25.1
Fixing the Beverton-Holt stock recruitment relationship's autocorrelation parameter, which is recommended by the Stock Synthesis 3 (SS3) developers (Johnson et al. (2016) and SS3 Manual)	25.2
Using the catchability float option in SS3 rather than using fixed values for the EBS shelf bottom trawl survey and the EBS slope bottom trawl survey	25.3

The changes made in Models 25.1 – 25.3 represent cumulative changes as the model number increases. The models do not include any iterative re-weighting.

As recommended by the Plan Team and the SSC, the surveyISS R package was used to obtain the input sample sizes (ISS) for the EBS shelf and slope bottom trawl surveys used in models 25.1 – 25.3 (Table A.1). The assumed ISS was 100 in model 16.4c; therefore, this approach increases the ISS for the shelf survey in the majority of years, while the ISS for the slope survey (2002 – 2016) was more consistent with what was used in 16.4c (i.e. 400 in all years between 2002 and 2016). The pre-2002 slope survey ISS was unchanged, as these years are not include in the surveyISS bootstrapping routine. The ISS for all other fleets were unchanged as the bootstrapping routine is not available for the fishery data or the longline survey.

It has been recommended that the SRR autocorrelation parameter not be estimated in SS3 models as SS3 poorly estimates the autocorrelation parameter (Johnson et al 2016). It is now recommended to use an external estimate the autocorrelation from the recruitment deviations and fix it in the model, if there is evidence of autocorrelation. This parameter was fixed to 0.437 in 25.2 based on the results for Pleuronectids from a study that developed a hierarchical model to evaluate variability and autocorrelation in a number of fish families (Thorson et al. 2014).

Catchability for the EBS shelf and EBS slope surveys were fixed in the last accepted assessment model. Since the 2015 assessment, the catchability values used in the model were $\log(q_{shelf}) = -0.485$ and $\log(q_{slope}) = -0.556$. The survey-specific catchability values were estimated from the 2015 Model 14.0 fit without the 2007 - 2015 data. This was meant to eliminate the effects of the 2007 through 2010 year classes. During the CIE review in 2021, the CIE reviewers indicated that the "practice of using estimates from an older model run is not recommended, because it uses the first part of the data twice". As an alternative, we used the float option in SS3 where the model derives an analytical solution for the survey-specific catchability.

Results

The fit to the AFSC longline survey RPNs improved when the time series from the interpolation method was included in the model (d5) while the fit to the bottom trawl survey biomass degrades a bit (Table 1, Figure 6). Up weighting the bottom trawl length composition data (25.1) leads to worse fits to the indices, as expected; however, this is improved upon by using the catchability float option in SS3 (25.3) for the surveys. The fit to the longline survey RPNs is improved (lower RMSE) for model 25.3 compared to model 16.4c, whereas, the fit to the bottom trawl surveys is better for model 16.4c than model 25.3.

It is difficult to compare the length likelihoods from models 25.1 – 25.3 to model 16.4c given that the data weighting used in model 16.4c differs from models 25.1 – 25.3. The fits to the bottom shelf survey degrades in years where the ISS is lower than model 16.4c and improved upon in years when the ISS is higher (Figure 7). The fits to the slope survey lengths are similar across models, with some year-specific improvement when ISS was change (Figure 8). The fits to the AFSC longline survey are similar across models (Figure 9). The fits to the trawl fishery are generally similar across models, but there is some improvement when model changes were made (Figure 10). The fits to the fixed gear fishery is similar across models (Figure 11). The greatest changes in selectivity were observed for the shelf and slope bottom trawl surveys, where the change in data weights indicate a narrower domed shelf survey selectivity curve, especially for females, and slightly higher logistic slope selectivity for females (Figure 12, Figure 13). Selectivity for the longline survey, the trawl fishery, and the longline survey were fairly consistent across models, which is not unexpected given that their input sample sizes were unchanged (Figure 14 - Figure 16).

Growth estimation, informed by the mean length-at-age data, was similar across the models (Figure 17).

Likelihood profiles

Likelihood profiles for R0 were conducted between values of 8 and 10 in increments of 0.1 highlighted model instability (Figure 18). There were convergence issues for log(R0) < 8.6, therefore only the right-hand-side of the profile is shown. Log likelihood was lowest for log(r0)=8.6, compared to the model's estimate of 8.7. Likelihood profiling also demonstrates several conflicts between data sources, especially for the R0 parameter (Figure 18 - Figure 22). The length likelihood component dominates the total likelihood when estimating R0 and the length data from the EBS shelf survey is particularly influential as this survey samples smaller fish and is important in informing recruitment in the model (Figure 18a). The EBS shelf survey length likelihood is lowest at a small R0 whereas all other length likelihood components do not have a single low over the profile range (Figure 18b). Additionally, the length at age likelihood conflicts with the length likelihood, where the slope length at age likelihood component is lowest at a larger R0 value (Figure 18c). There is more agreement between the sources of data when profiling over the von Bertalanffy growth parameters (Figure 19-Figure 22).

Francis iterative reweighting

We explored Francis reweighting using model 25.3 given the data conflicts present in the data and model 16.4c included variance adjustment on the length composition data (Table 3). The sample size of the fixed gear fishery remained unchanged, while the adjusted sample size was greatly reduced for the shelf and slope surveys for model 25.3. The resulting mean sample size is as follows:

	Mean sample size after variance adjustment (var adjustment)	
Fleet	16.4c	25.3 Francis reweighting
Trawl fishery	12.5 (0.25)	10.8 (0.22)
Fixed gear fishery	25 (0.5)	50 (1)
EBS shelf BTS	59 (0.25)	5.4 (0.02)
EBS slope BTS	122 (0.5)	10.4 (0.04)
AFSC longline survey	15 (0.5)	10.3 (0.17)

Iterative re-weighting leads to an improved fit to the longline survey RPNs and little improvement in the fits to the shelf and slope bottom trawl survey biomass even though the length composition from the bottom trawl surveys was significantly down weighted (Table A.1, Figure A.1). The fits to the length composition data remain relatively unchanged (Figure A.2 - Figure A.6). The intention of iterative reweighting of composition data is to improve the fits to survey indices used to scale the population, which this does. The extreme down weighting of two fishery-independent sources of data is concerning as these are considered highly reliable sources of information about recruitment and the adult population. Given the extreme down weighting, **we recommend that iterative re-weighting be revisited during the next full assessment when approaches to obtain better input sample sizes for the fishery and longline survey can be re-evaluated.**

Sensitivity runs using model 25.3

We completed two sensitivity runs to address SSC comments. Using model 25.3 without data reweighting, we ran the following models:

Model change
Include the sex-specific longline survey length composition data (2021 – 2023)
Include EBS shelf BTS age composition data

The inclusion of the sex-specific longline survey length composition data had little impact on the model (Appendix B). Three years of sex-specific length data are available. The data were included as part of the longline survey and selectivity was assumed to be the same for females and males. When the next full assessment is done for this stock we will have another data point (i.e., a total of four if the longline survey is completed in 2026) **and we recommend including a time block on selectivity, where females and males are assumed to have the same selectivity prior to 2021 and estimate sex-specific selectivity for 2021 forward.**

The inclusion of the EBS shelf bottom trawl survey age composition data also had little impact on the model. The intention of including the shelf survey age composition data was to improve the estimation of recruitment. The first year of age data available from the shelf survey is 1998 and earliest age composition data is mainly young fish (<5 years old) and this particular data set generally sees fish < 15 years old (Figure B. 8). Although these data could inform recruitment from the 1970s and 1980s, the greatest uncertainty about recruitment is the 1960s. Therefore, we recommend including these data in the model as ghost data, as has been done in previous years.

The cumulative changes represented in model 25.3 result in similar fits to the data and assessment outcomes and the changes represent decisions based on stock assessment best practices. **As such, we recommend bringing forward model 25.3 in November.** We also use this model to explore changes in when to initiate the population in the model.

Later model start date

Estimation of early recruitment deviations, particularly those during 1963 and 1964 and 1974-1976 had been identified as a major source of uncertainty in this model because two years of recruitment deviations during this period are an order of magnitude larger than any subsequent recruitment deviation that is informed by the size composition data. One approach to exploring the impact of the large recruitment deviations estimated in 1963-1964 and 1974-1976 is to start the model later, when there are more data to inform the population dynamics. The tradeoff to starting the model later is the need to account for large catches prior to the model start data by estimating and initial F and providing the model with an average catch removed from the population prior to the model start year. In 2024 the Plan Team recommended exploring a more recent model start year, later than 1945 as used in the current model for management. Past assessments used 1945 as the model start date because catch data, available starting 1960, informed large removals between 1960 and 1980. One result of this choice was that the model estimated a relatively small unfished population in the start year (1945), and a few large recruitment events in the 1960s and 1970s to support the observed catch. One concern with this model specification is the fact that the earliest available length data, which constitute the primary source of information for recruitment, are from the late 1970s. Length data are only informative about previous recruitment given variation in size at age. Given where the growth curves have their asymptote for this species (Figure 4), length data from the late 1970s may not be informative of recruitment events in the 1960s and early 1970s. In addition, early fishery length data is available as raw lengths while lengths are standardized in later data (starting in 1997 for the longline survey and 2002 for the slope bottom trawl survey), and early survey length data is from US-Japan cooperative surveys where differences in survey design prevents us from standardizing the composition data using the design based estimators. These limitations are a source of concern that the early length data may less reliably inform the 1960s and 1970s recruitment estimates that are larger than all subsequent recruitments estimated with more data.

To explore these concerns, we performed a sensitivity analysis on model 25.3 to evaluate the effects of removing the early length data from the model one fleet at a time. The purpose of this analysis was to evaluate the effects of the early length data on the 1960s and 1970s recruitment events. Removing early length data required removing the first selectivity blocks for the fishery fleets.

We developed the following models:

a1: removed 1978-1988 from trawl fishery length data; remove first block.

a2: remove 1979-1990 from longline fishery length data; remove first block.

a3: remove 1979-1991 from EBS slope survey length data.

a4: remove 1979-1994 from AFSC longline survey length data.

a5: a1 + a2.

a6: a3 + a4.

a7: a1 + a2 + a3 + a4.

Removing early length data had limited effects on the estimation of the largest recruitment event in the model - the 1960s peak - with the exception of the model where length data were removed from all fleets simultaneously (Figure 23). Length data removal had significant effects on stock SSB and depletion, in part because of the different recruitment estimates for the peak in the 1970s, but also because of differences in estimates of selectivity deriving from the various block structures.

This analysis suggests that the 1960s recruitment peak, which is the largest in the model, is likely being estimated to support the large removals informed by the catch data, rather than being informed by the length data. Length data appeared more influential in estimating the 1975 recruitment peak. These findings support starting the model at a later date than 1945. The year 1977 was chosen mainly because this is the year recognized as the start the Bering Sea regime shift and many BSAI models start in this year. Also given the relative weights placed on the early length composition data and the results from the analysis removing the length data from the assessment model, the model should have information to estimate recruitment in 1977. Model 25.3 (and previous assessments) predicts positive recruitment deviations in 1977, highlighting a period of favorable conditions but after the 1975 peak and near the first year of length composition data (1979).

As such, the following models were explored:

Key difference in model assumptions	Model number
Model 25.3, where start year = 1977, equilibrium catch = average of the 1960 – 1977 observed catch, initial fishing mortality estimated, main recruitment deviation estimation = 1977 – 2020, mean recruitment = 2020-onward	25.4
Model 25.3, where start year = 1977, catch starts in 1977 as observed, main recruitment deviation estimation = 1977 – 2020, mean recruitment = 2020-onward	25.5

Models 25.4 and 25.5 reflect different assumptions in the conditions of the stock in 1977. Model 25.4 captures the large removals from the 1960-1977 period, which are informative of the size of this stock during that period. Model 25.5 assumes that the productivity of this stock changed in the mid-1970s, possibly due to a larger regime shift in the Bering Sea in this period (Wooster and Zhang 2004), and does not assume that the stock will return to the productivity that supported the 1960-1977 catches.

Fits to the data were similar across models 25.3 – 25.5 (Table 1, Table 2, Figure 24 - Figure 29). These models had a better fit to the longline survey RPNs than model 16.4c, whereas model 16.4c better fit the bottom trawl surveys. Fits to the length composition data were also similar, with some subtle improvements in comparison to model 16.4c. Estimates of selectivity and growth were also similar across models (Figure 30 - Figure 34).

A crucial difference between the models is the hypotheses they represent. More specifically, models 16.4c and 25.3 represent the hypothesis that the Greenland turbot population was relatively small initially and highly productive, whereas models 25.4 and 25.5 represent the hypothesis that the population was initially large and less productive (Figure 37). Models 16.4c and 25.3 estimate a peak in recruitment in the 1960s that is orders of magnitude greater than has been observed during the period with which we have data. When starting the model in 1945 this period of recruitment is mainly informed by the catch and less so from the available length data. Meanwhile when starting the model in 1977, the recruitment estimated in 1977 is higher than that estimated by 16.4c and 25.3; however the estimates are within each other's confidence limits and our analysis here shows this recruitment is better informed by the length data (Figure 23). Ultimately, this shows that there is considerable uncertainty about the initial population size for this stock.

Likelihood Profiles

Likeihood profiling was conducted on R0 using values between 10 and 12 in increments of 0.1 for model 25.5. This model appeared less prone to encountering convergence issues than model 25.3 when exposed to changes in unfished recruitment. The minimum of the log likelihood was located at log(R0)=11.5, which coincides with the MLE estimated by the model. Contribution to total model likelihood and to each likelihood component by fleet followed the same patterns outlined for model 25.3 (Figure 36).

Retrospective Analysis

A retrospective analysis was completed for models 25.3, 25.4, and 25.5 and was conducted in SS3 by removing data systematically by year for 10 years. All models have a positive pattern indicating that with each new year of data the estimates of SSB are less optimistic than the previous assessment. Model 16.4c had the smallest retrospective pattern in SSB of all the models. Mohn's rho was 0.14 for Model 16.4c, whereas the alternative models had Mohn's rho values of > 0.2 (Table 4). This larger retrospective pattern was largely driven by the solution for the EBS shelf bottom trawl survey catchability that declined with each peel effectively increasing the estimate of SSB (Figure 38 - Figure 40).

Jitter Analysis

A jitter analysis was conducted using a step of 0.02. A total of 100 jitters were carried out for each model. The number of converged runs per model is as follows:

Model	Runs converged (out of 100)
25.3	73
25.4	100
25.5	100

Model 25.3 showed signs of instability. Parameters associated with the longline fishery selectivity appeared the main source of instability in the jitter analysis for model 25.3. Of the converged runs, 44 had a lower NLL than the original model, with the lowest NLL in the set being 5964.29, compared to the value of 5967.78 from the original model (a difference of 3.49 units). To explore whether a new minimum in the likelihood surface had been identified by the jitter analysis, we attempted to further explore the model with the lowest NLL in the set. However, the Hessian for that model could not be inverted, highlighting the general instability of model 25.3.

Model 25.4 appeared to be more stable than model 25.3, and all 100 runs converged. There were 68 runs in the set with a lower NLL than the original model. Model 25.5 also appeared to be more stable than model m0c, and all 100 runs converged.

Biologically informed Recommended Distributions

In previous years, Biologically-informed Recommended Distributions (BRD) between the Bering Sea and the Aleutian Islands were determined assuming that the spawning stock includes adults on the continental slope of both areas. BRD in 2024 was based on Greenland turbot biomass from bottom trawl data from the EBS slope and the AI from three overlapping years (2010, 2012, 2016). That methodology led to an estimated 15.7% of the stock's biomass in the AI region.

We developed a REMA model (Sullivan et al. 2022) to determine the BRD using the bottom trawl data as in previous years, but also an additional relative index of abundance for each area. The REMA model is stratified by management area (EBS and AI). The biomass index data used were EBS slope bottom trawl survey and AI bottom trawl survey; the relative abundance index data were the EBS longline survey and the AI longline survey, for which we used relative population weights. The model estimated one process error (PE) parameter because we assumed this to be the same population across the two areas; the model estimated unique scaling parameters q for each strata. The model converged and model fits to the indices suggested that the REMA model appropriately captured the population's process error (Figure 41). The model estimated an average of 11.2% of the adult stock's biomass to be in the AI over the last 10 years. It should be noted that the only year of overlap between the two bottom trawl surveys in the last 10 years was 2016; therefore, apportionment in recent years is largely driven by the relative abundance index (longline data).

References

Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., and Punt, A.E. 2016. Can autocorrelated recruitment be estimated using integrated assessment modelsand how does it affect population forecasts? Fisheries Research 183: 222–232. doi:10.1016/j.fishres.2016.06.004.

Methot, R. D. and C.R. Wetzel. 2013 Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research. 142:86-99.

Myers, R.A., K. G. Bowen, and N.J. Barrowman. 1999. Maximum reproductive rate of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56: 2404–2419

Sullivan, J., C. Monnahan, P. Hulson, J. Ianelli, J. Thorson, and A. Havron. 2022a. REMA: a consensus version of the random effects model for ABC apportionment and Tier 4/5 assessments. Plan Team Report, Joint Groundfish Plan Teams, North Pacific Fishery Management Council. 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Thorson, J. T., O. P. Jensen, and E. F. Zipkin. 2014. How variable is recruitment for exploited marine fishers? A hierarchical model for testing life history theory. Can. J. Fish. Aquat. Sci. 71:973-983.

Williams B, Hulson P (2024). *surveyISS: Survey composition input sample size*. R package version 1.0.0, https://benwilliams-noaa.github.io/surveyISS/, https://github.com/BenWilliams-NOAA/surveyISS/.

Wooster, W.S. and Zhang, C.I., 2004. Regime shifts in the North Pacific: early indications of the 1976–1977 event. Progress in Oceanography, 60(2-4), pp.183-200.

Tables

Table 1. Index RMSE values for model 16.4c, d5, and models 25.1 - 25.5.

Model	Fleet		
	AFSC_LONGLINE	SHELF	SLOPE
16.4c	0.437	0.306	0.188
d5_all data updates	0.290	0.303	0.194
25.1_afscISS	0.427	0.365	0.242
25.2_SRRauto	0.430	0.365	0.244
25.3_qfloat3.21	0.367	0.338	0.228
25.4_syr1977	0.383	0.341	0.224
25.5_rmearlycatch	0.375	0.342	0.225

Table 2. Likelihood values for model 16.4c, d5, and models 25.1 - 25.5. Shading indicated models that use the same data weighting.

		Fleet					
Model	Likelihood	ALL	FshTrawl	FshLL	SHELF	SLOPE	AFSC_LONGLINE
16.4c	Length	1282.9	161.9	106.5	503.8	290.5	220.1
d5_dataupdate		1294.1	161.4	79.3	548.6	290.1	214.7
25.1_afscISS		3360.8	617.2	161.0	1582.0	583.0	417.8
25.2_SRRauto		3356.5	615.5	160.9	1578.8	583.9	417.5
25.3_qfloat3.21		3399.4	622.4	167.2	1623.2	579.6	407.1
25.4_syr1977		3506.3	655.8	175.2	1644.2	591.7	439.3
25.5_rmearlycatch		3479.4	664.8	166.2	1637.6	591.6	419.3
16.4c	Length - at - age	2314.6	0	0	1749.1	565.5	0
d5_datupdate		2313.6	0	0	1749.7	563.9	0
25.1_afscISS		2464.5	0	0	1824.4	640.1	0
25.2_SRRauto		2463.9	0	0	1823.7	640.3	0
25.3_qfloat3.21		2468.2	0	0	1820.6	647.6	0
25.4_syr1977		2461.6	0	0	1814.1	647.5	0
25.5_rmearlycatch		2468.4	0	0	1821.0	647.4	0
16.4c	Survey	8.8	0	0	-7.9	-6.0	22.7
d5_dataupdate		-22.9	0	0	-7.7	-5.6	-9.6
25.1_afscISS		53.9	0	0	14.4	-0.1	39.6
25.2_SRRauto		55.4	0	0	14.1	0.2	41.1
25.3_qfloat3.21		9.2	0	0	-2.4	-1.7	13.3
25.4_syr1977		16.3	0	0	0.1	-2.0	18.1
25.5_rmearlycatch		13.4	0	0	-0.2	-1.9	15.6

Table 3. Input sample size and variance adjustment used in 16.4c and the variance adjustment from Francis reweighting applied to model 25.3.

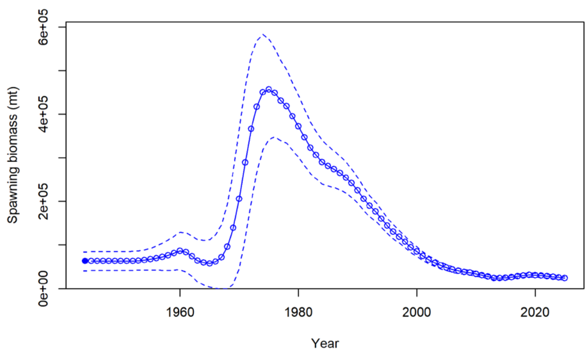
	16.4c		25.3
Fleet	Sample size	Variance adjustment	Variance adjustment
Trawl fishery	50	0.25	0.22
Longline fishery	50	0.5	1
EBS shelf survey	200	0.25	0.02
EBS slope survey	400	0.5	0.04
AFSC longline survey	60	0.5	0.17

Table 4. Summary of Mohn's rho retrospective analysis for models 16.4c and 25.3-25.5.

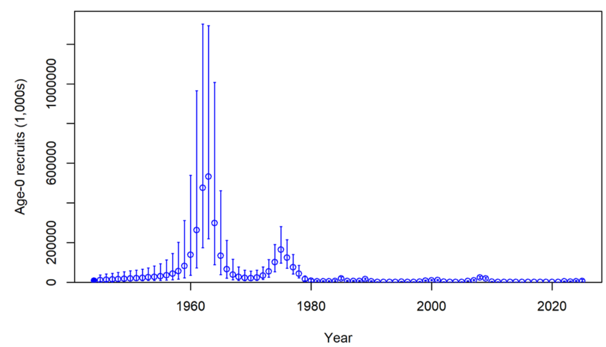
model	AFSC_Hurtado_SSB	AFSC_Hurtado_Rec	AFSC_Hurtado_F	AFSC_Hurtado_Bratio
16.4c	0.142	11.489	-0.200	-0.299
25.3	0.259	3.891	-0.223	-0.102
25.4	0.377	7.147	-0.276	0.356
25.5	0.345	5.980	-0.264	0.324

Figures

a)



b)



c)

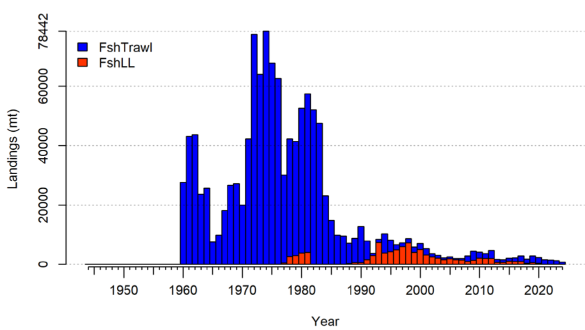
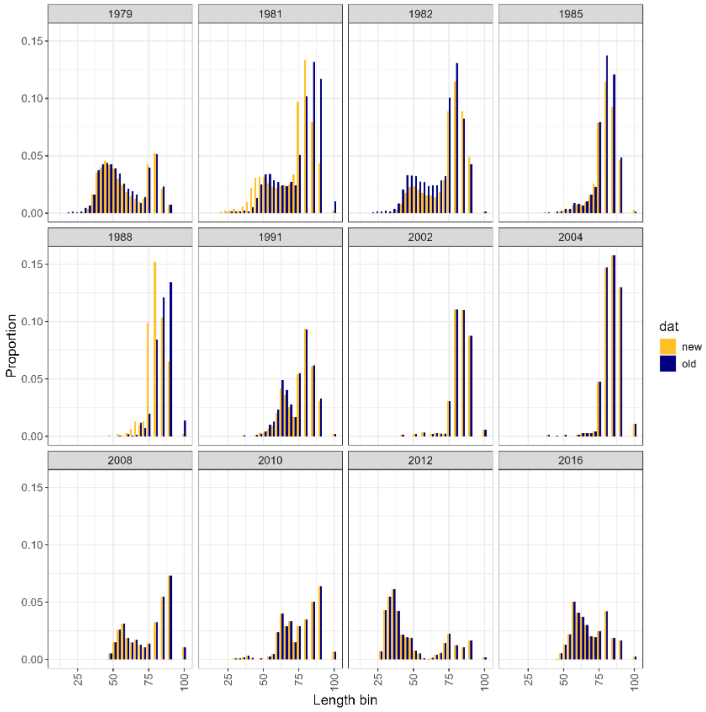


Figure 1. a) Spawning stock biomass and b) recruitment from model 16.4c from the 2024 assessment model and c) commercial fishery catch.

a)



b)

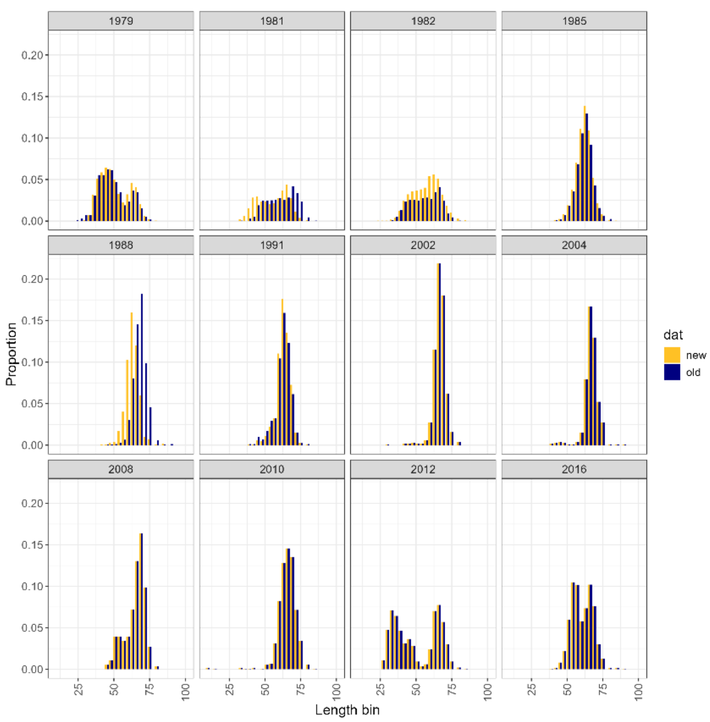


Figure 2. AFSC longline survey length composition data update. 1988-1992 have not been included in the model.

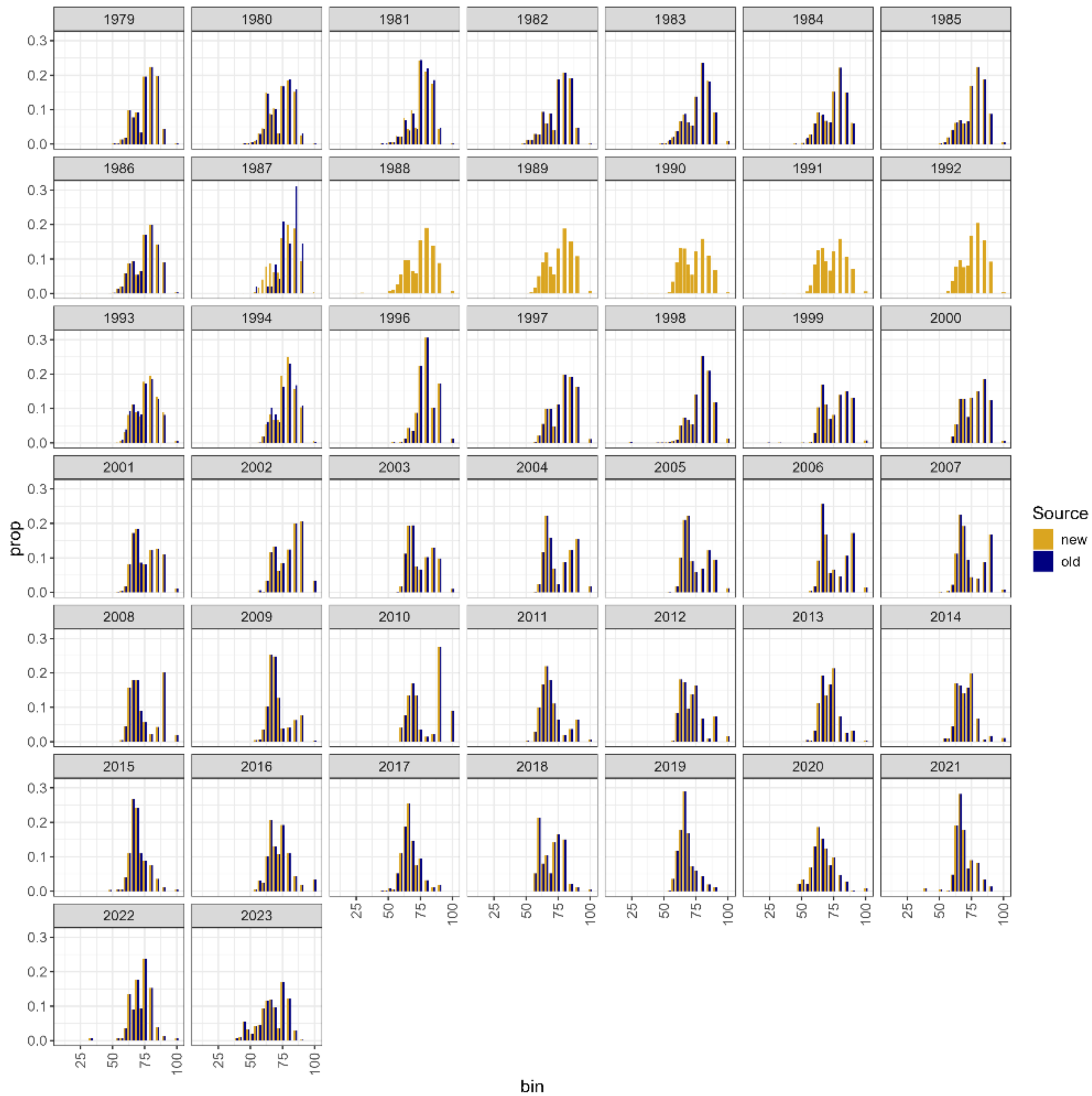


Figure 3. AFSC longline survey length composition data update. 1988-1992 have not been included in the model.

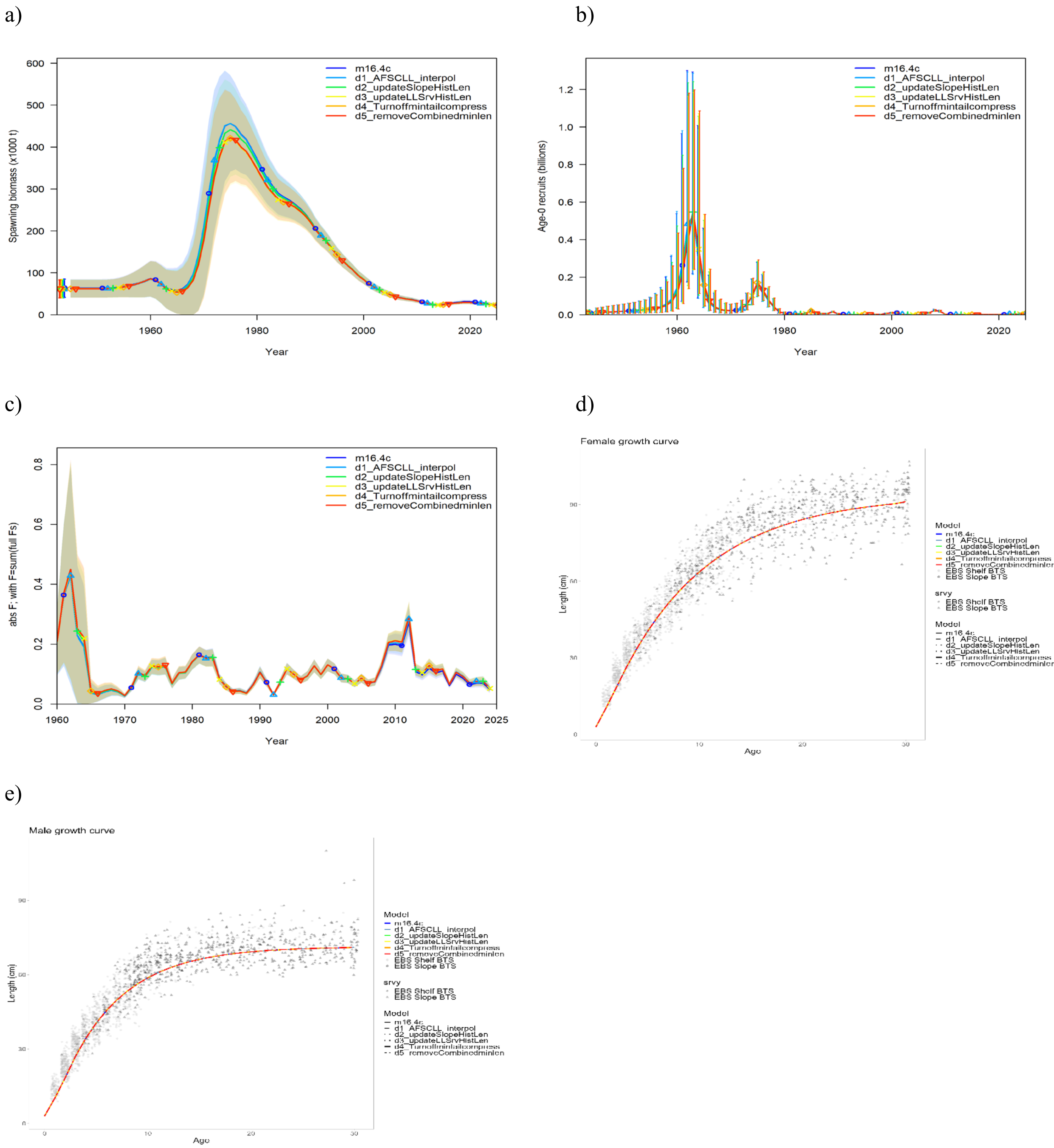


Figure 4. a) Spawning stock biomass, b) recruitment, c) fishing mortality, d) female growth curve, and e) male growth curve.

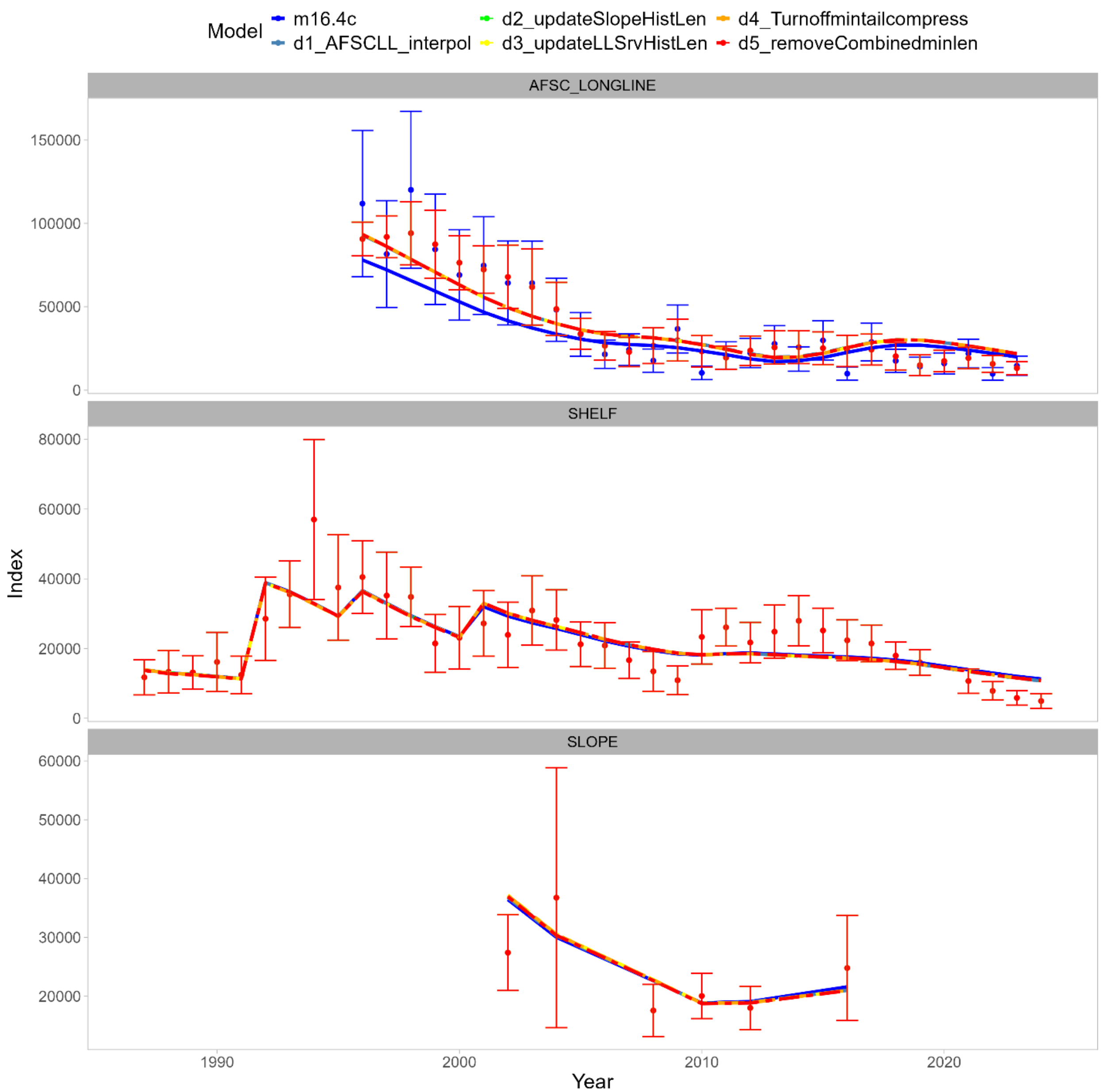


Figure 5. Fits to the AFSC longline survey RPNs, EBS shelf survey (SHELF), and EBS slope survey (SLOPE). The data in d1-d4 differ from model 16.4c and use the RPNs from the interpolation method.

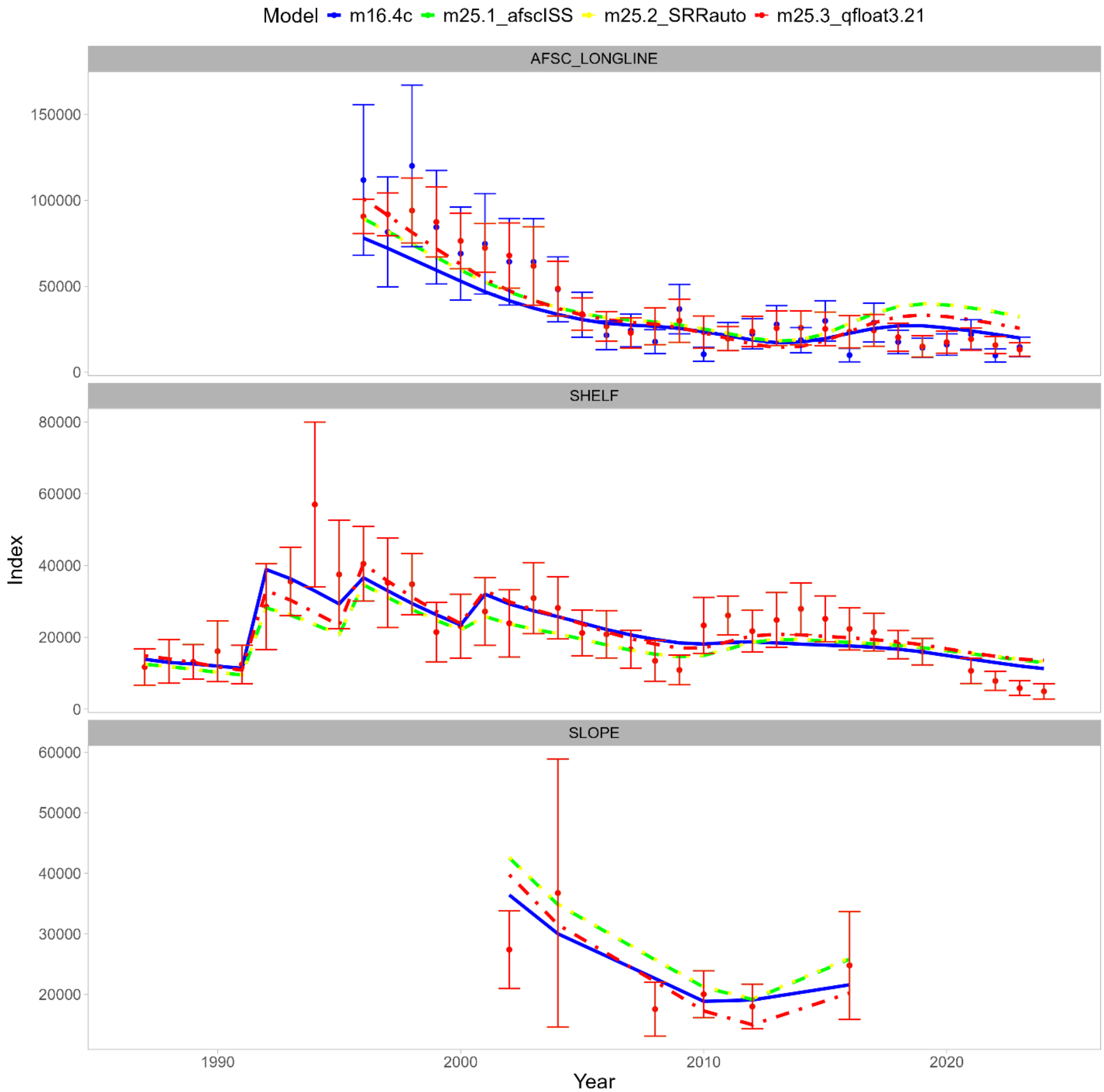


Figure 6. Fits to the longline survey RPNs and bottom trawl survey biomass for models 16.4c and 25.1 – 25.3.

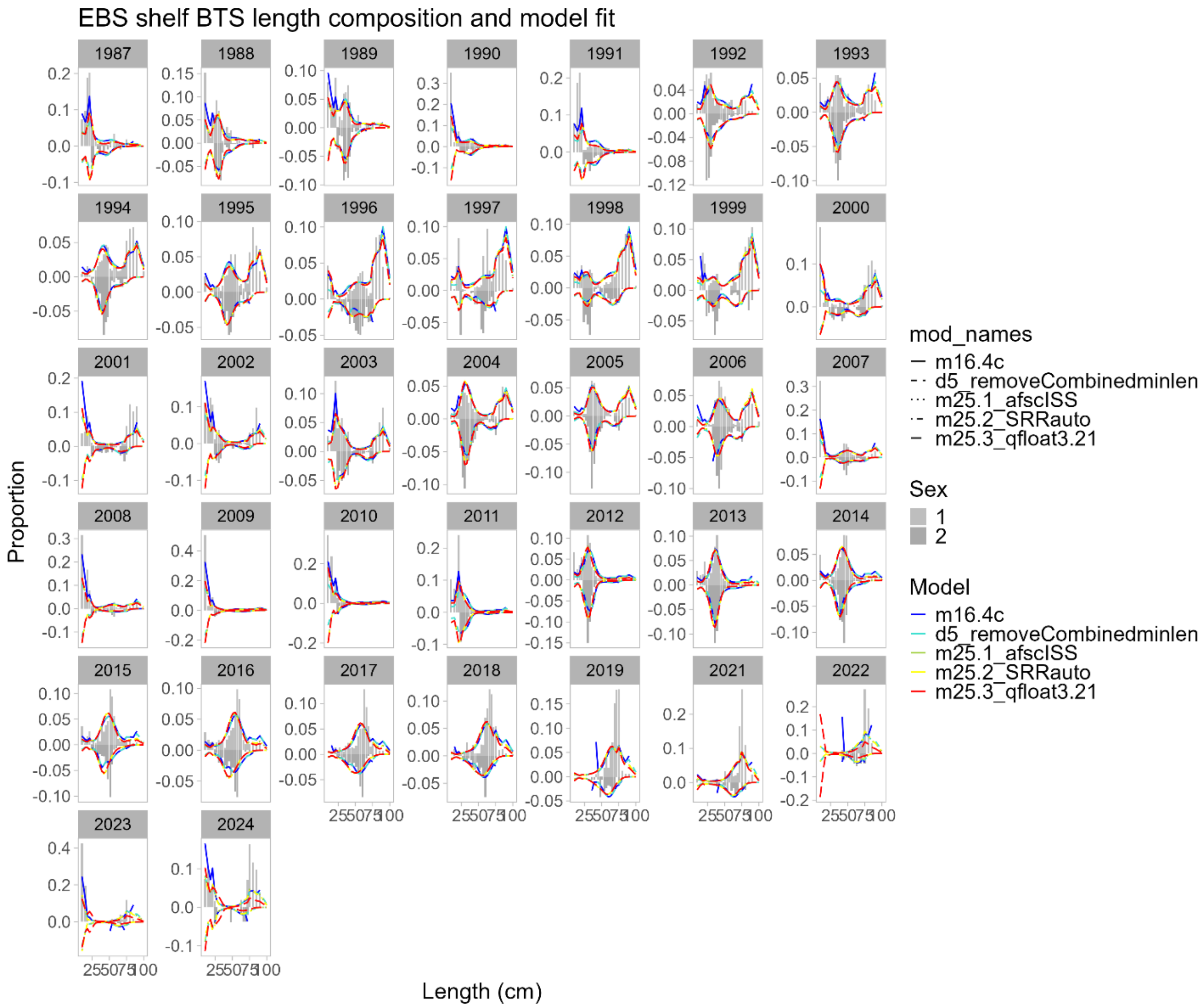


Figure 7. Fits to the EBS shelf bottom trawl survey length composition data for models 16.4c and 25.1 – 25.3.

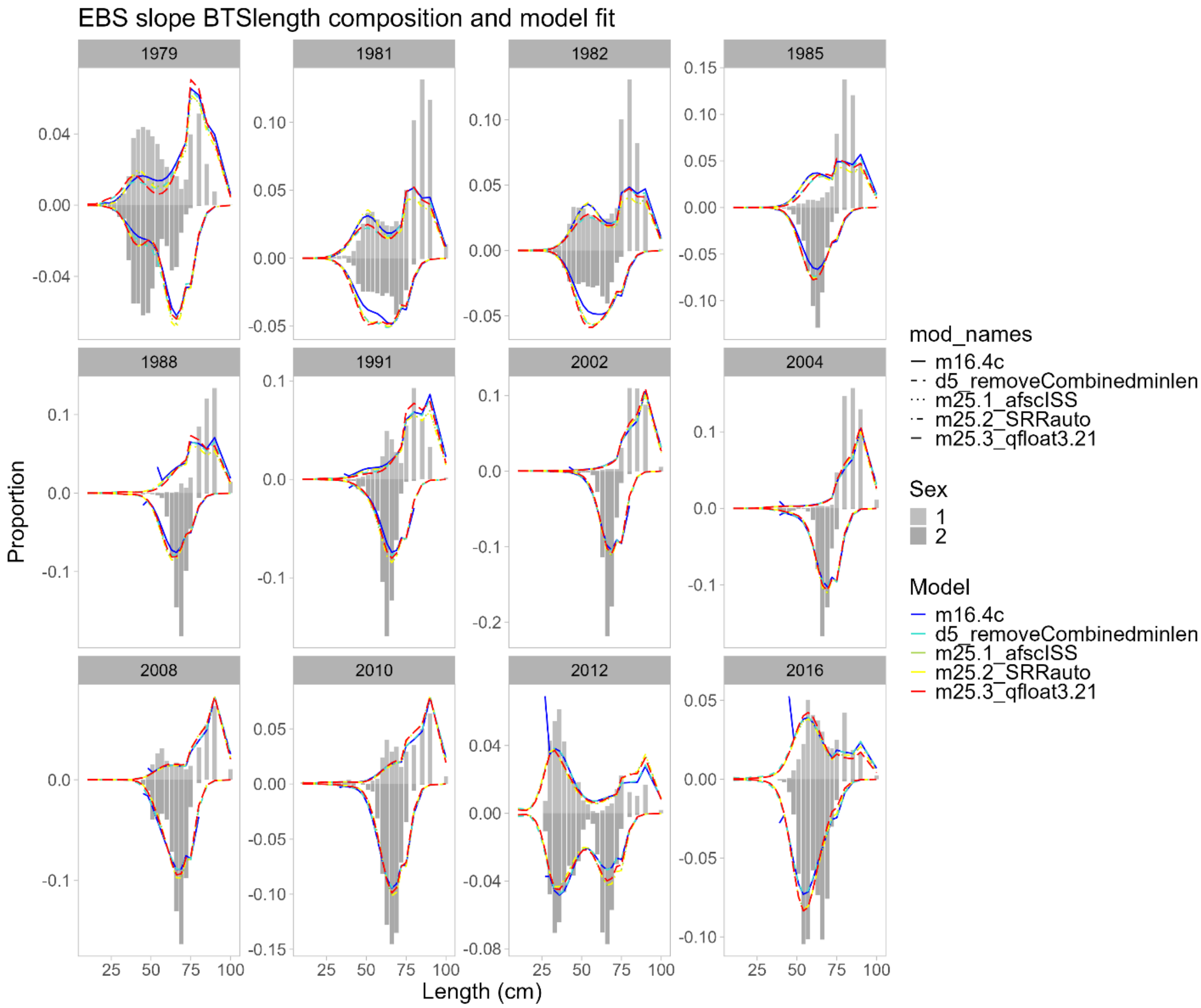


Figure 8. Fits to the EBS slope bottom trawl survey length composition data for models 16.4c and 25.1 – 25.3.

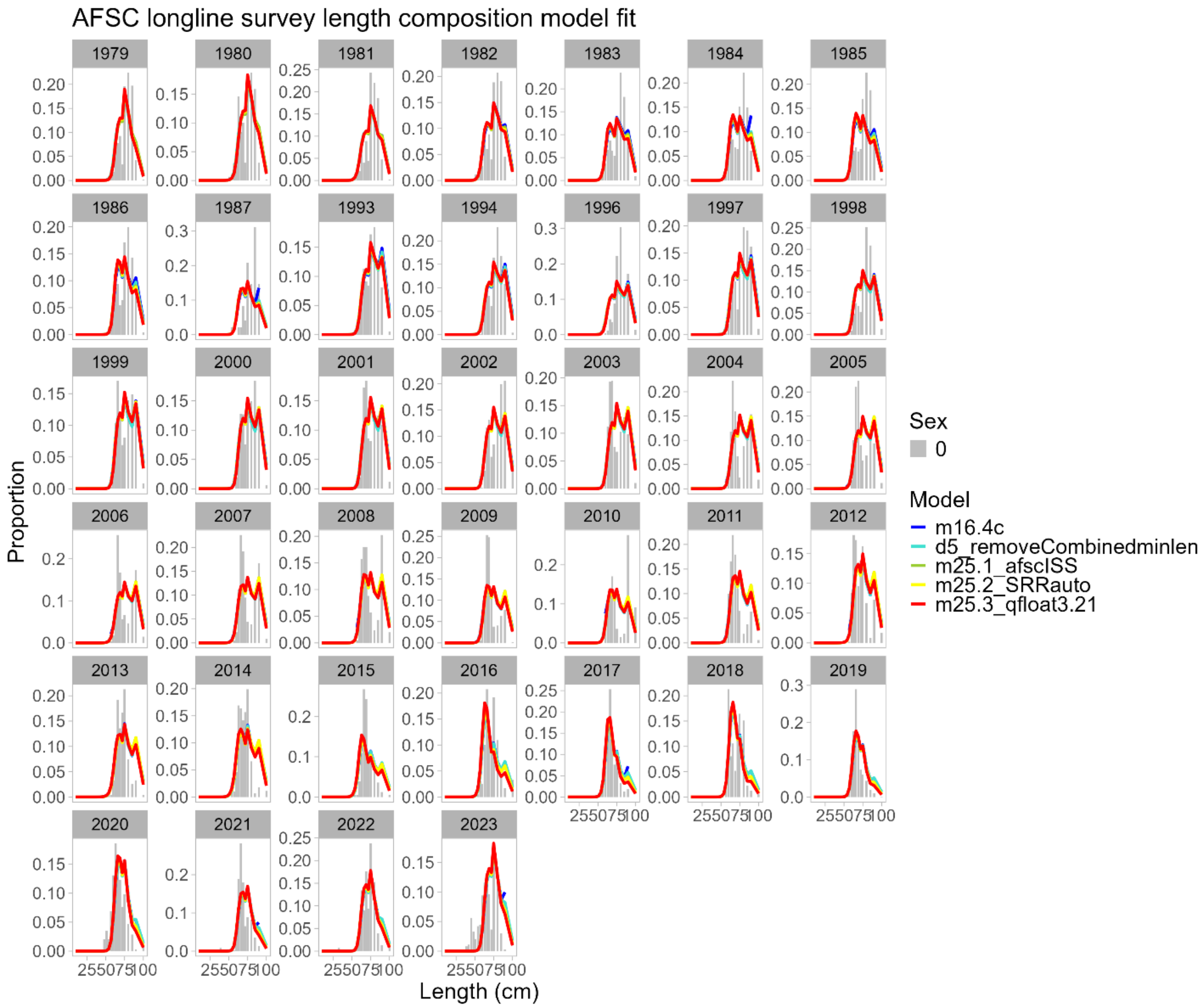


Figure 9. Fits to the AFSC longline survey length composition data for models 16.4c and 25.1 – 25.3.

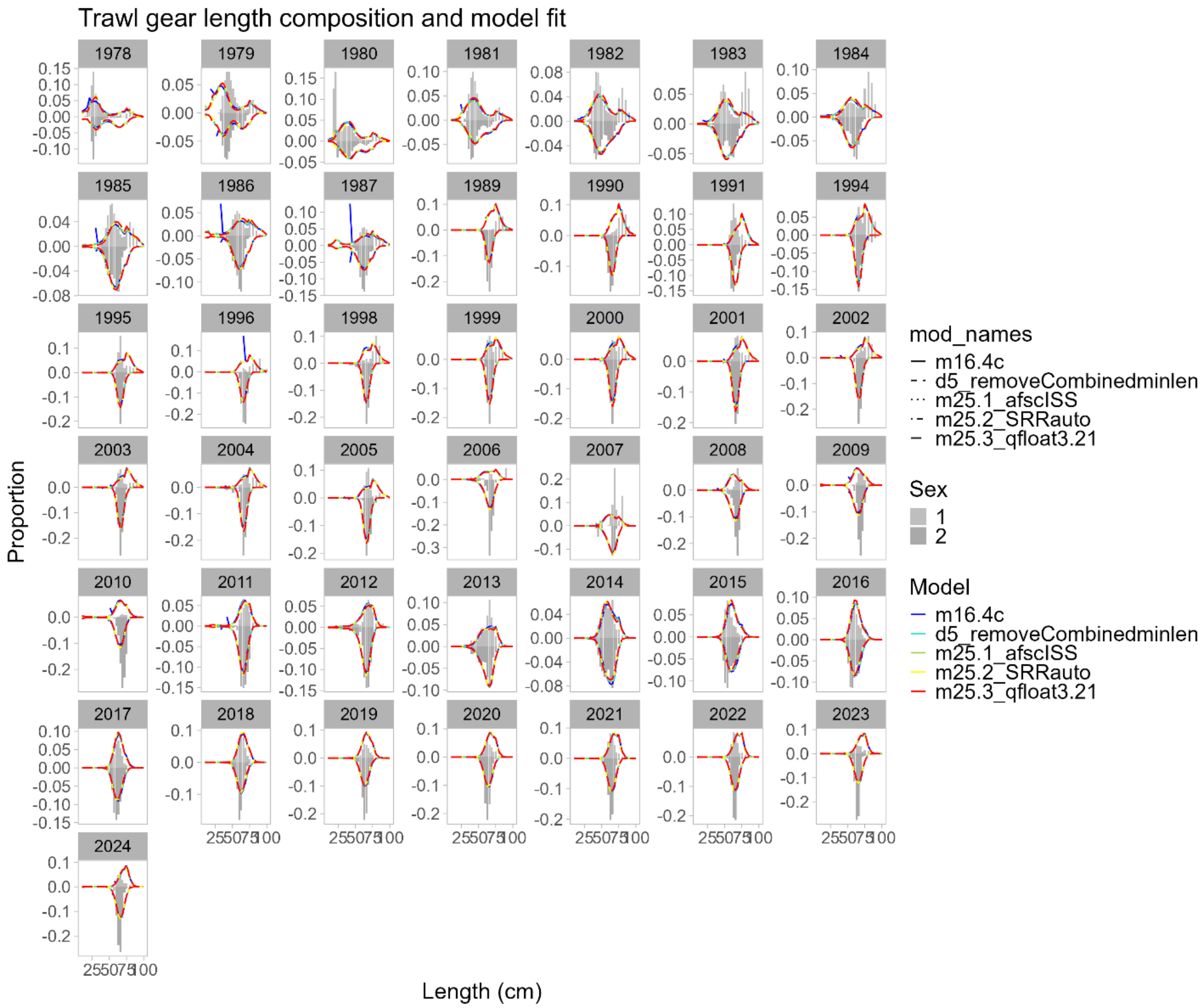


Figure 10. Fits to the trawl fishery length composition data for models 16.4c and 25.1 – 25.3.

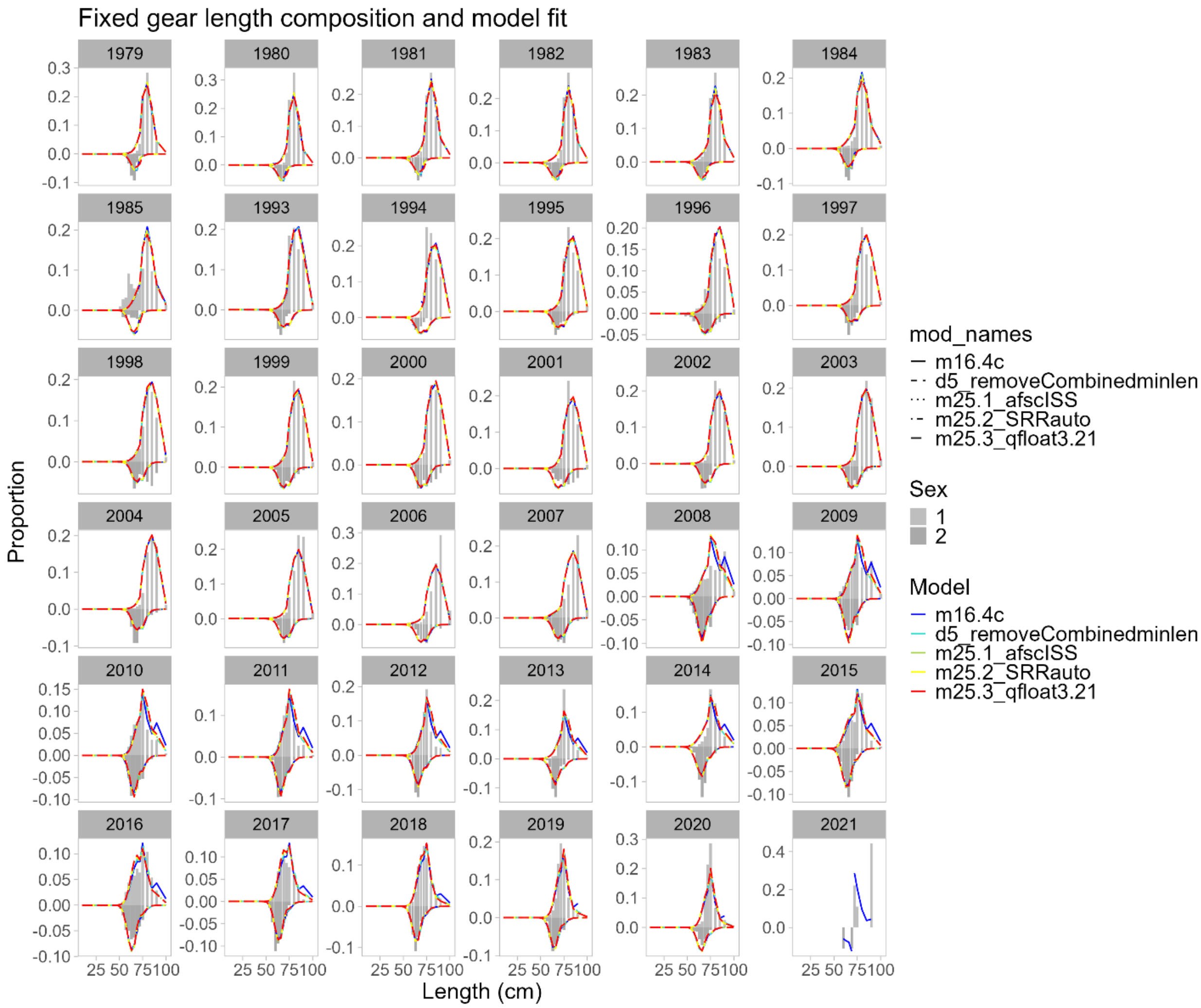


Figure 11. Fits to the fixed gear fishery length composition data for models 16.4c and 25.1 – 25.3.

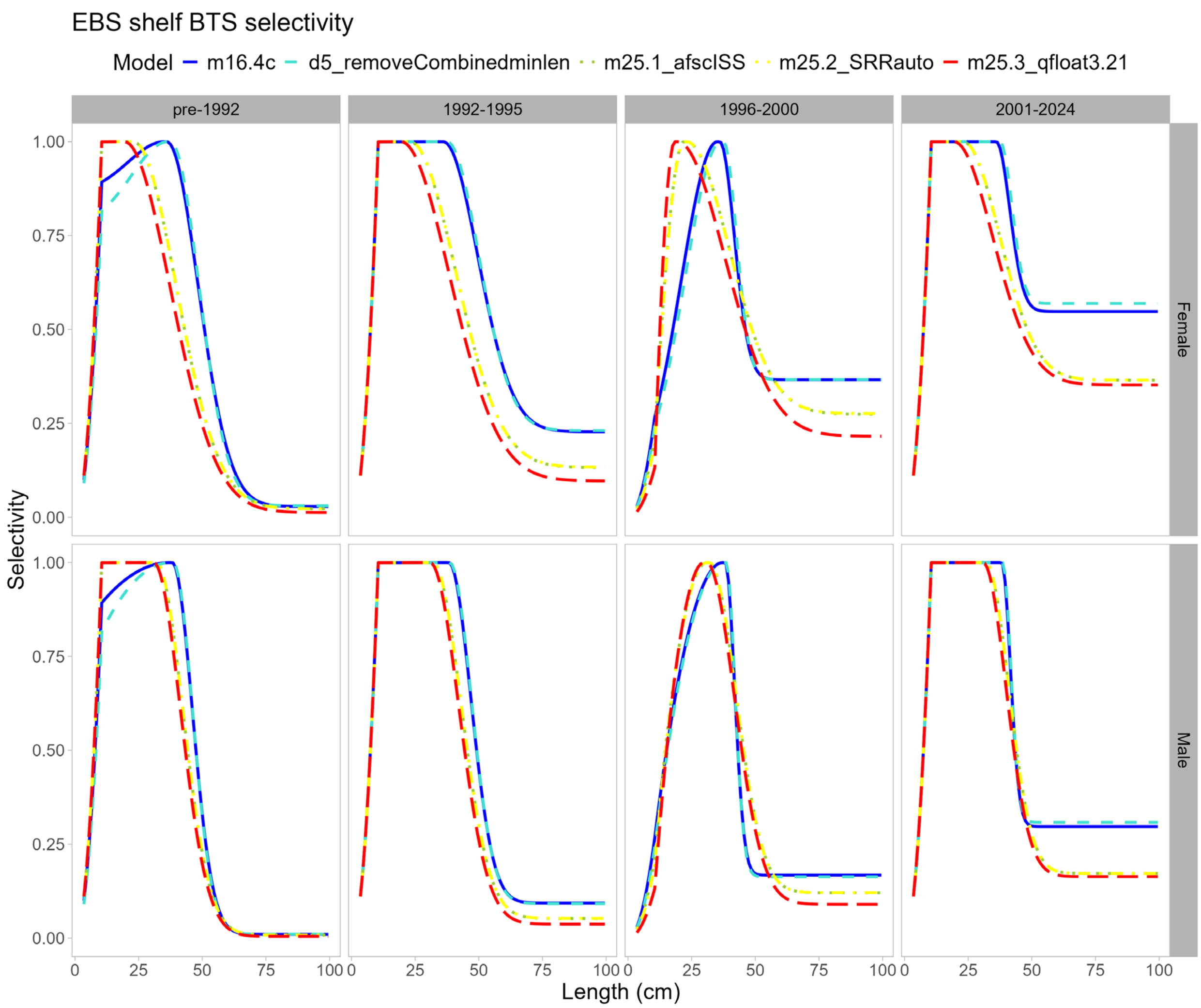


Figure 12. EBS shelf selectivity for models 16.4c, d5, and 25.1-25.3.

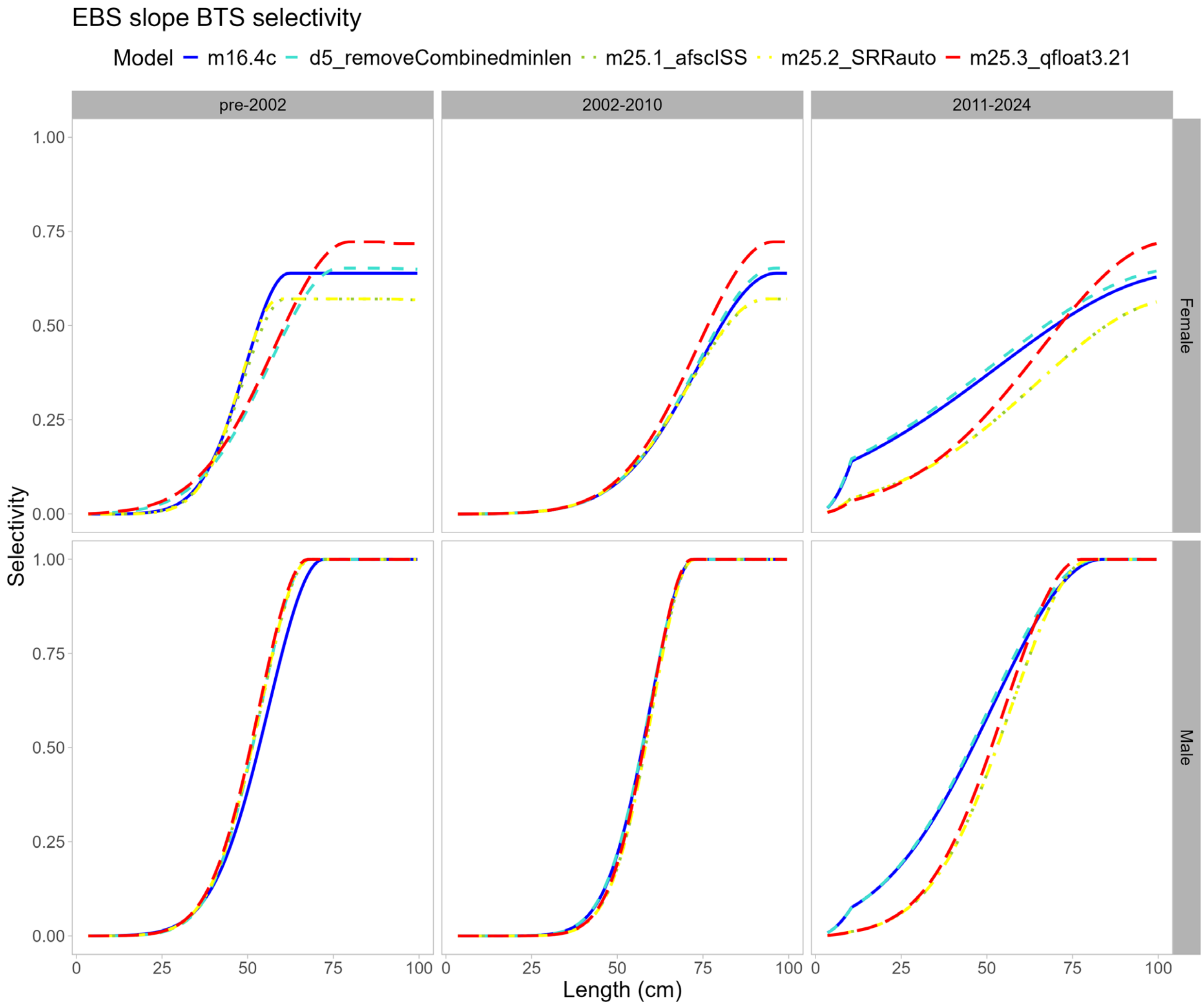


Figure 13. EBS slope selectivity for models 16.4c, d5, and 25.1-25.3.

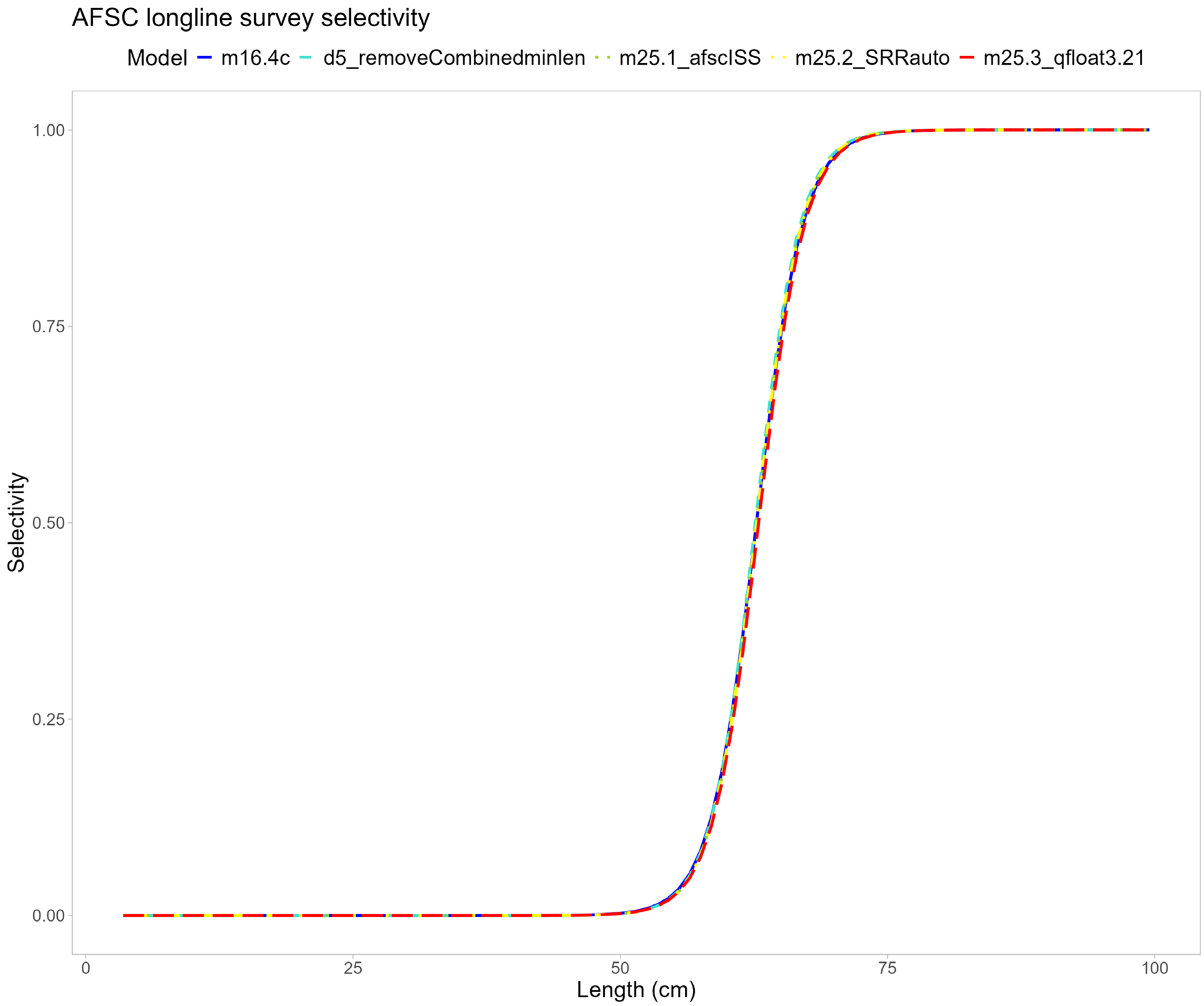


Figure 14. Longline survey selectivity for models 16.4c, d5, and 25.1-25.3.

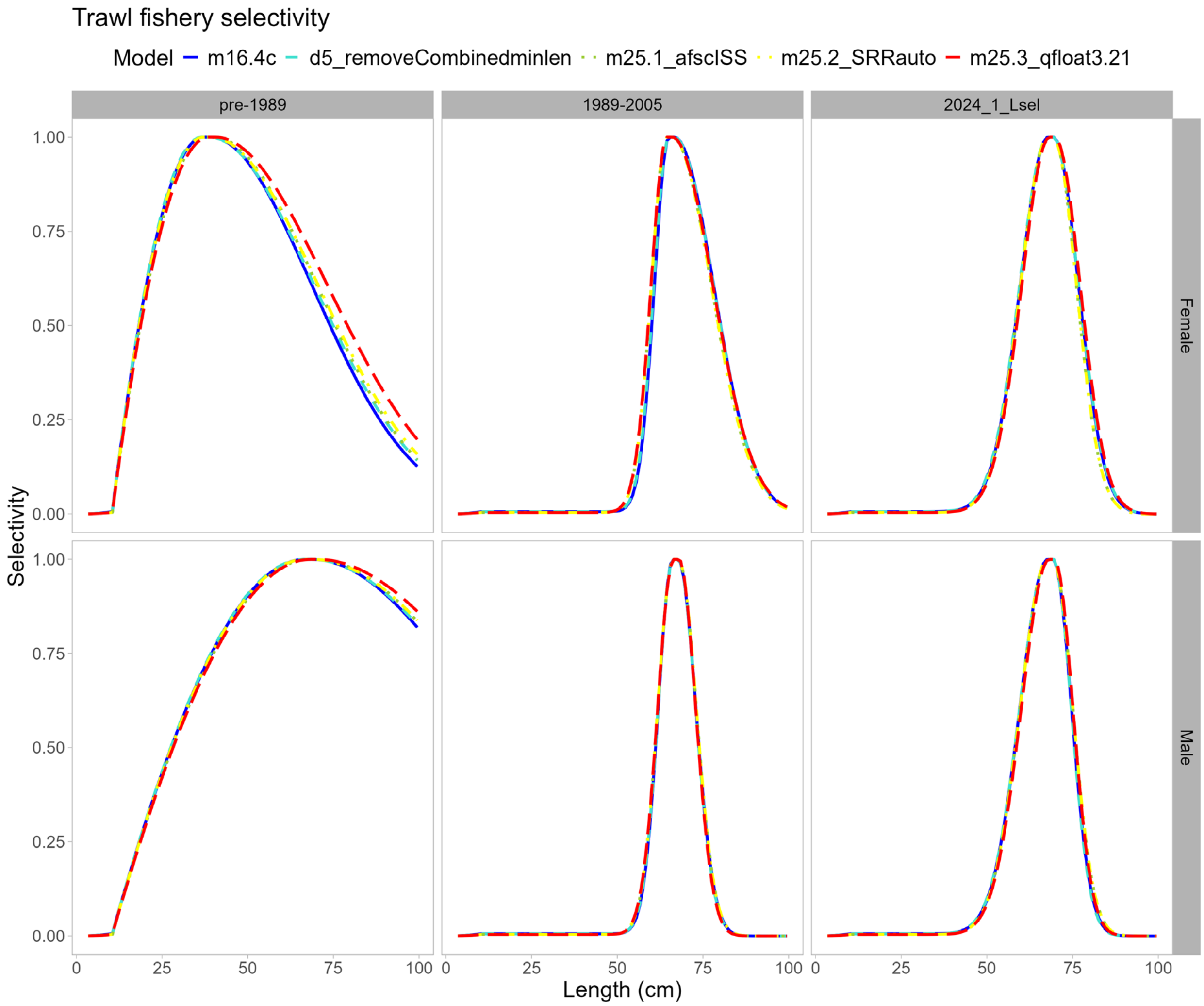


Figure 15. Trawl fishery selectivity for models 16.4c, d5, and 25.1-25.3.

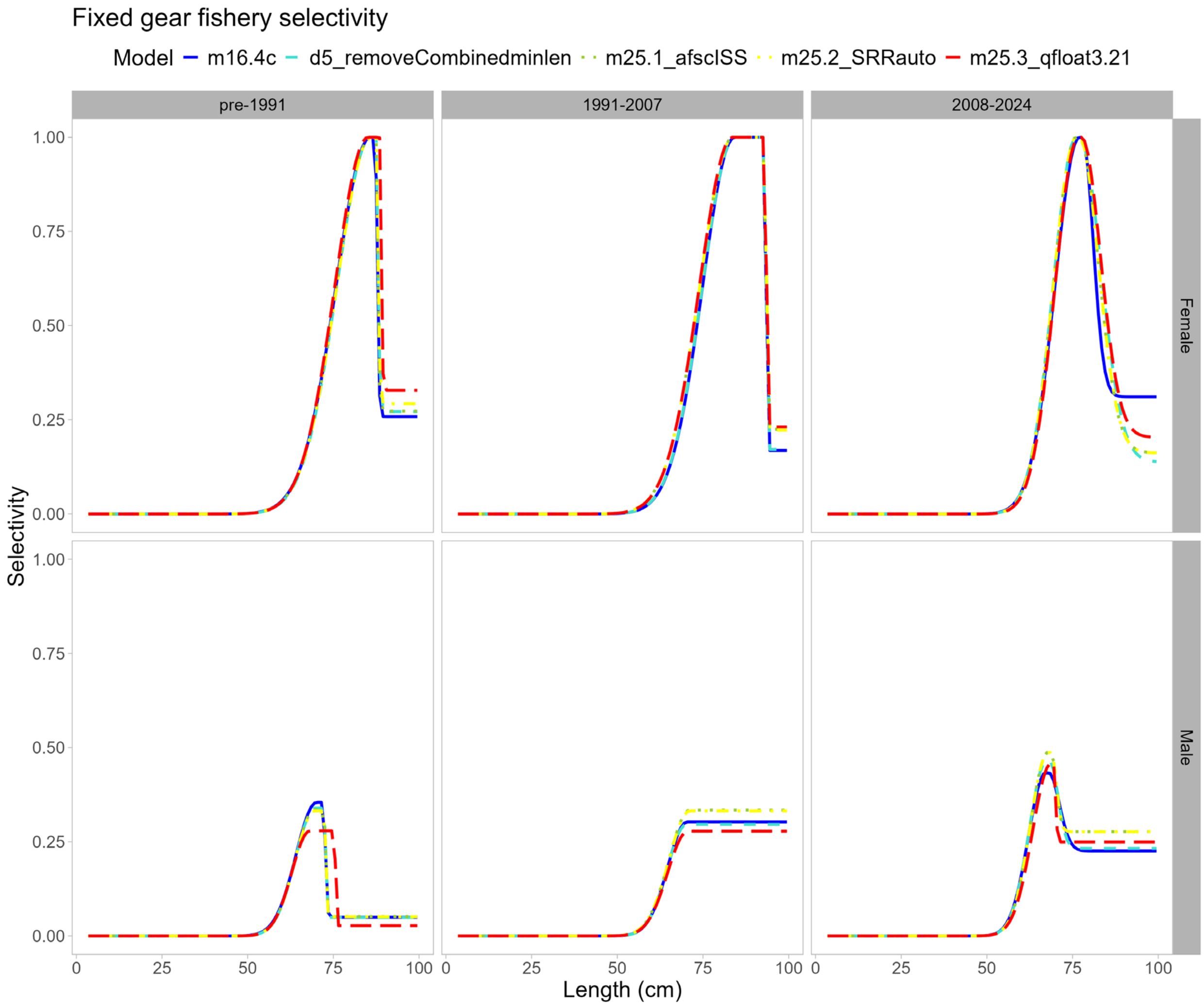
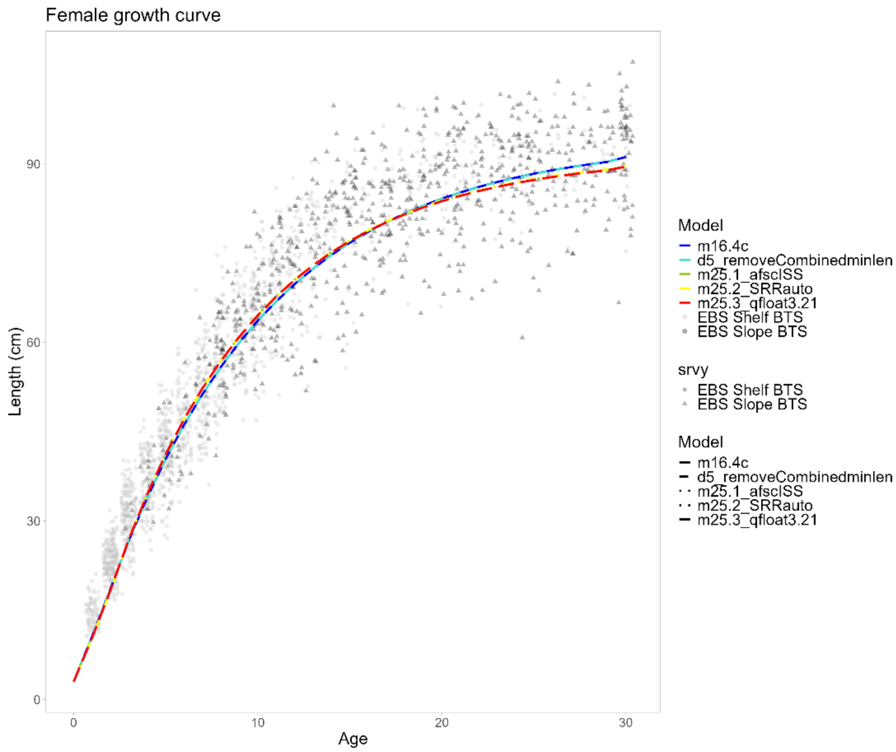


Figure 16. Logline fishery selectivity for models 16.4c, d5, and 25.1-25.3.

a)



b)

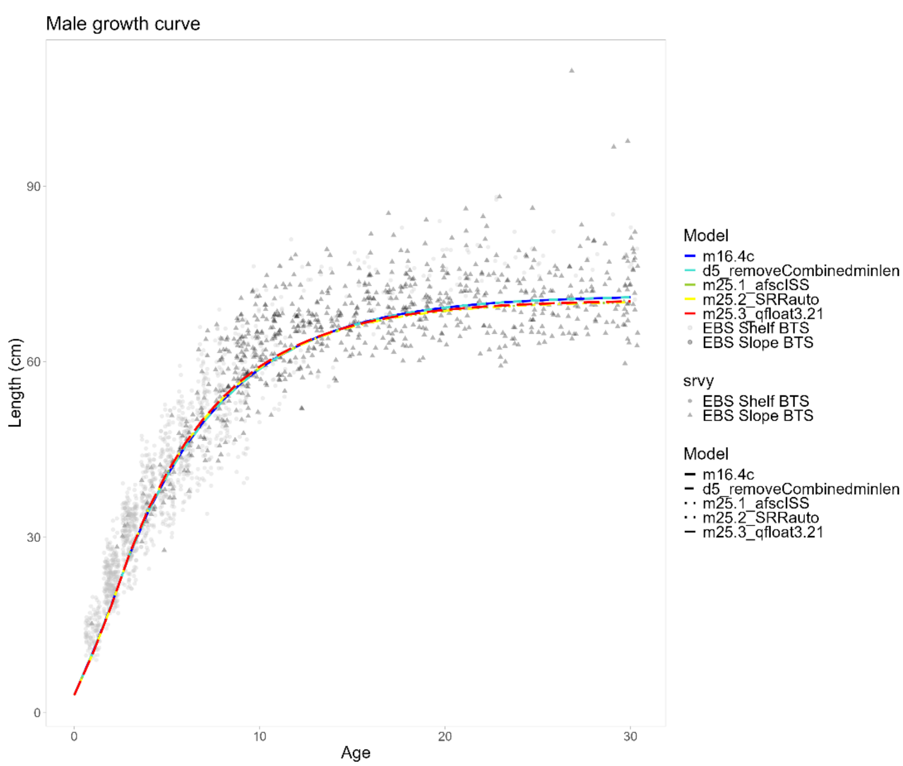


Figure 17. von Bertalanffy growth estimated by models 16.4c, d5, and 25.1 – 25.3; a) female and b) male.

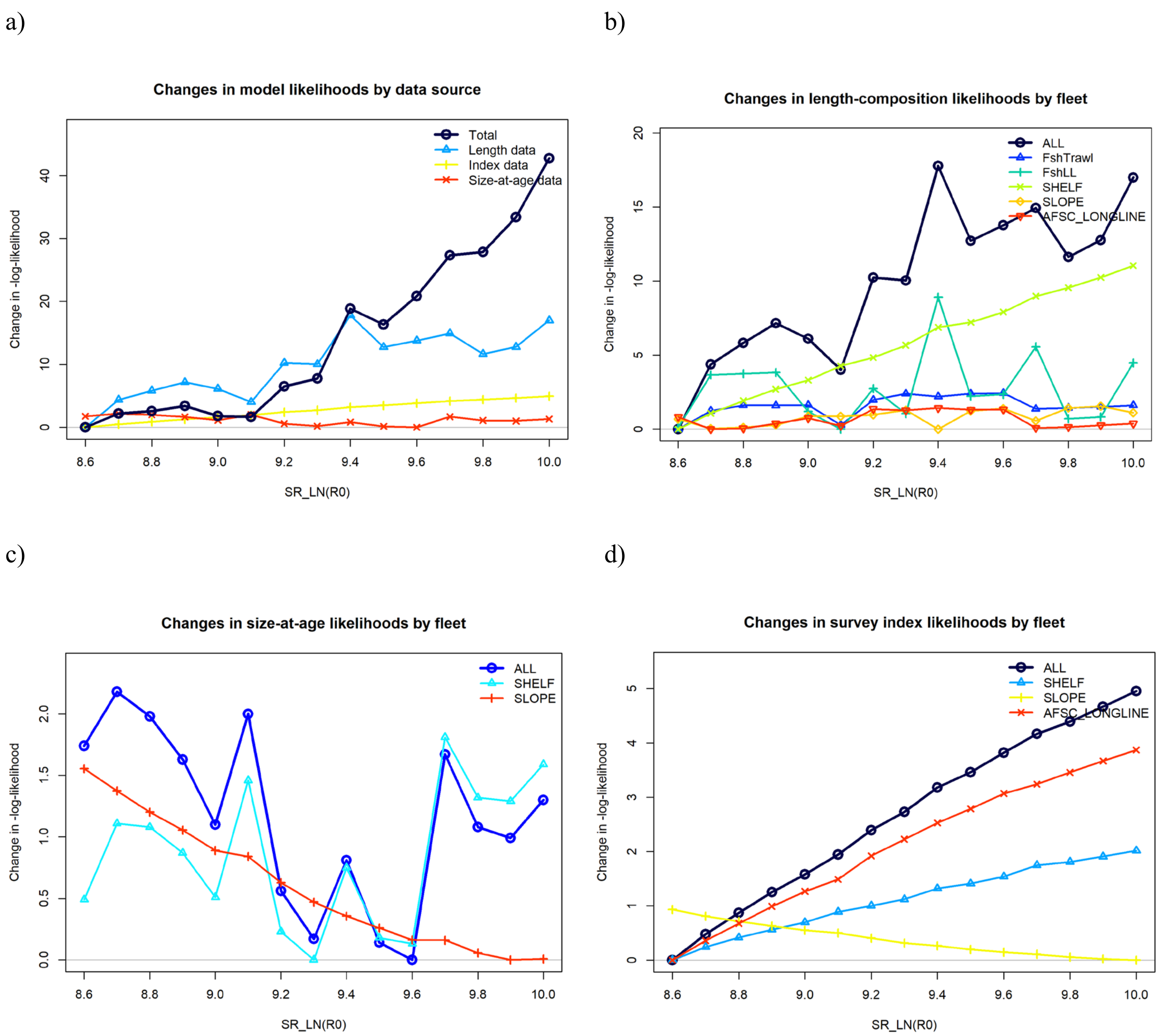


Figure 18. Likelihood profile on R0 for model 25.3 a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

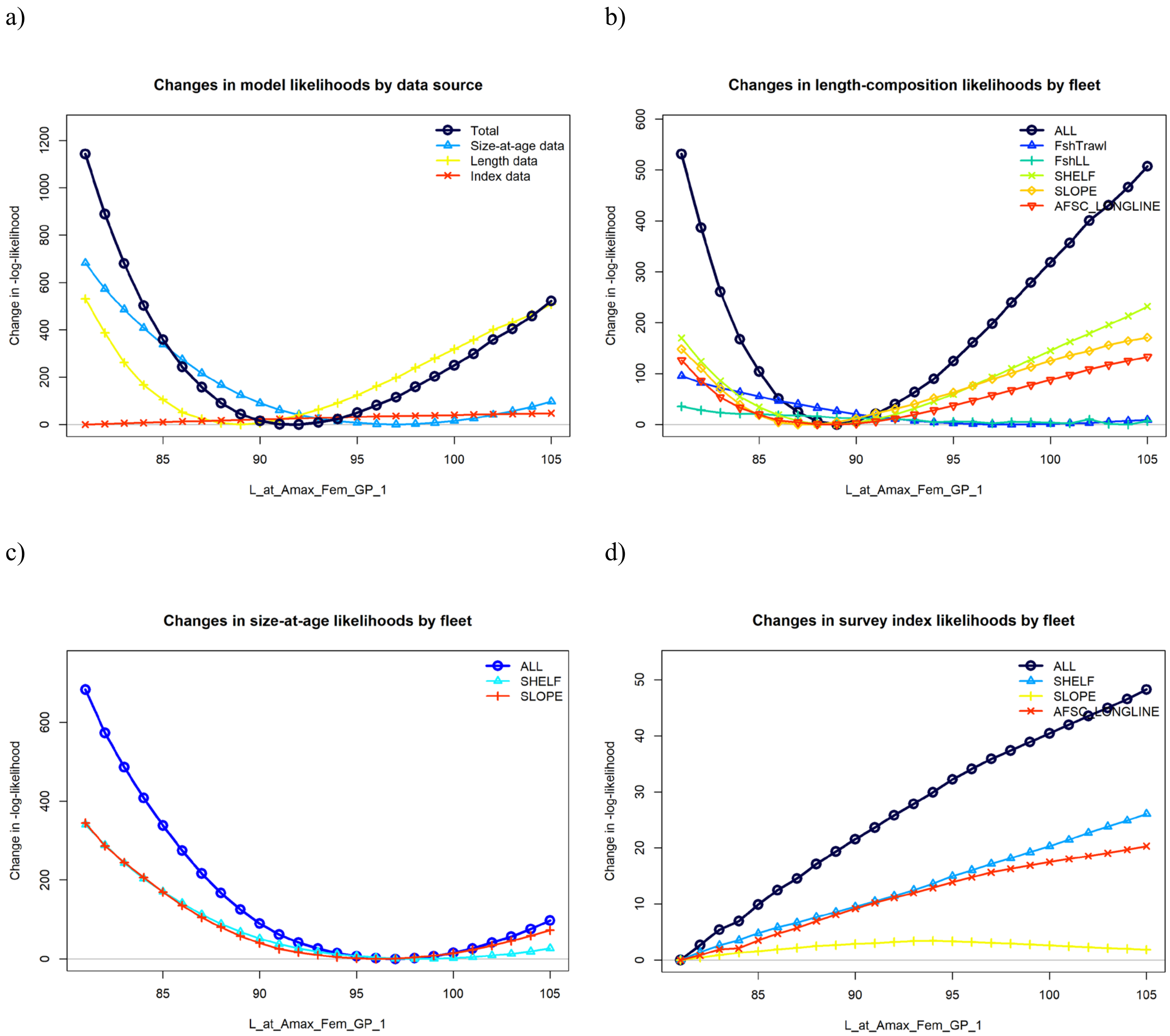


Figure 19. Likelihood profile on female Linf for model 25.3. a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

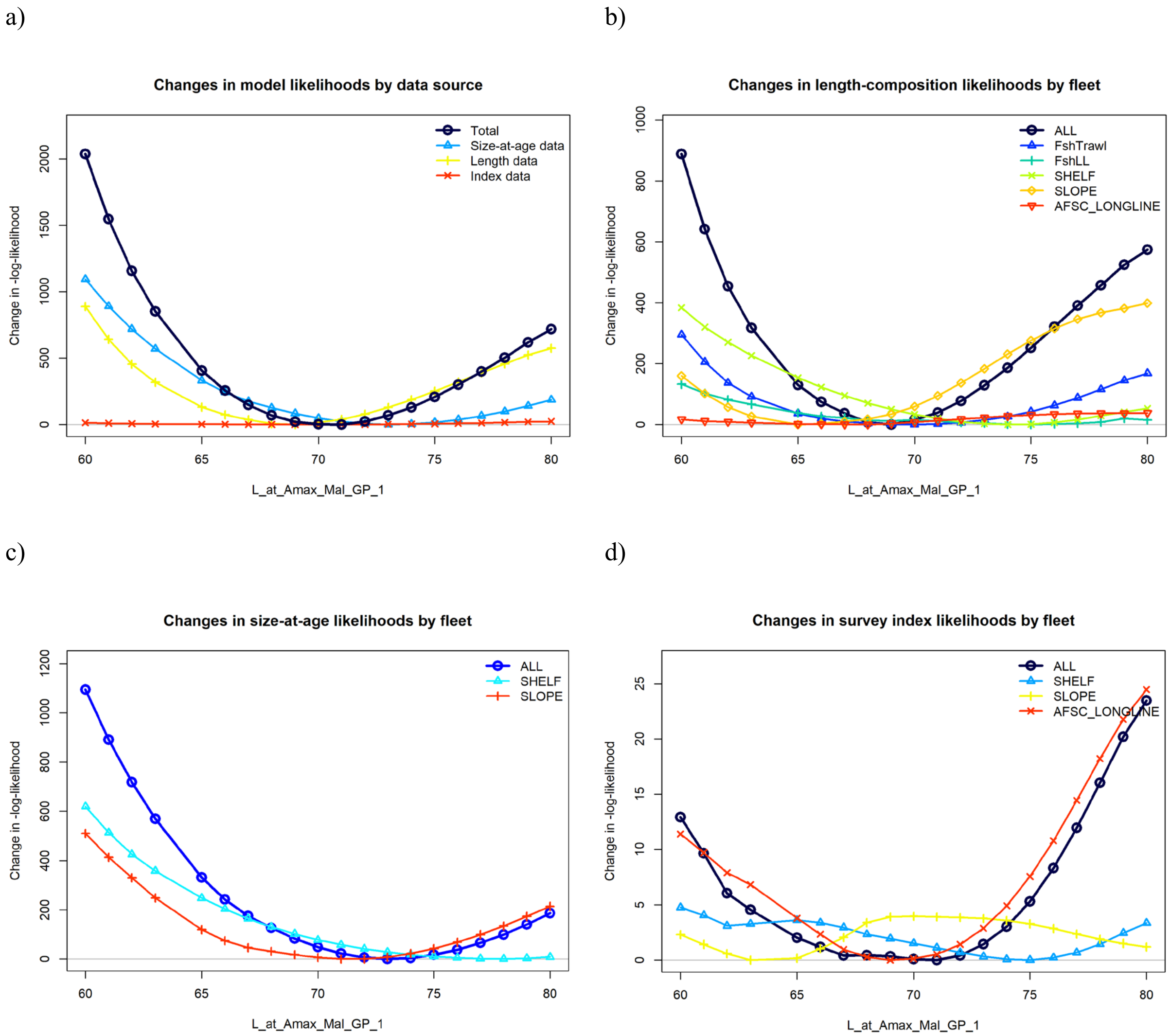


Figure 20. Likelihood profile on male Linf for model 25.3. a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

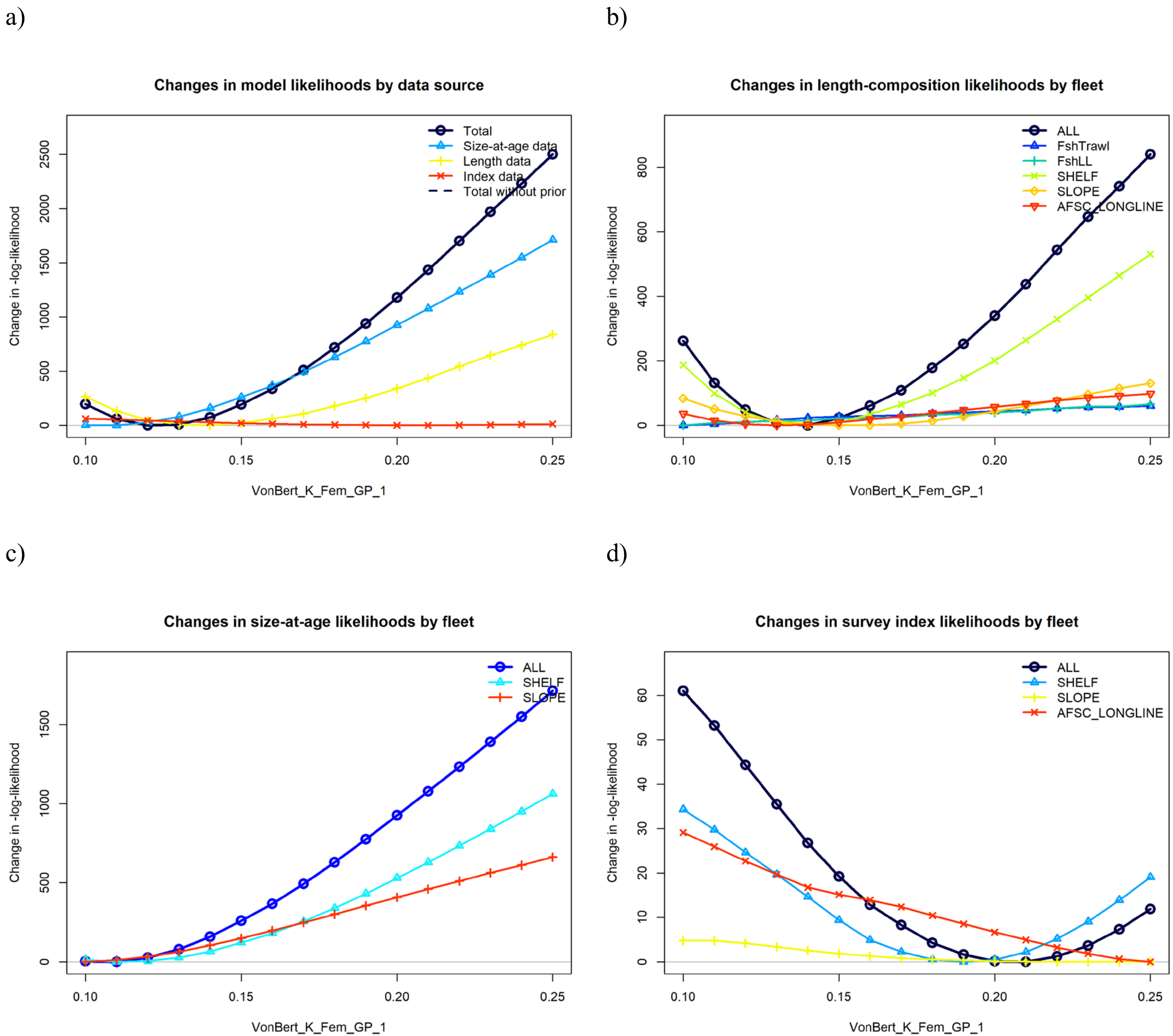
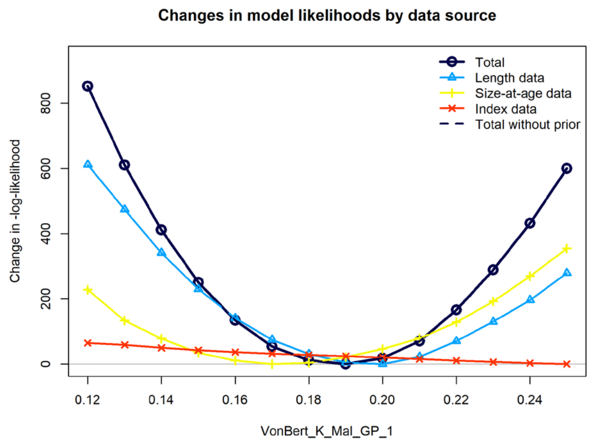
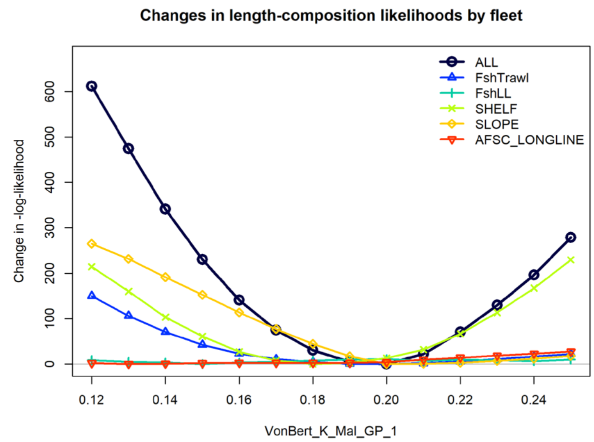


Figure 21. Likelihood profile on female von Bertalanffy growth coefficient for model 25.3. a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

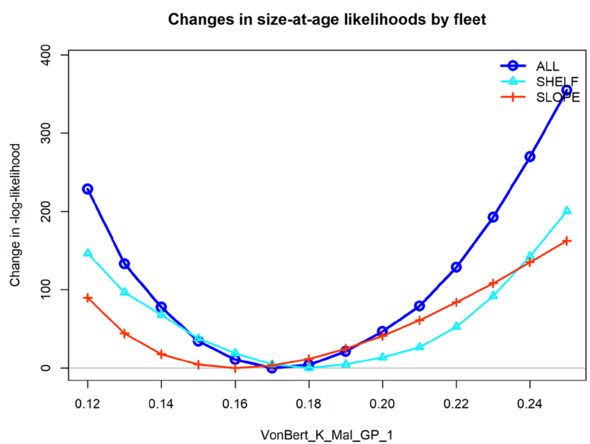
a)



b)



c)



d)

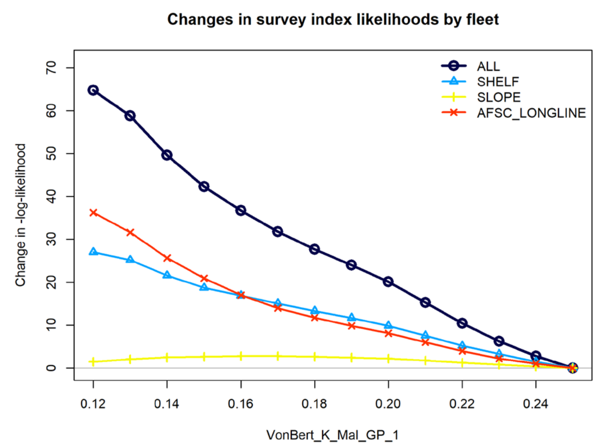


Figure 22. Likelihood profile on male von Bertalanffy growth coefficient for model 25.3. a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

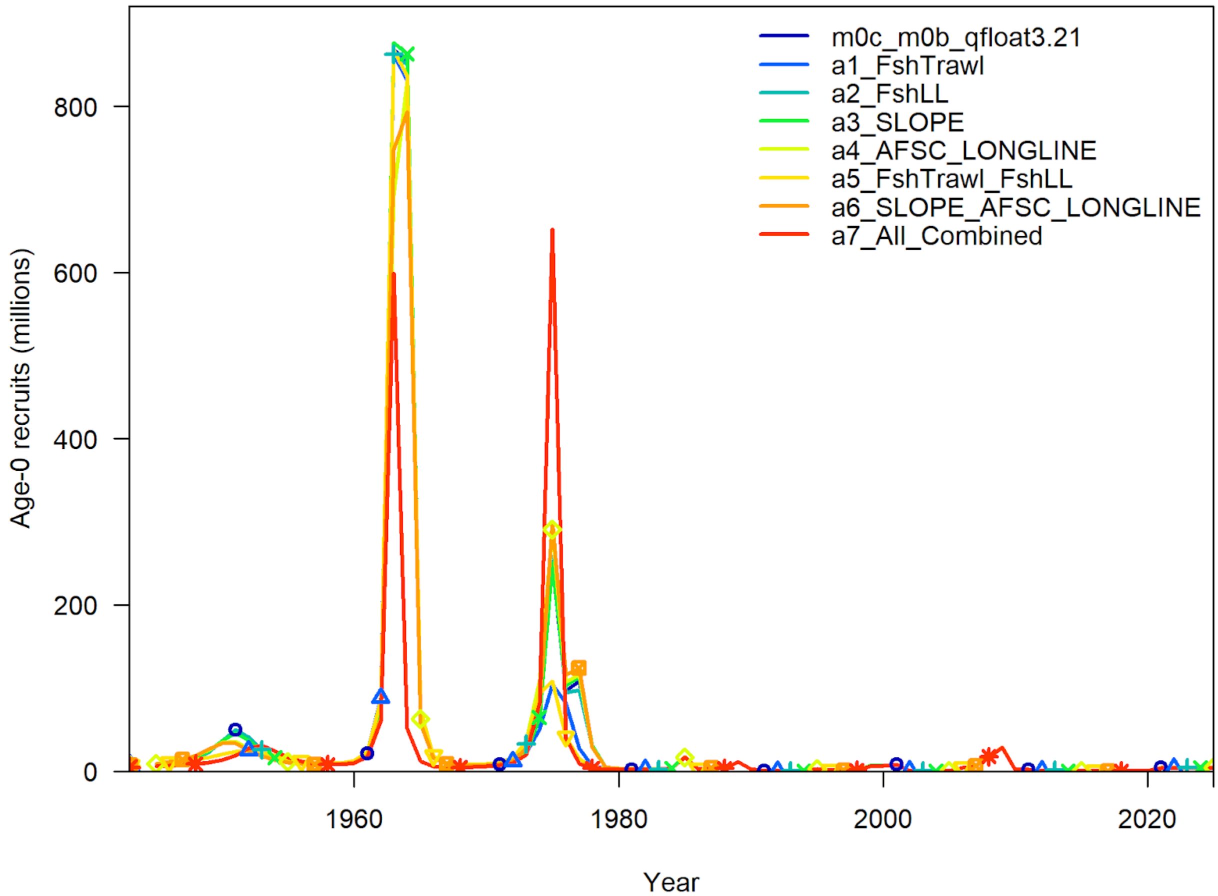


Figure 23. Age-0 recruits estimated by models where early length data were removed one fleet at a time and then in combination.

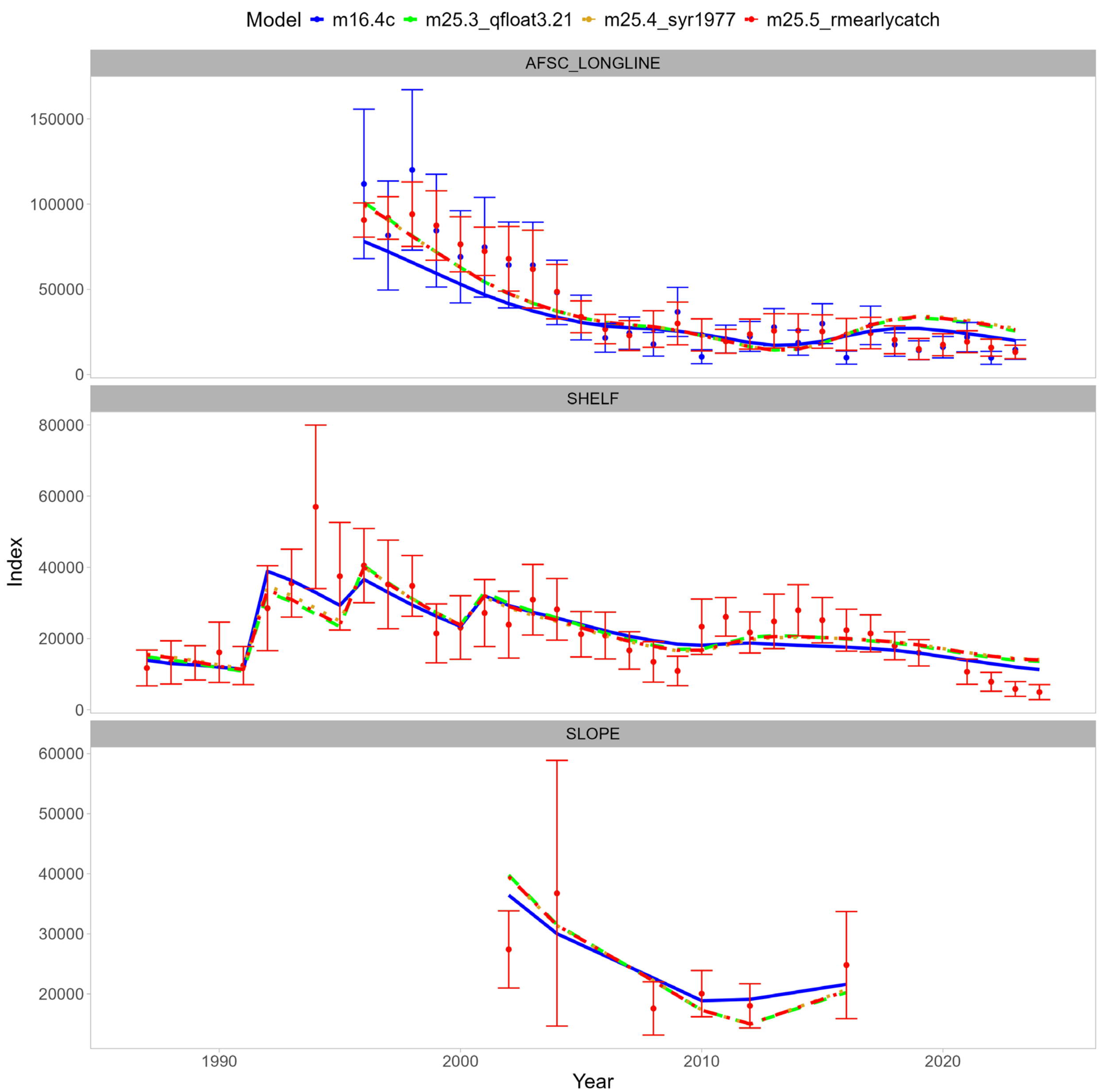


Figure 24. Fits to the longline survey RPNs and the bottom trawl survey biomass; models 16.4c, 25.3-25.5.

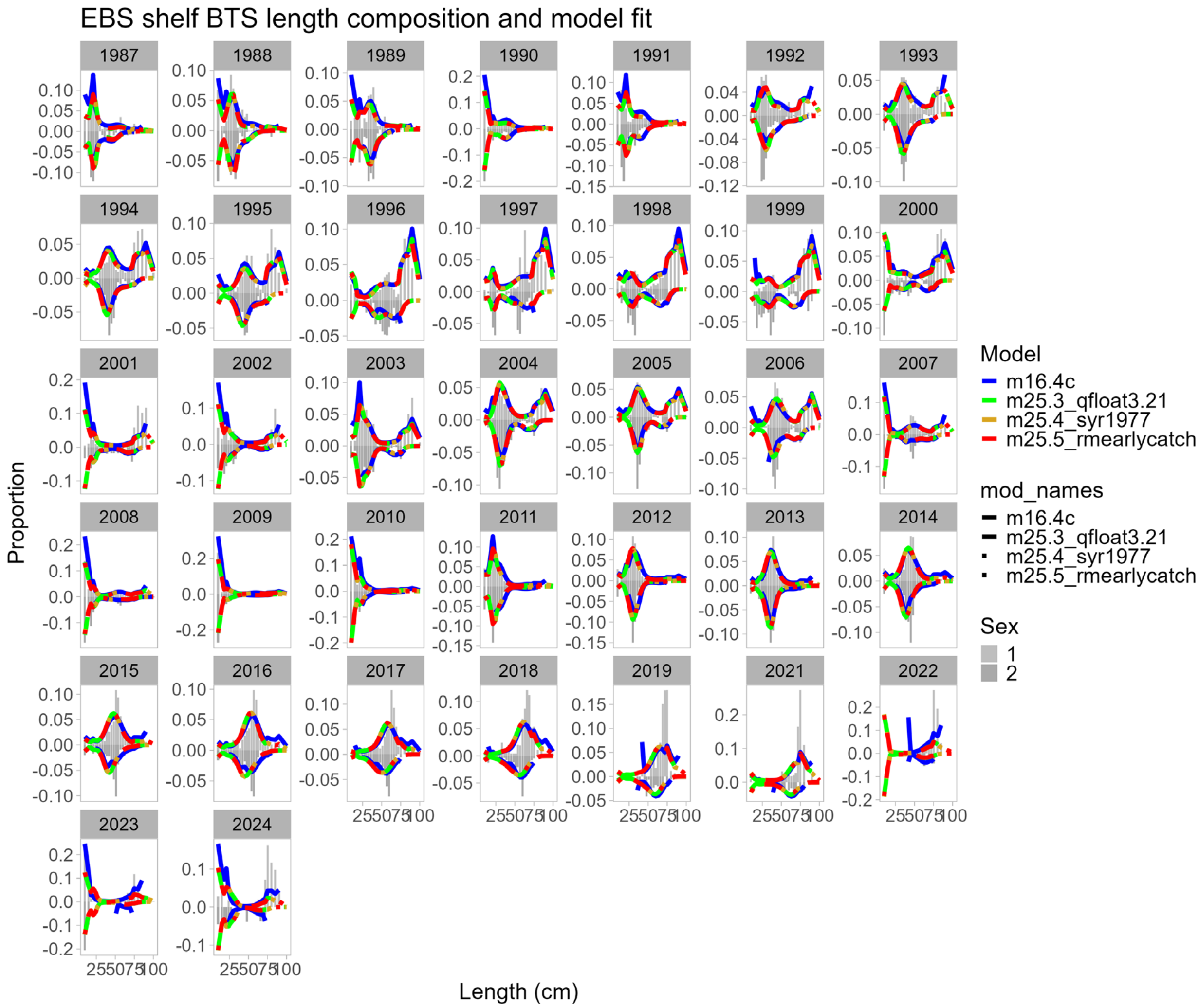


Figure 25. Fits to the EBS shelf bottom trawl survey length composition data; models 16.4c, 25.3-25.5.

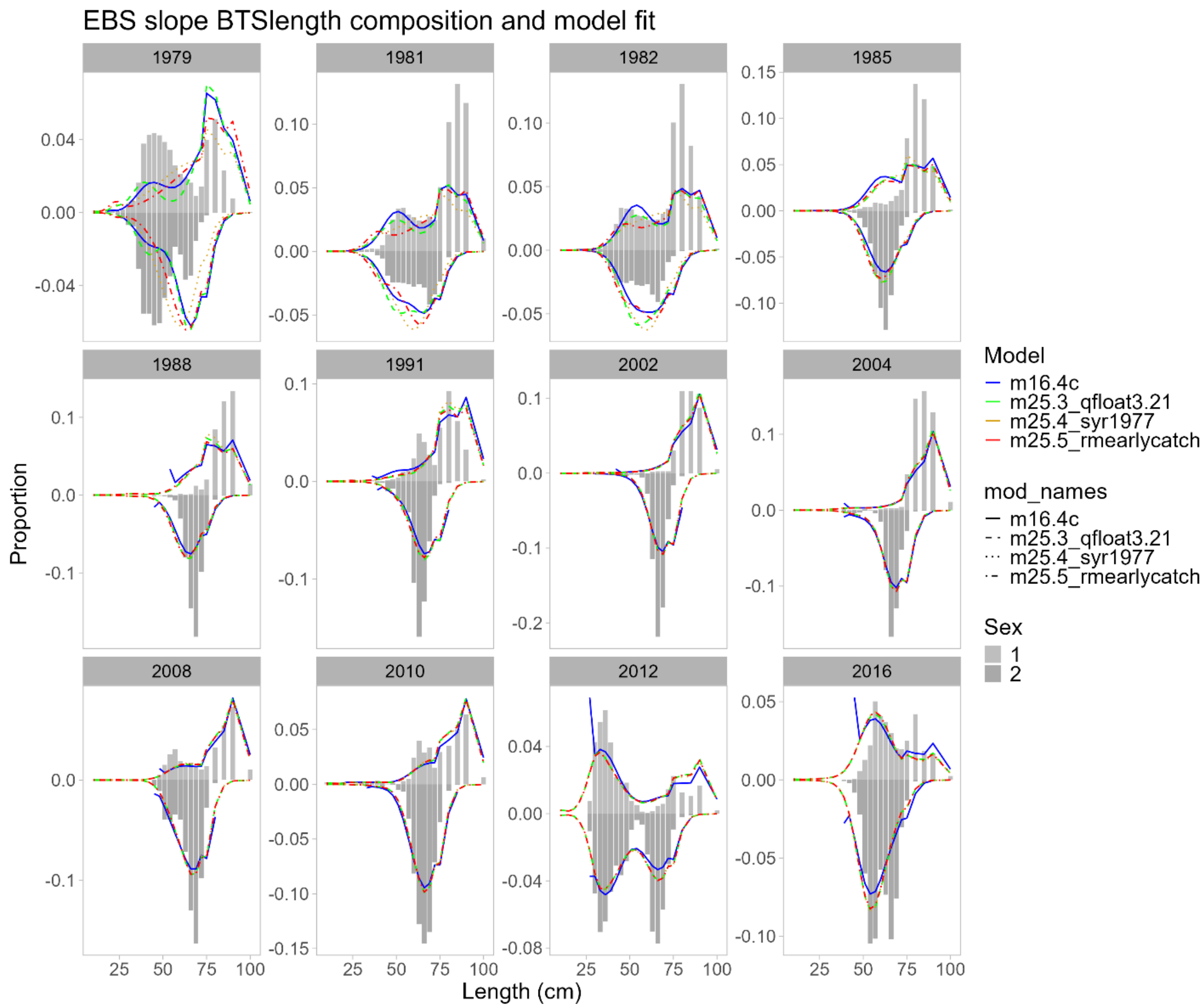


Figure 26. Fits to the EBS slope bottom trawl survey length composition data; models 16.4c, 25.3-25.5.

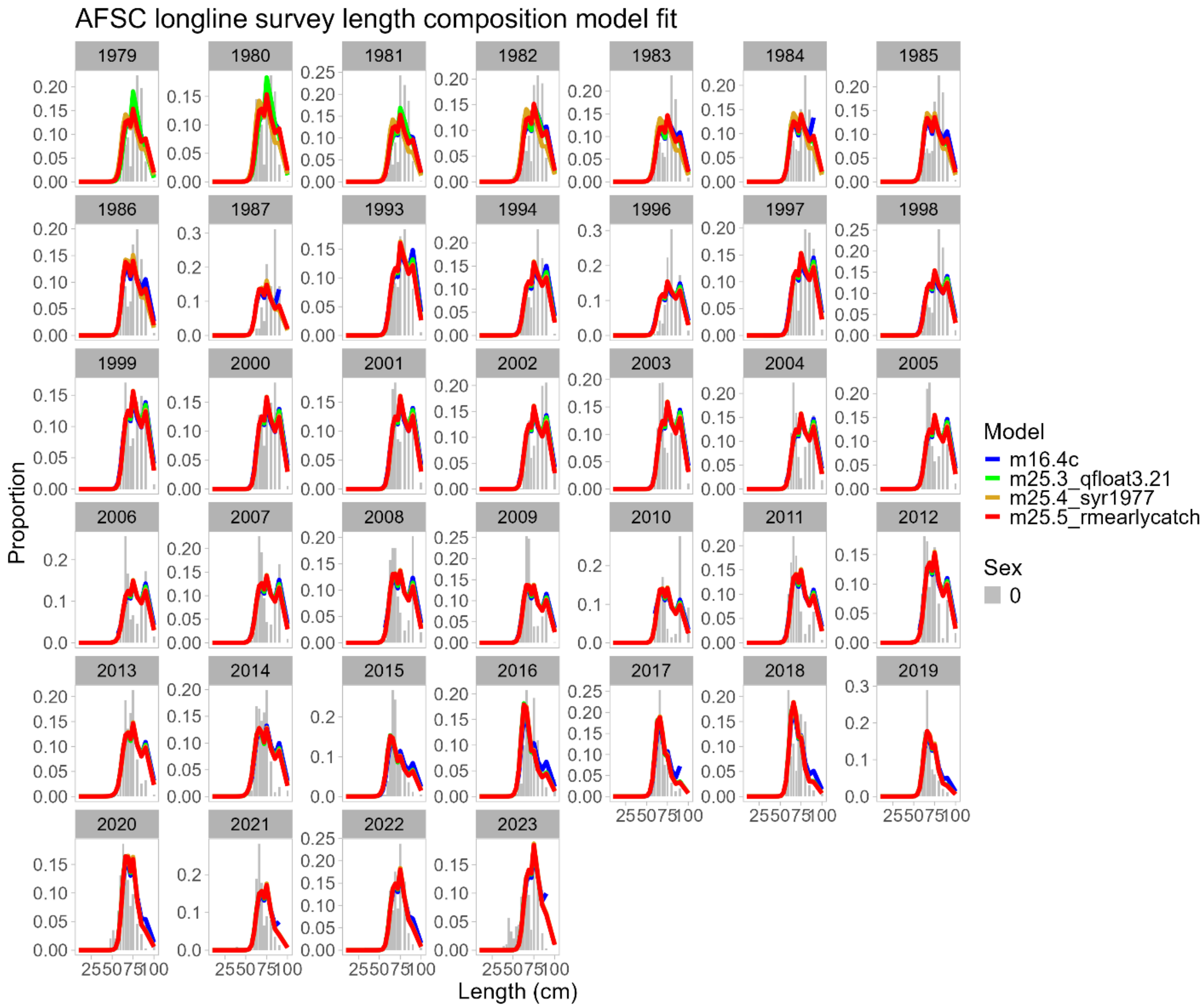


Figure 27. Fits to the longline survey length composition data; models 16.4c, 25.3-25.5.

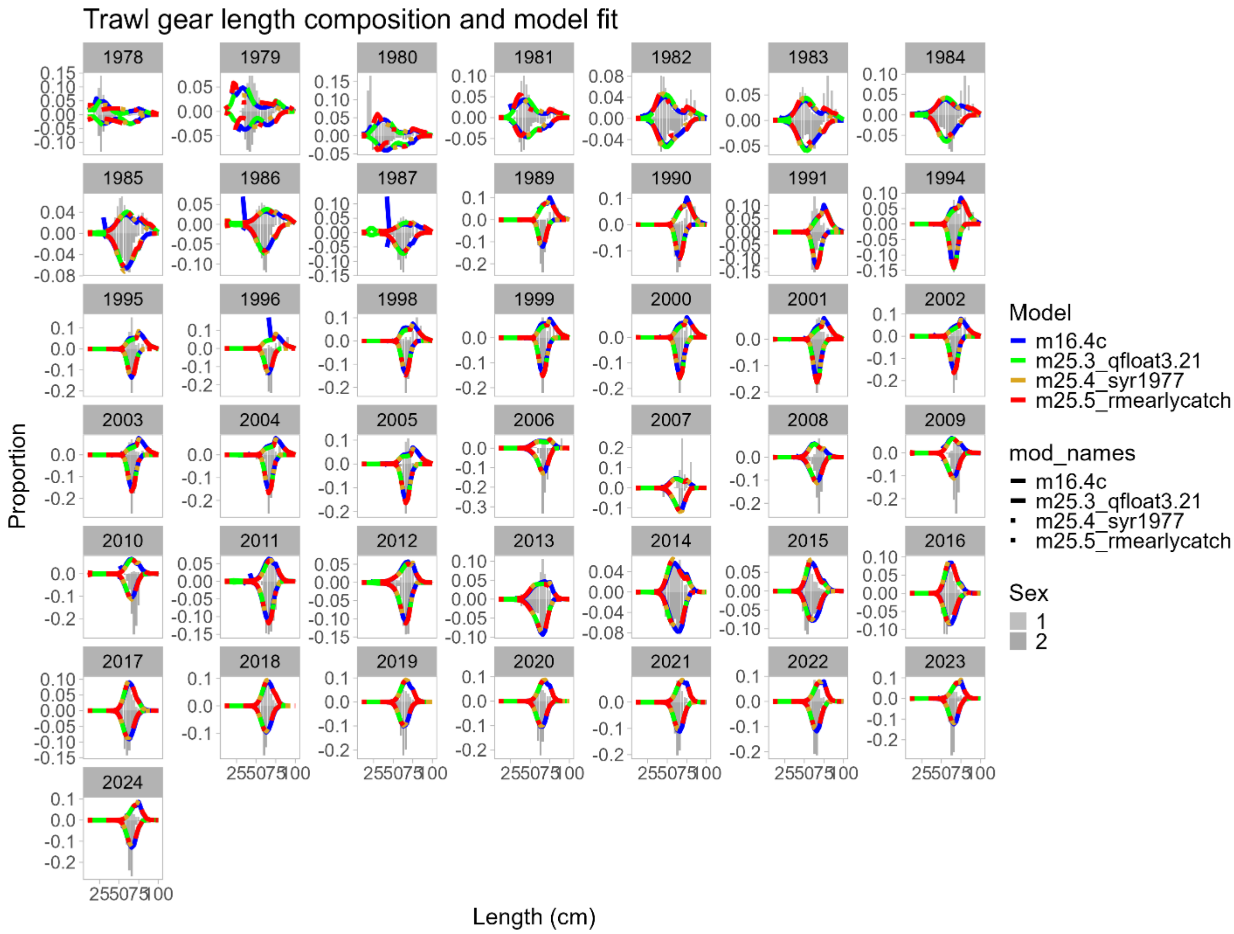


Figure 28. Fits to the trawl fishery length composition data; models 16.4c, 25.3-25.5.

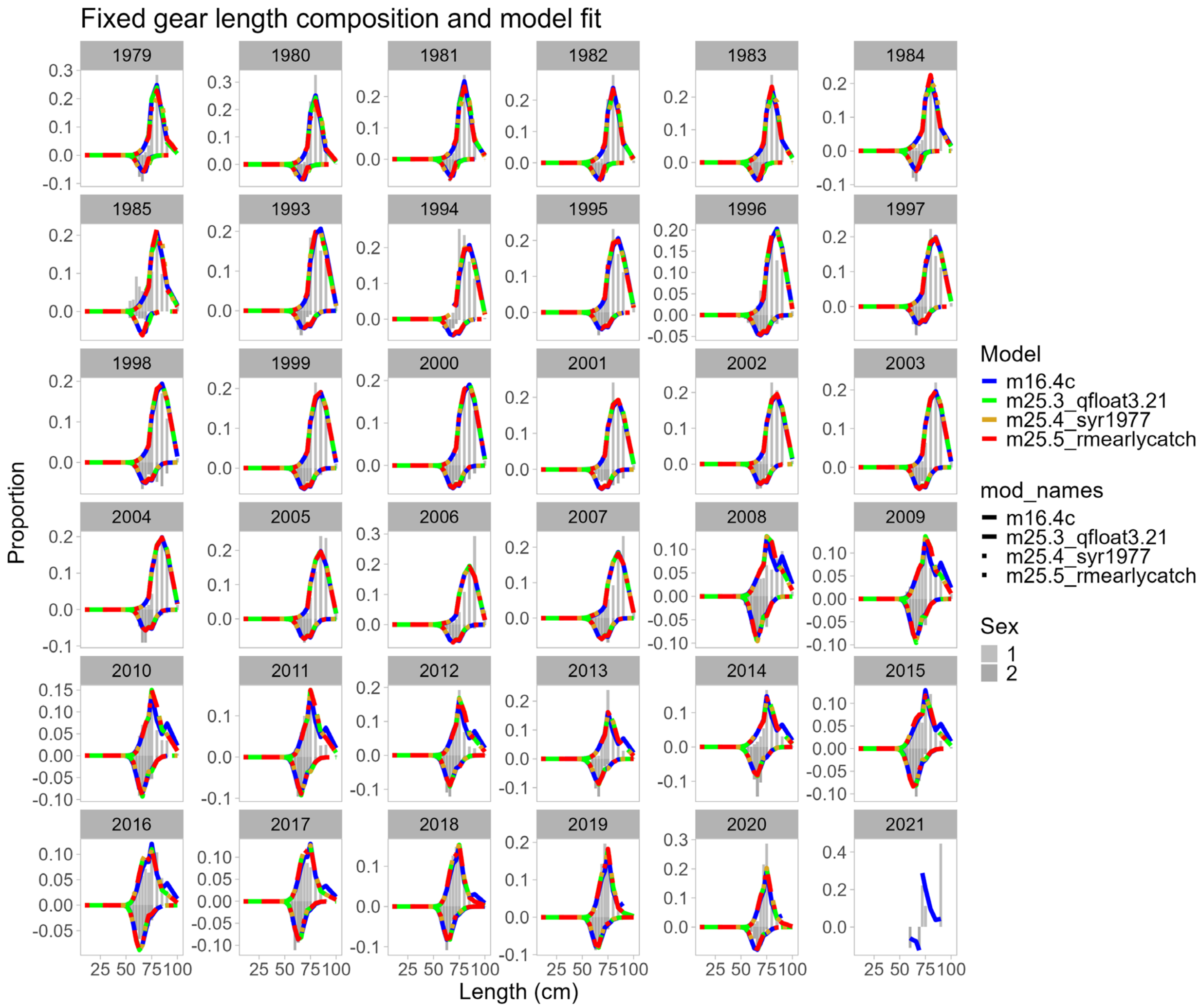


Figure 29. Fits to the longline fishery length composition data; models 16.4c, 25.3-25.5.

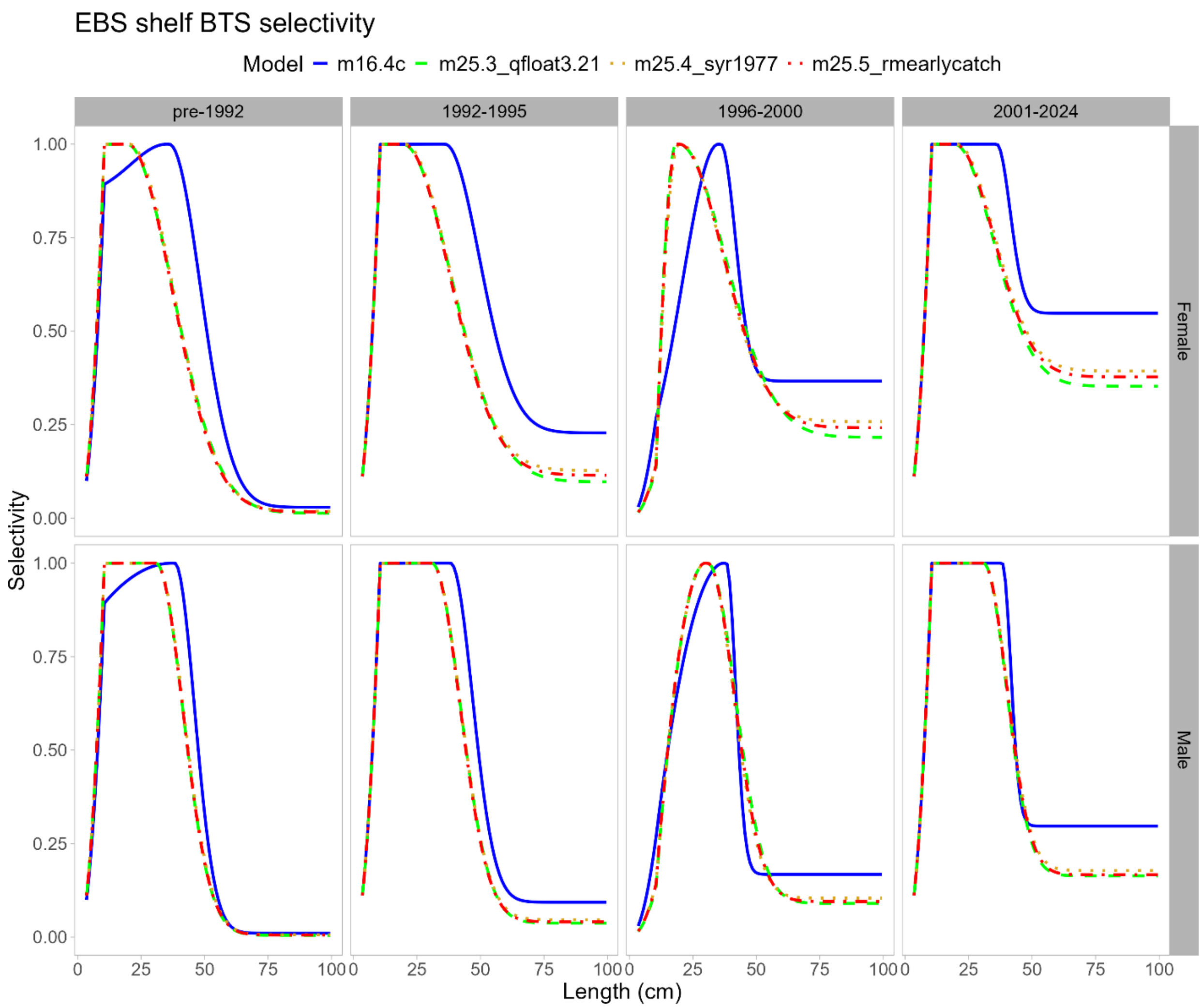


Figure 30. EBS shelf bottom trawl survey selectivity; models 16.4c, 25.3-25.5.

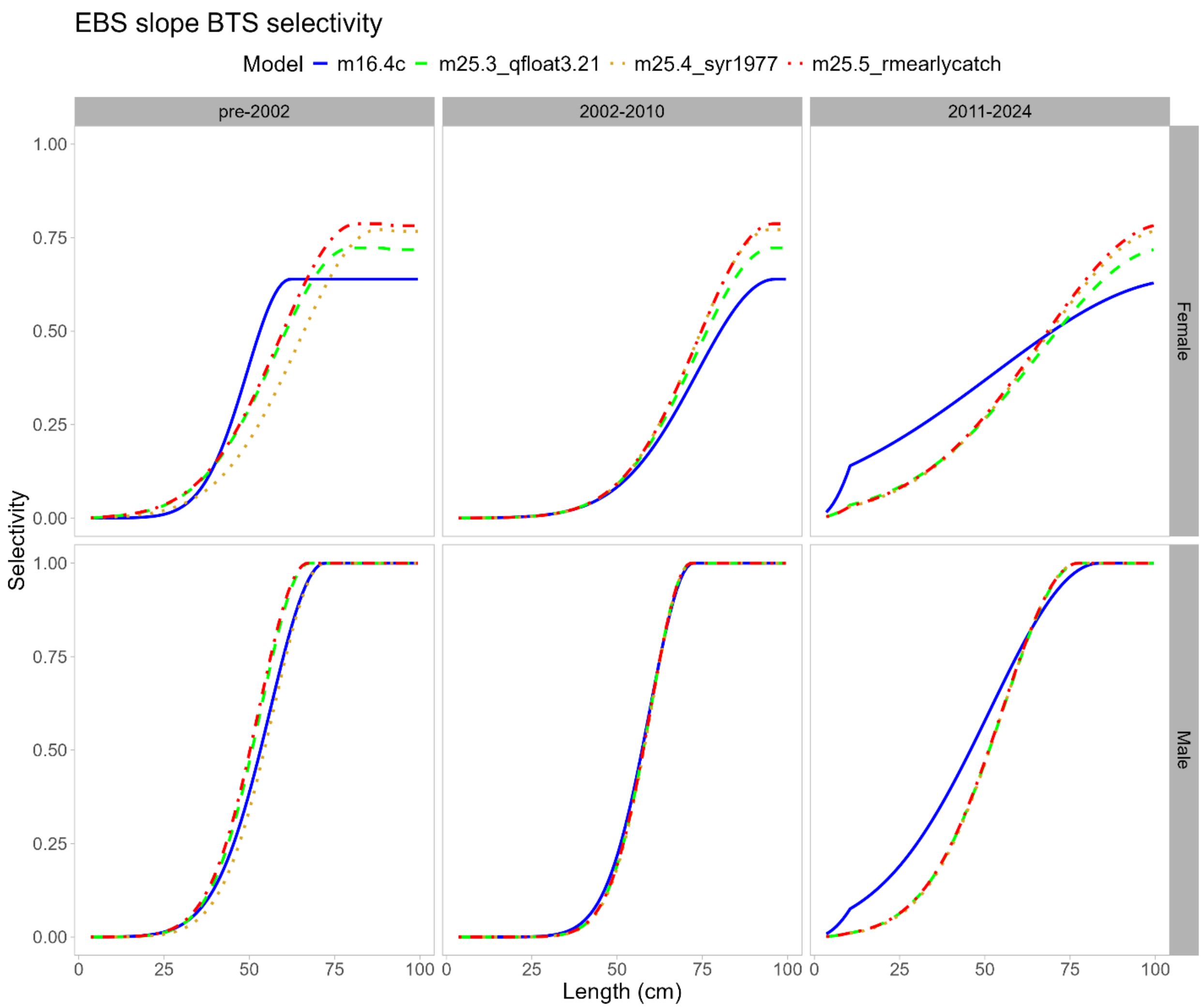


Figure 31. EBS slope bottom trawl survey selectivity; models 16.4c, 25.3-25.5.

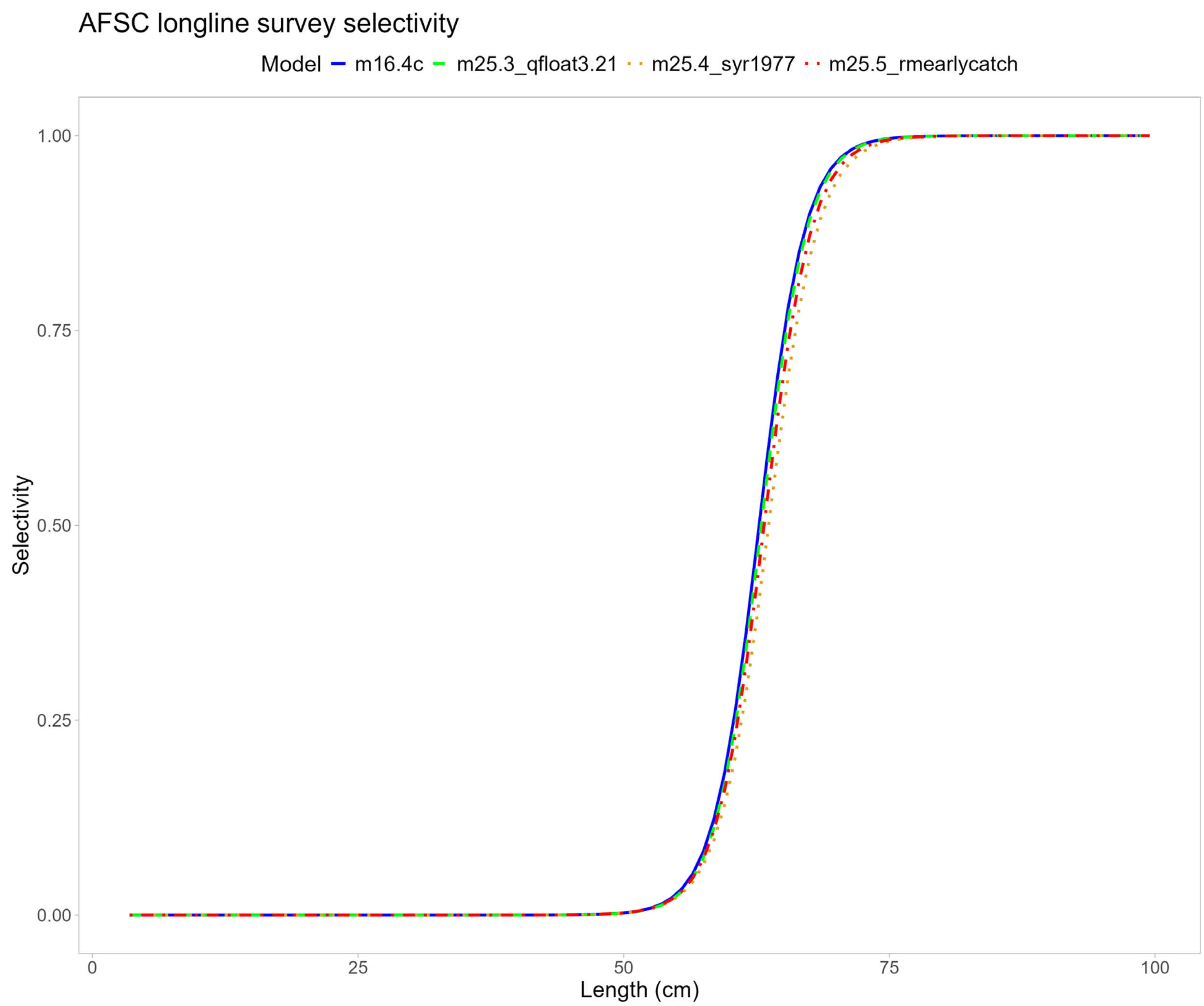


Figure 32 Longline survey selectivity 16.4c, 25.3-25.5.

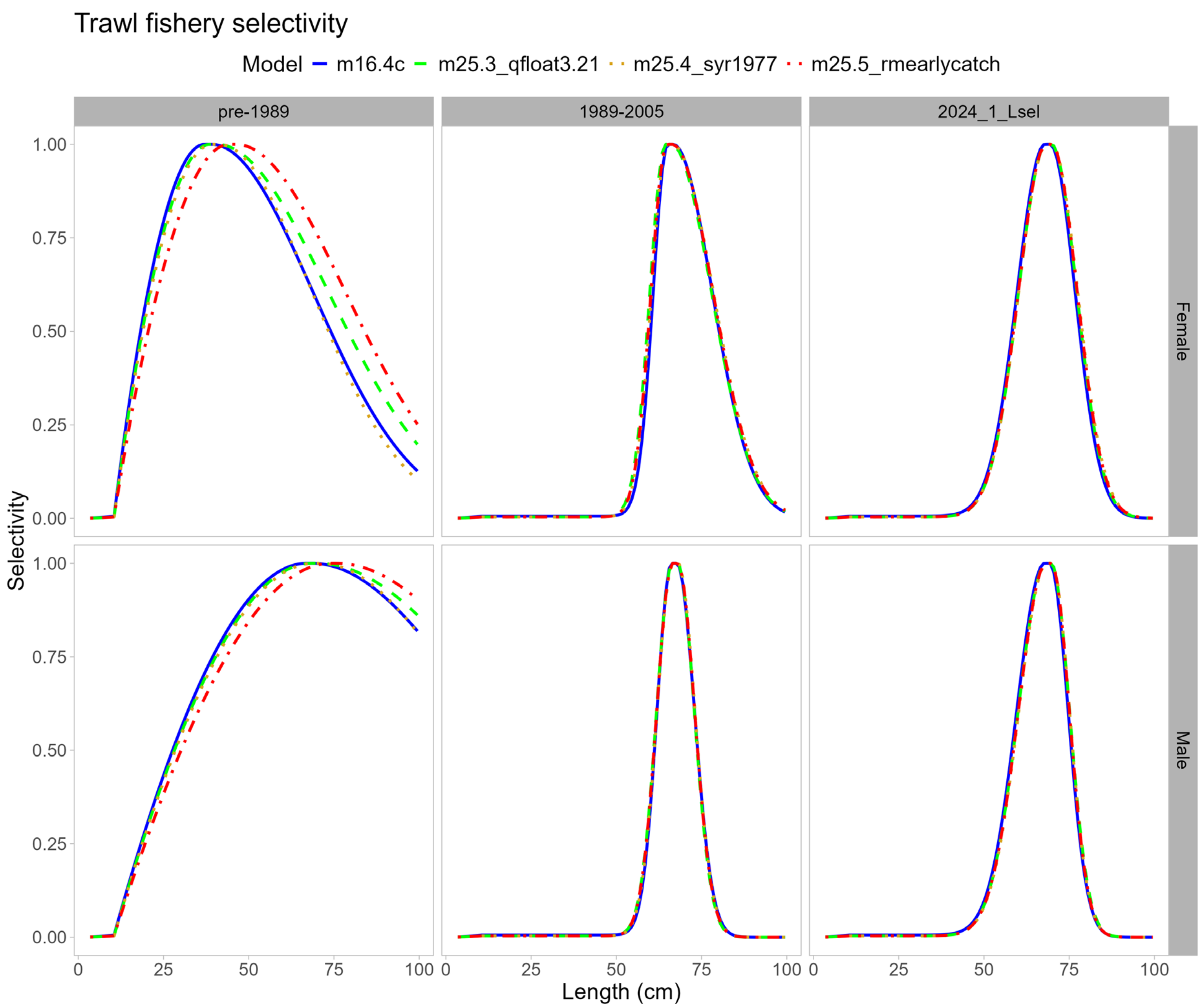


Figure 33. Trawl fishery selectivity; models 16.4c, 25.3-25.5.

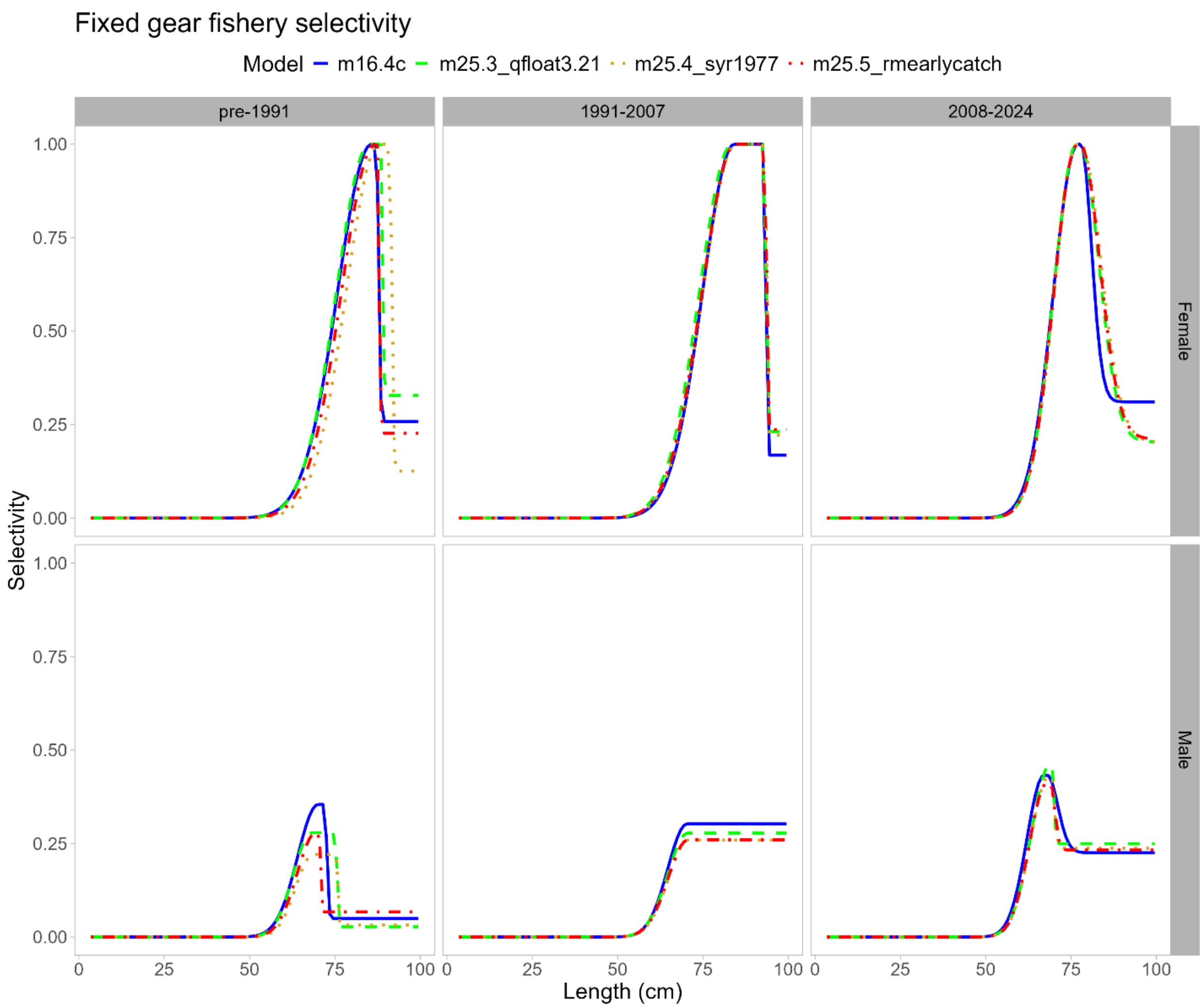
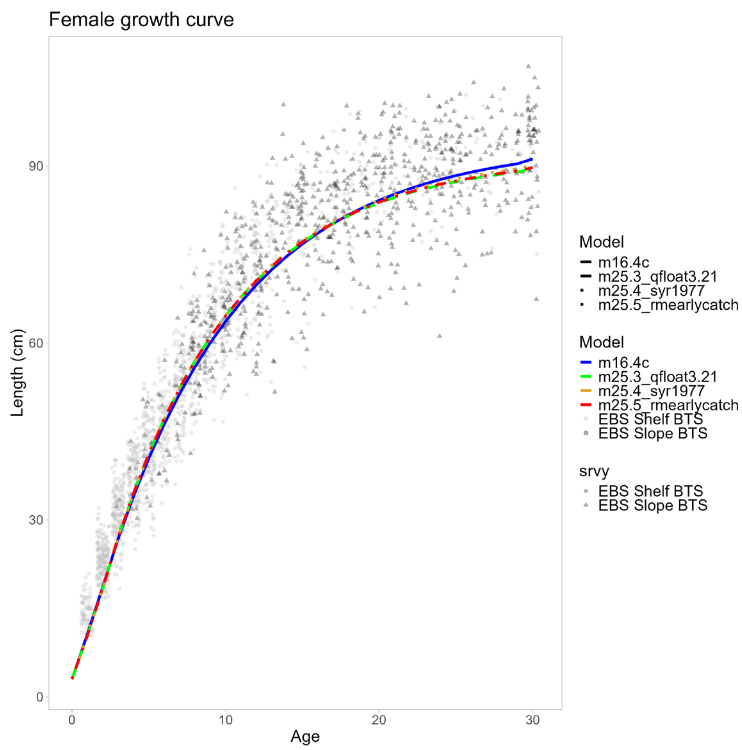


Figure 34. Longline selectivity; models 16.4c, 25.3-25.5.

a)



b)

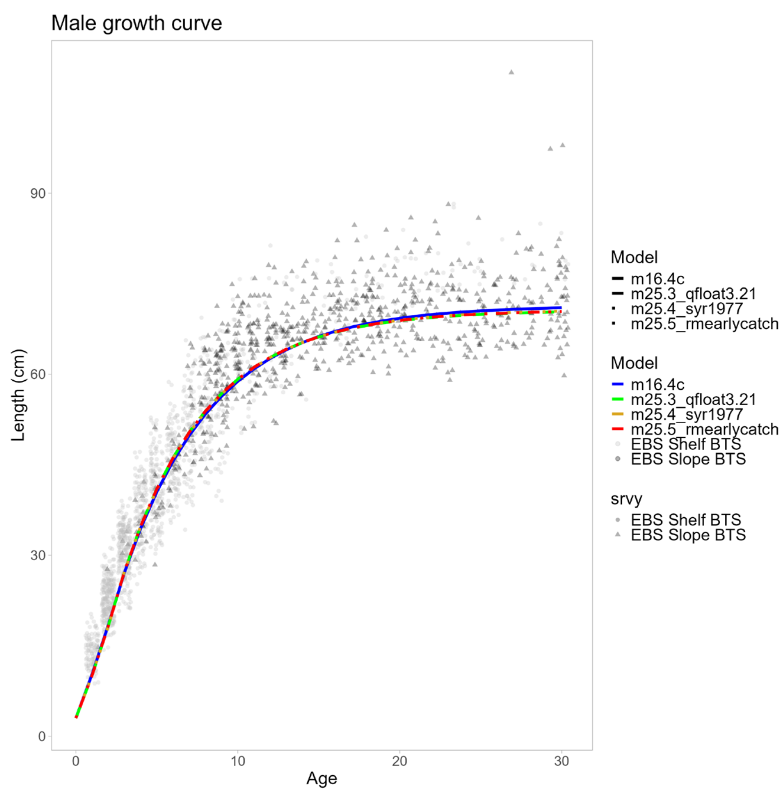


Figure 35. Estimates of vonBertalanffy growth for models 16.4c, 25.3-25.5 and a) females, and b) males.

a) b)

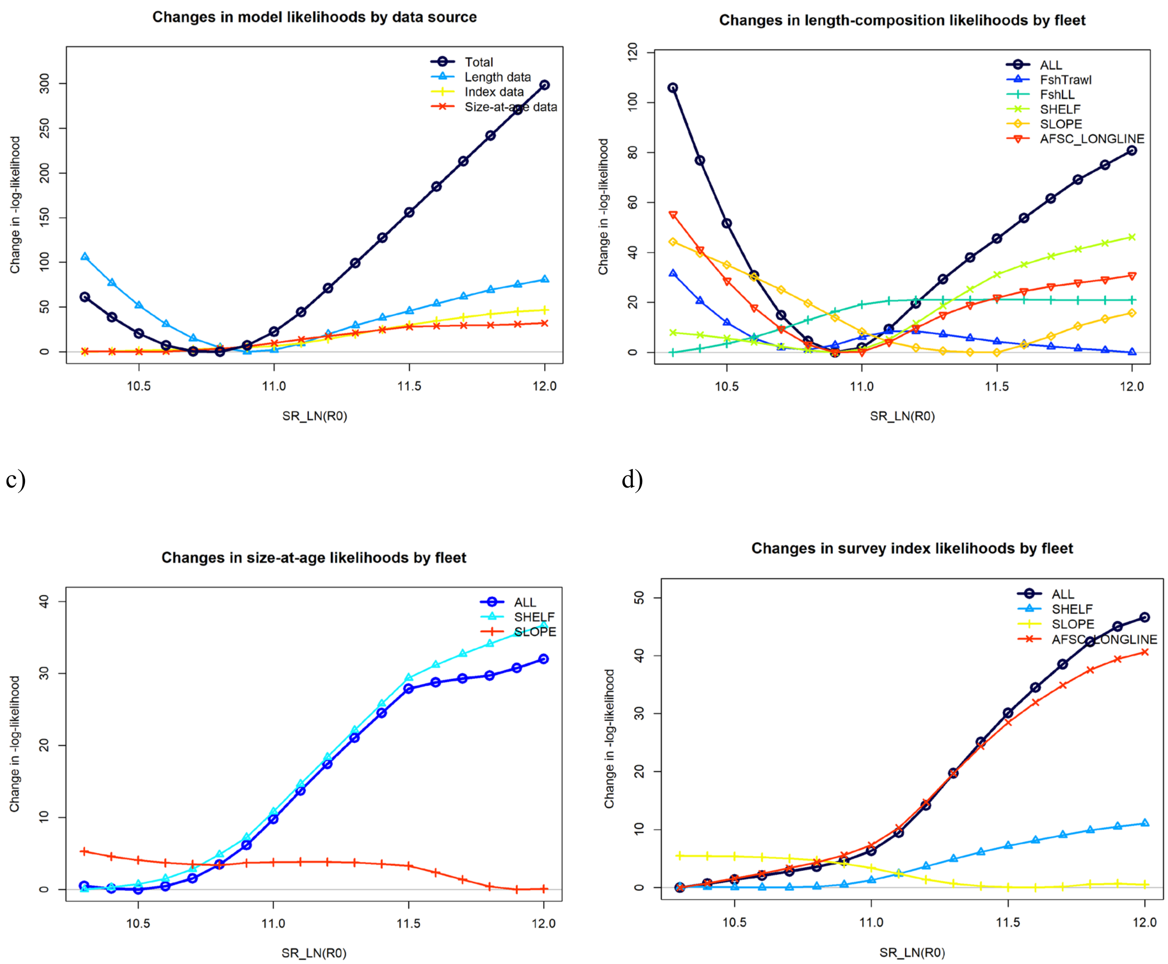


Figure 36. Likelihood profile on R0 for model 25.5 a) Total likelihood and total likelihood component b) fleet specific length likelihood, c) fleet-specific length at age likelihood, and d) fleet specific survey likelihood.

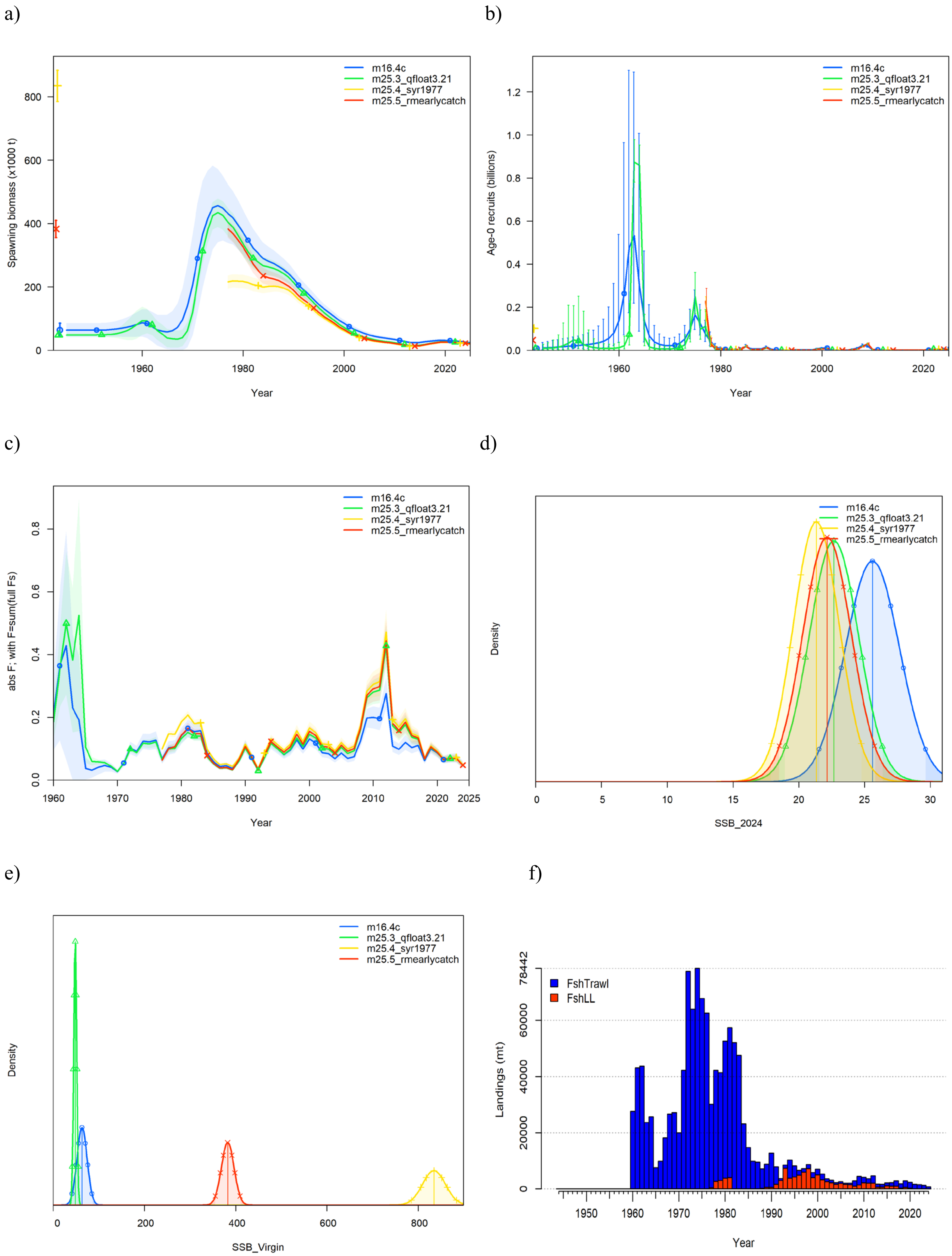
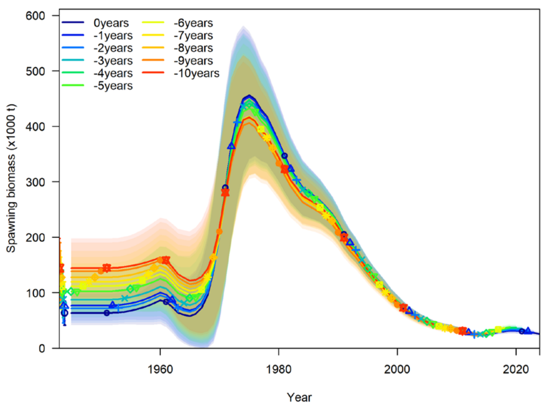
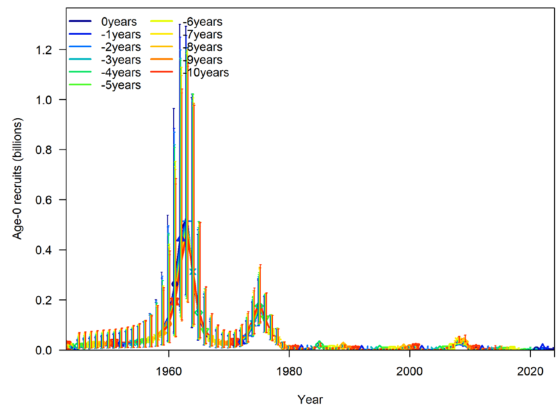


Figure 37. a) Spawning stock biomass, b) age-0 recruits, c) fishing mortality, d) SSB 2024 density, and e) unfished SSB density for models 164c and 25.3-25.5.

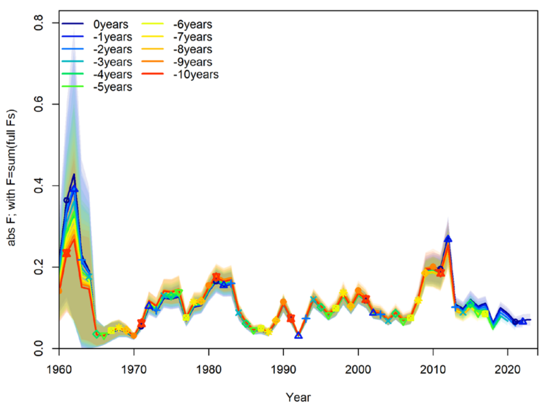
a)



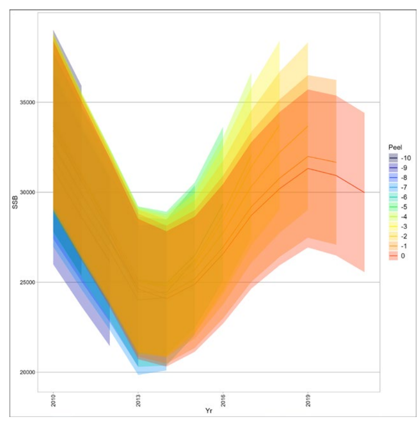
b)



c)



d)



e)

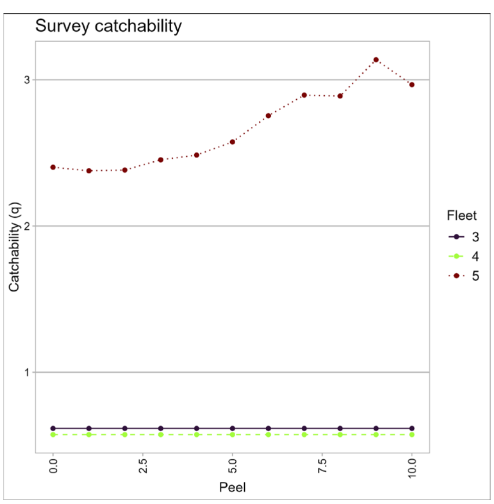


Figure 38. Retrospective analysis of a) spawning stock biomass, b) age-0 recruits, c) fishing mortality, d) zoomed SSB, and e) survey catchability for model 16.4c. Fleet 3 = EBS shelf BTS, fleet 4 = EBS slope BTS, and fleet 5 = longline survey.

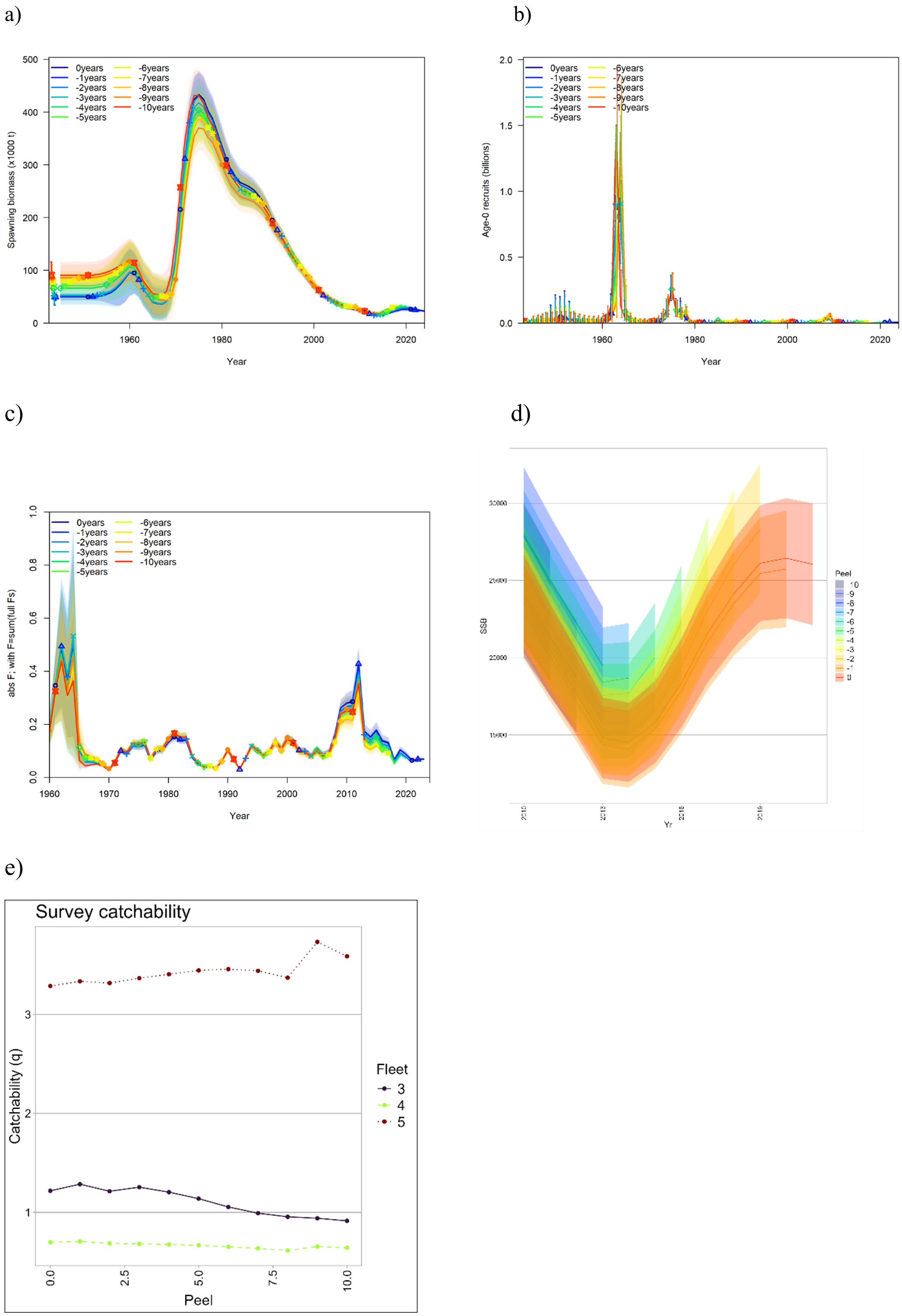
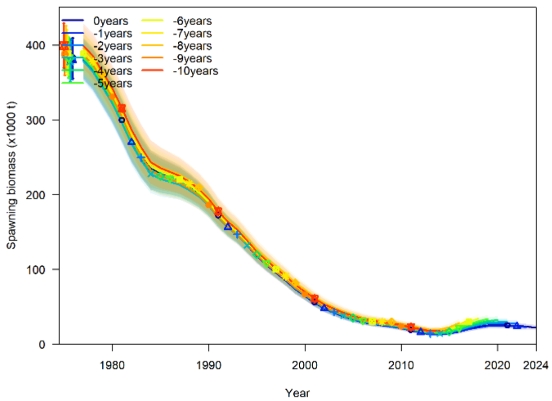
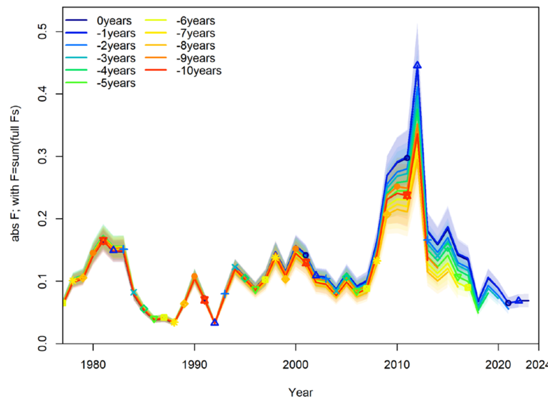


Figure 39. Retrospective analysis of a) spawning stock biomass, b) age-0 recruits, and c) fishing mortality for model 25.3. Fleet 3 = EBS shelf BTS, fleet 4 = EBS slope BTS, and fleet 5 = longline survey.

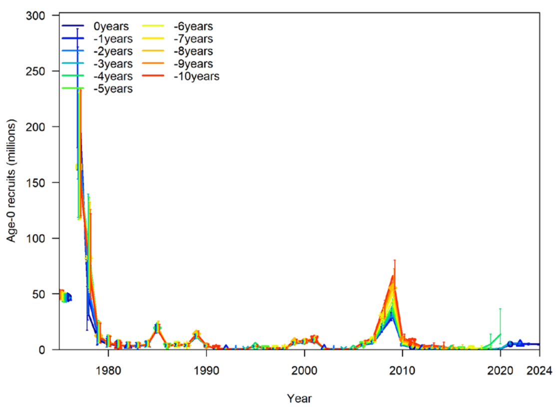
a)



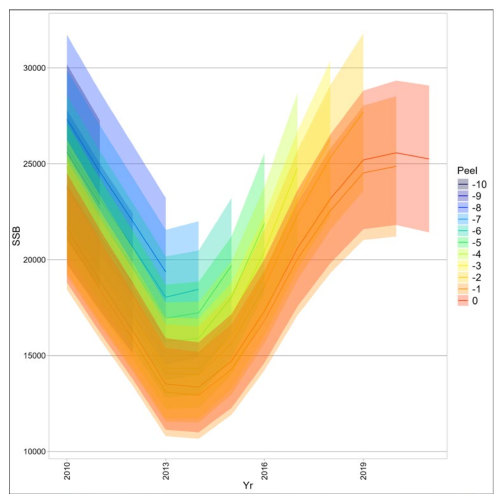
b)



c)



d)



e)

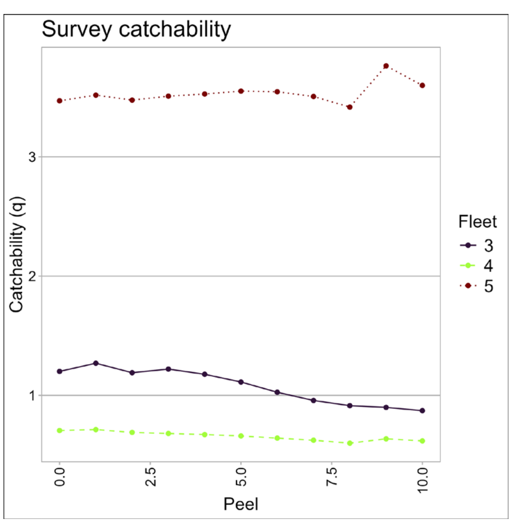


Figure 40. Retrospective analysis of a) spawning stock biomass, b) age-0 recruits, c) fishing mortality, d) zoomed SSB, and e) survey catchability for model 25.5.

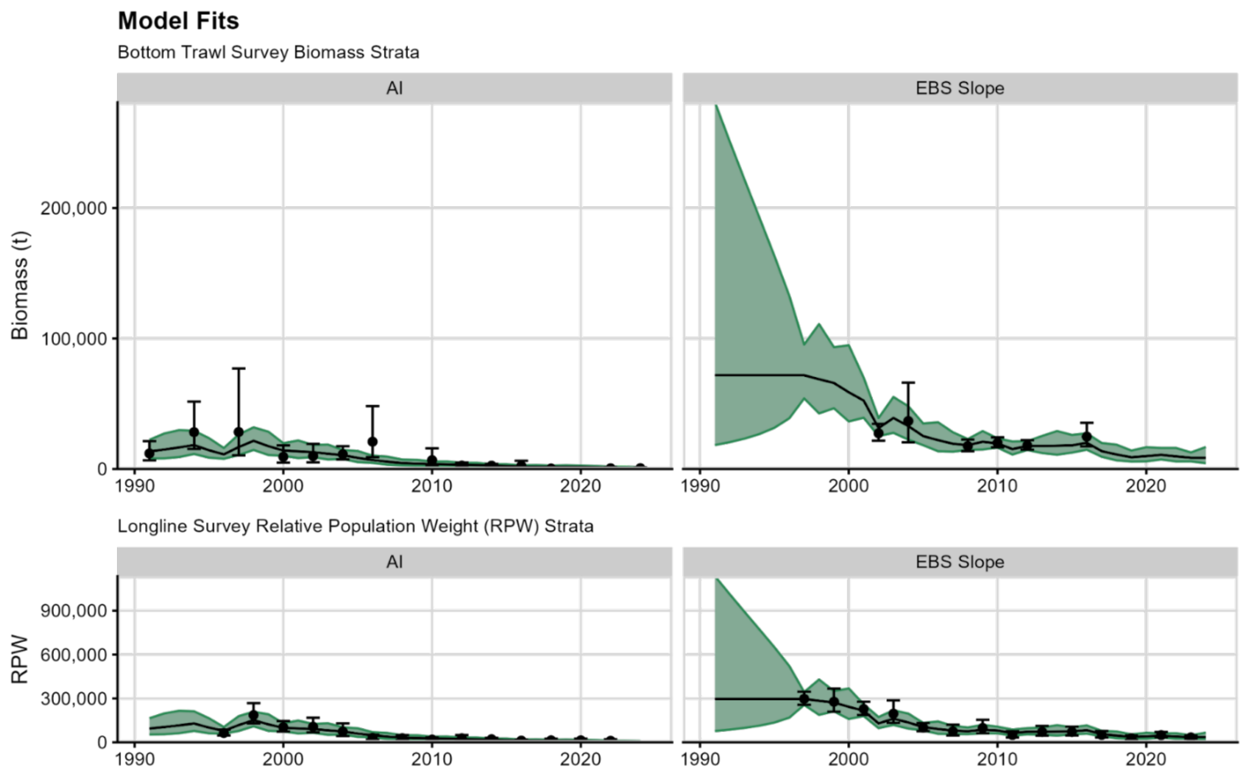


Figure 41. REMA model fits to bottom trawl survey biomass data (top row) and longline relative population weight indices (bottom row) stratified by area (AI and EBS slope).

Appendix A. Francis reweighting of model 25.3

Table A.1. Input sample size and adjusted sample size after iterative re-weighting for the shelf and slope surveys. The slope input sample size for 1979, 1981, 1982, 1985, 1988, and 1991 was set equal to 25.

	Shelf Survey		Slope Survey	
Year	ISS	Adjusted	ISS	Adjusted
1987	90	1.8		
1988	104	2.08		
1989	111	2.22		
1990	187	3.74		
1991	194	3.88		
1992	184	3.68		
1993	276	5.52		
1994	191	3.82		
1995	153	3.06		
1996	236	4.72		
1997	72	1.44		
1998	209	4.18		
1999	71	1.42		
2000	128	2.56		
2001	141	2.82		
2002	146	2.92	444	17.76
2003	276	5.52		
2004	267	5.34	373	14.92
2005	217	4.34		
2006	209	4.18	493	19.72
2007	213	4.26		
2008	163	3.26	492	19.68
2009	291	5.82		
2010	757	15.14		
2011	1218	24.36		
2012	709	14.18	488	19.52
2013	438	8.76		
2014	372	7.44		
2015	339	6.78		
2016	240	4.8	505	20.2
2017	190	3.8		
2018	150	3		
2019	95	1.9		
2021	67	1.34		
2022	54	1.08		
2023	63	1.26		
2024	49	0.98		

Table A.2. RMSE values by fleet and model.

	Models		
Fleets	16.4c	25.3_qfloat3.21	25.3_qfloat3.21 Francis_ReWgt
AFSC_LONGLINE	0.437	0.367	0.222
SHELF	0.306	0.338	0.326
SLOPE	0.188	0.228	0.206

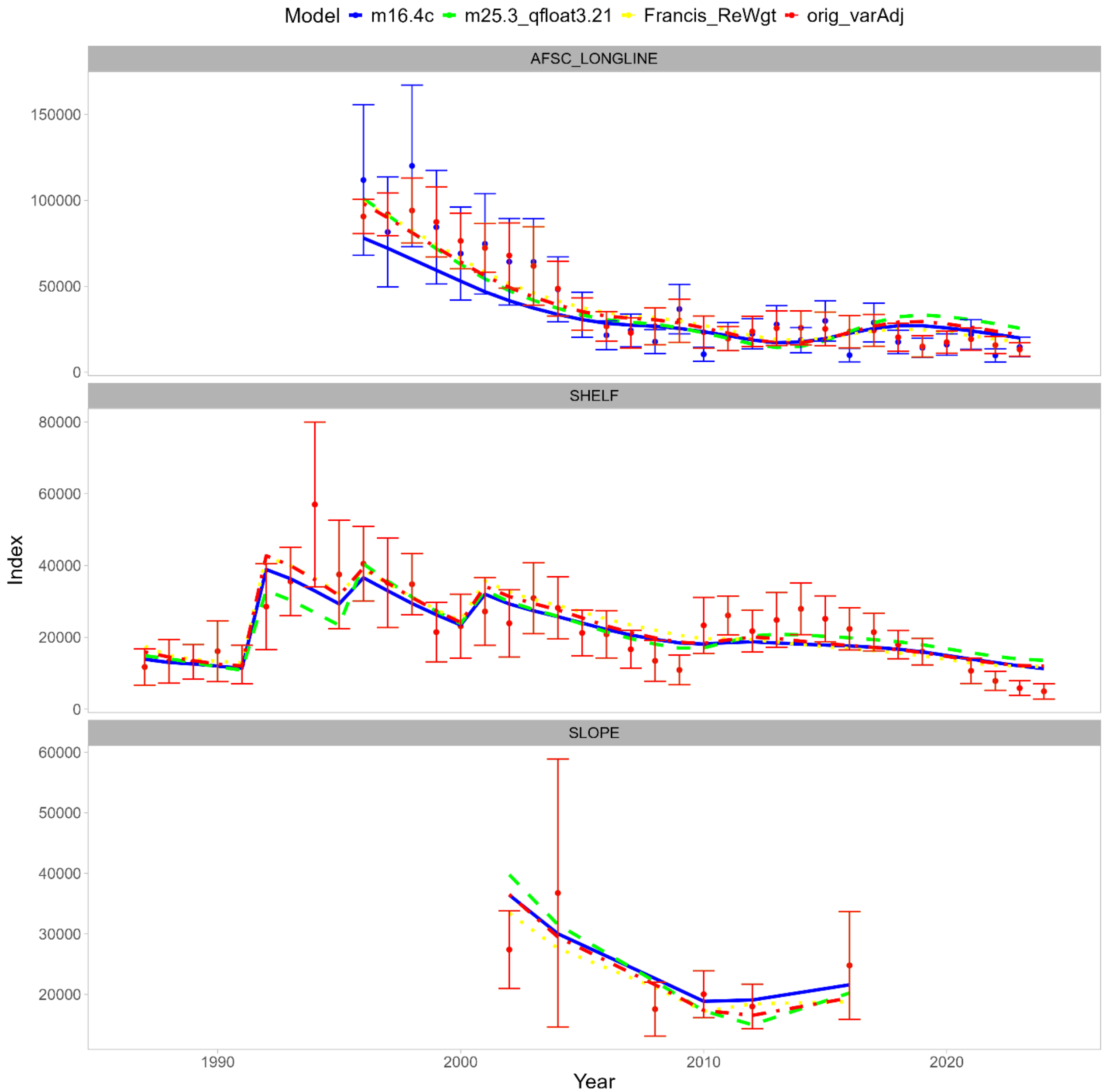


Figure A.1. Fits to the longline survey RPNs and the bottom trawl survey biomass.

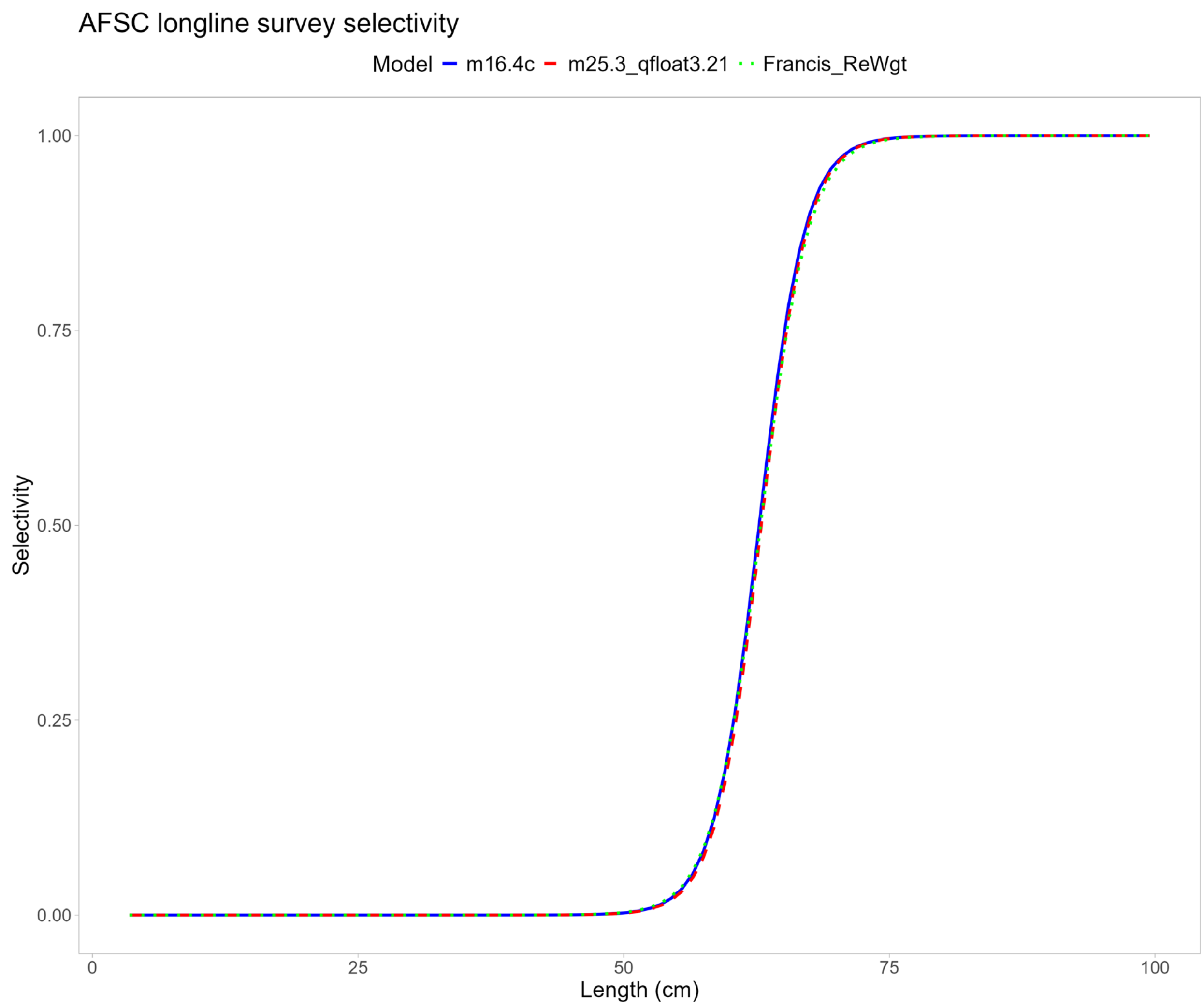


Figure A.2. Fits to the longline survey length composition data.

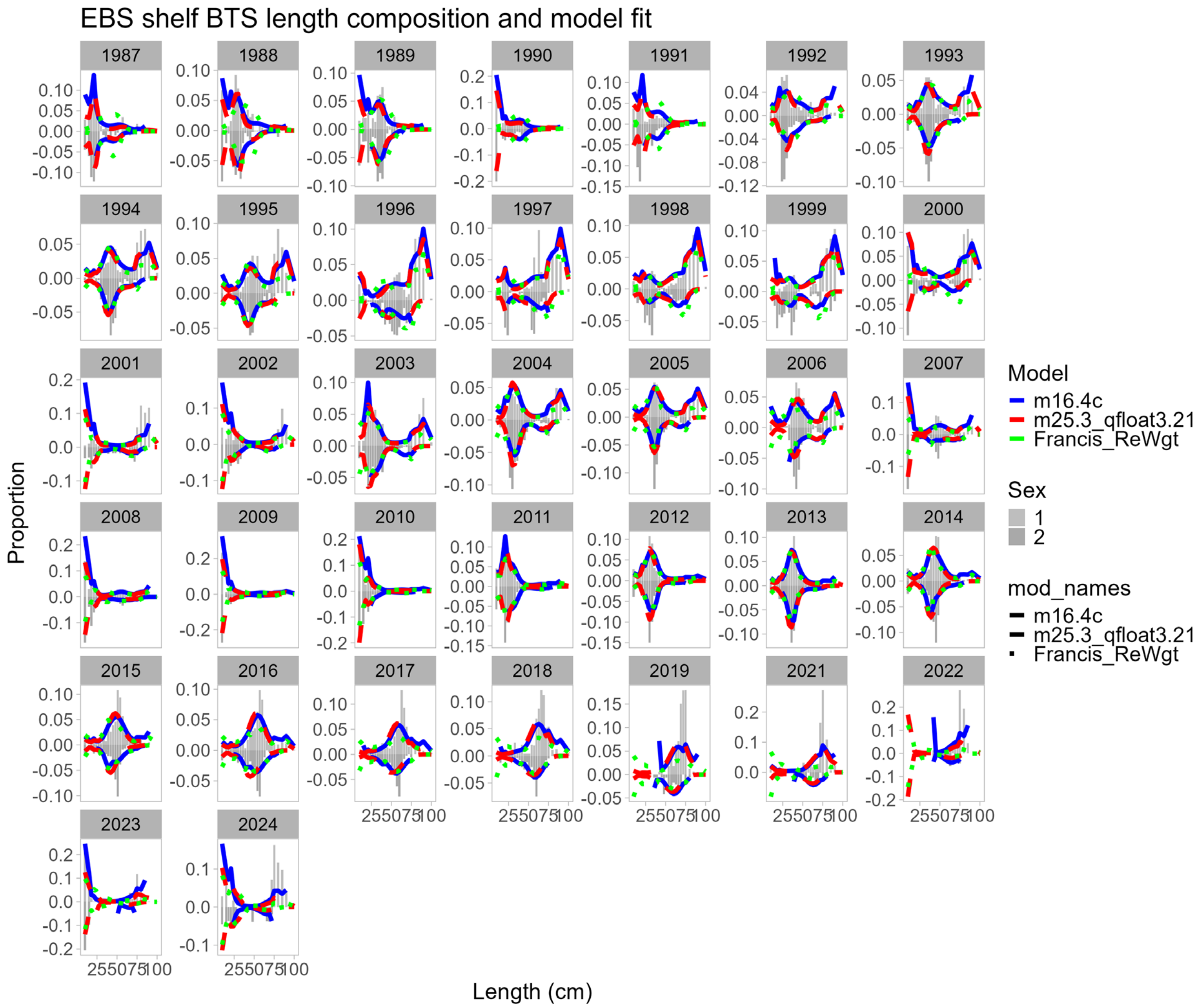


Figure A.3. Fits to the shelf survey length composition data.

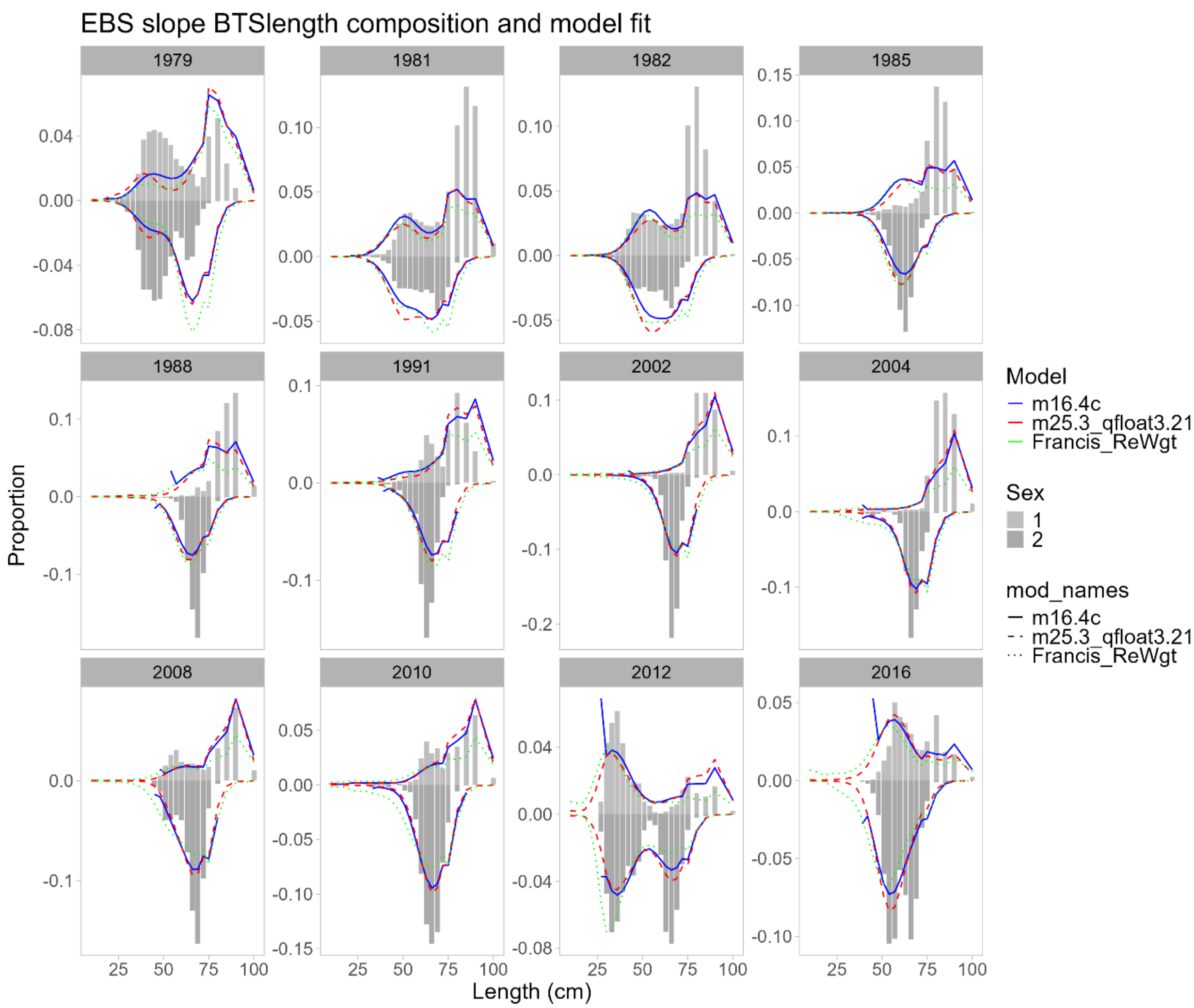


Figure A.4. Fits to the slope length composition data.

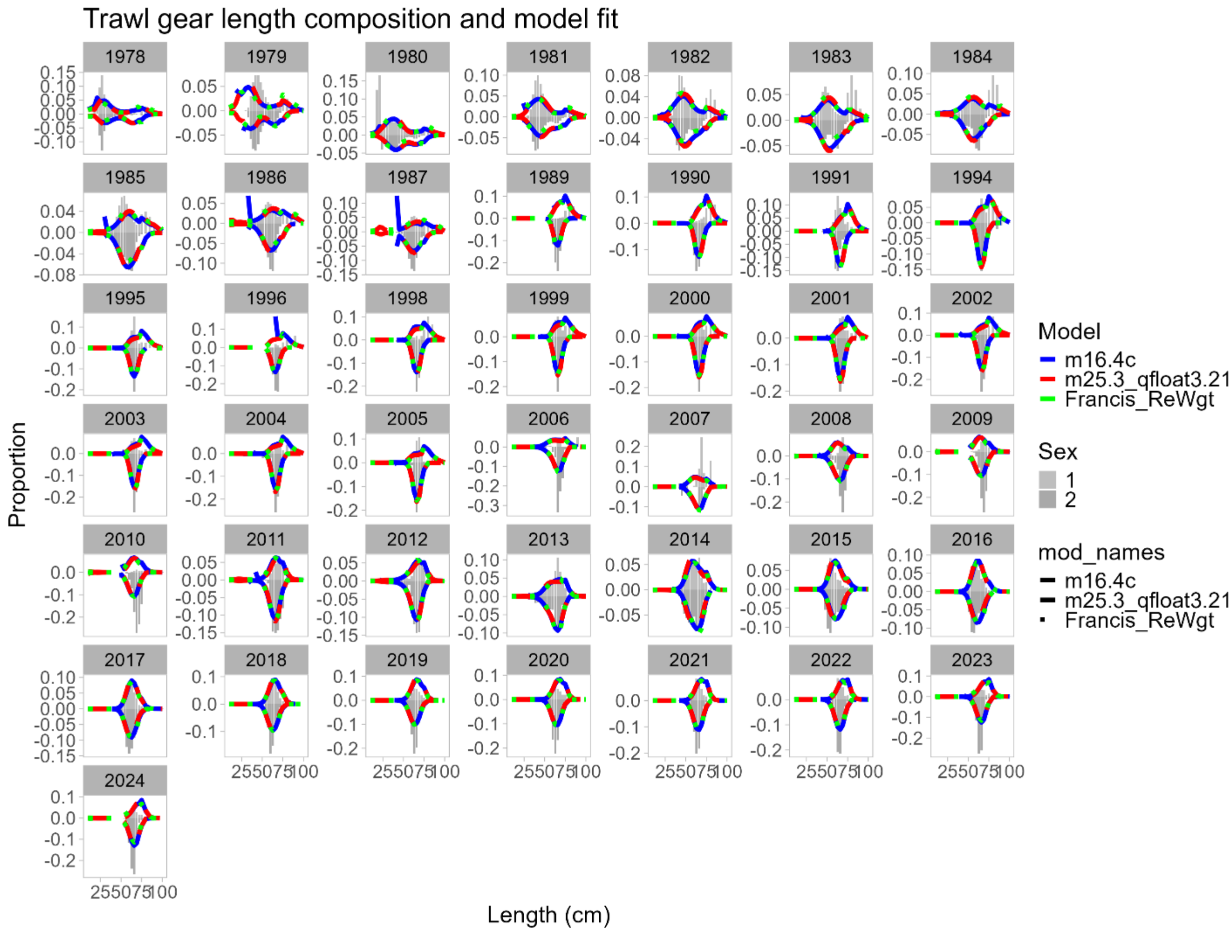


Figure A.5. Fits to the trawl fishery length composition data.

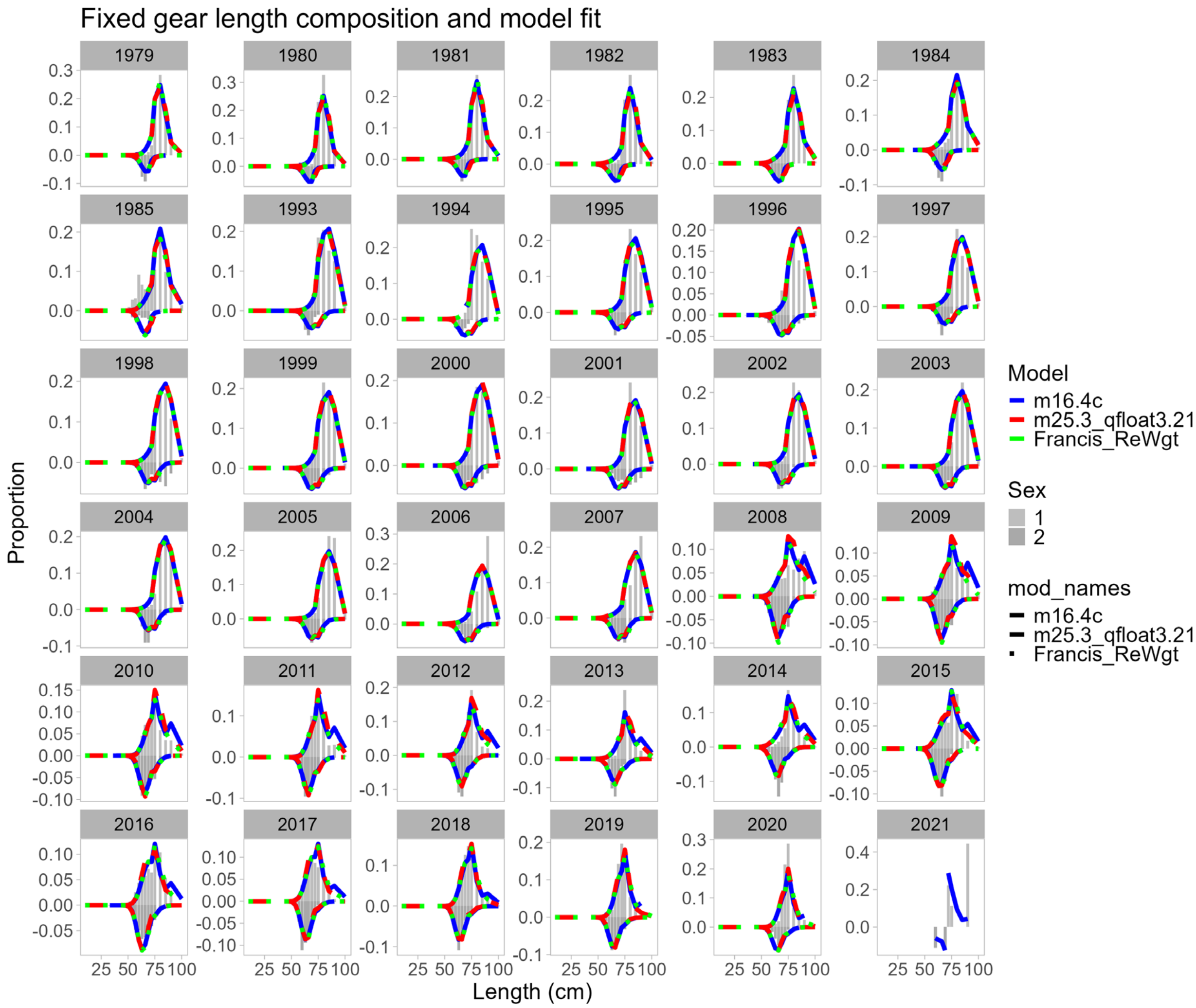


Figure A.6. Fits to the longline fishery length composition data.

Appendix B. Sensitivity runs using model 25.3

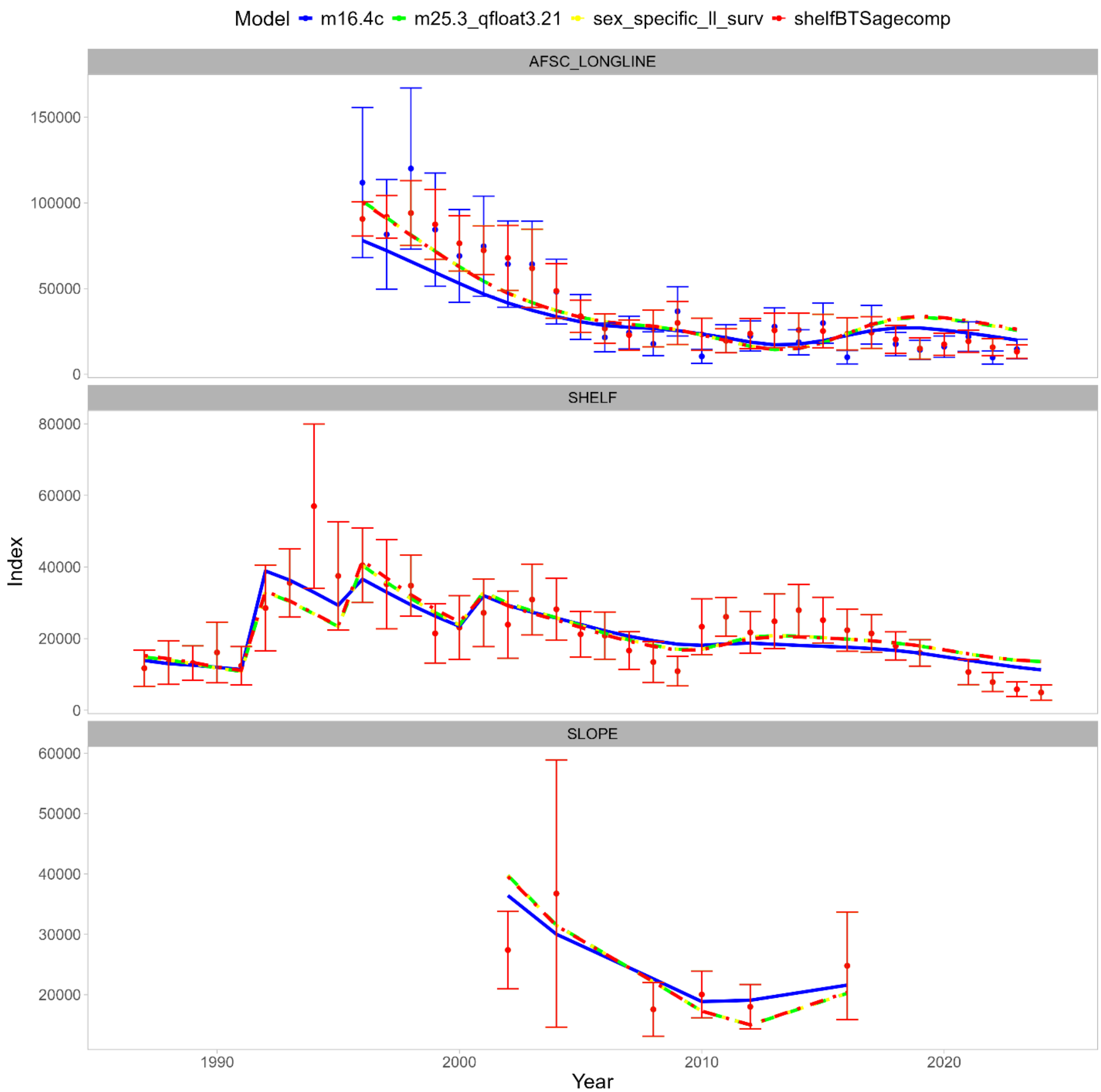


Figure B. 1. Fits to the longline survey RPNs and the EBS shelf and slope bottom trawl survey biomass from model 25.3 sensitivities including longline survey sex-specific length composition data and shelf survey age composition data.

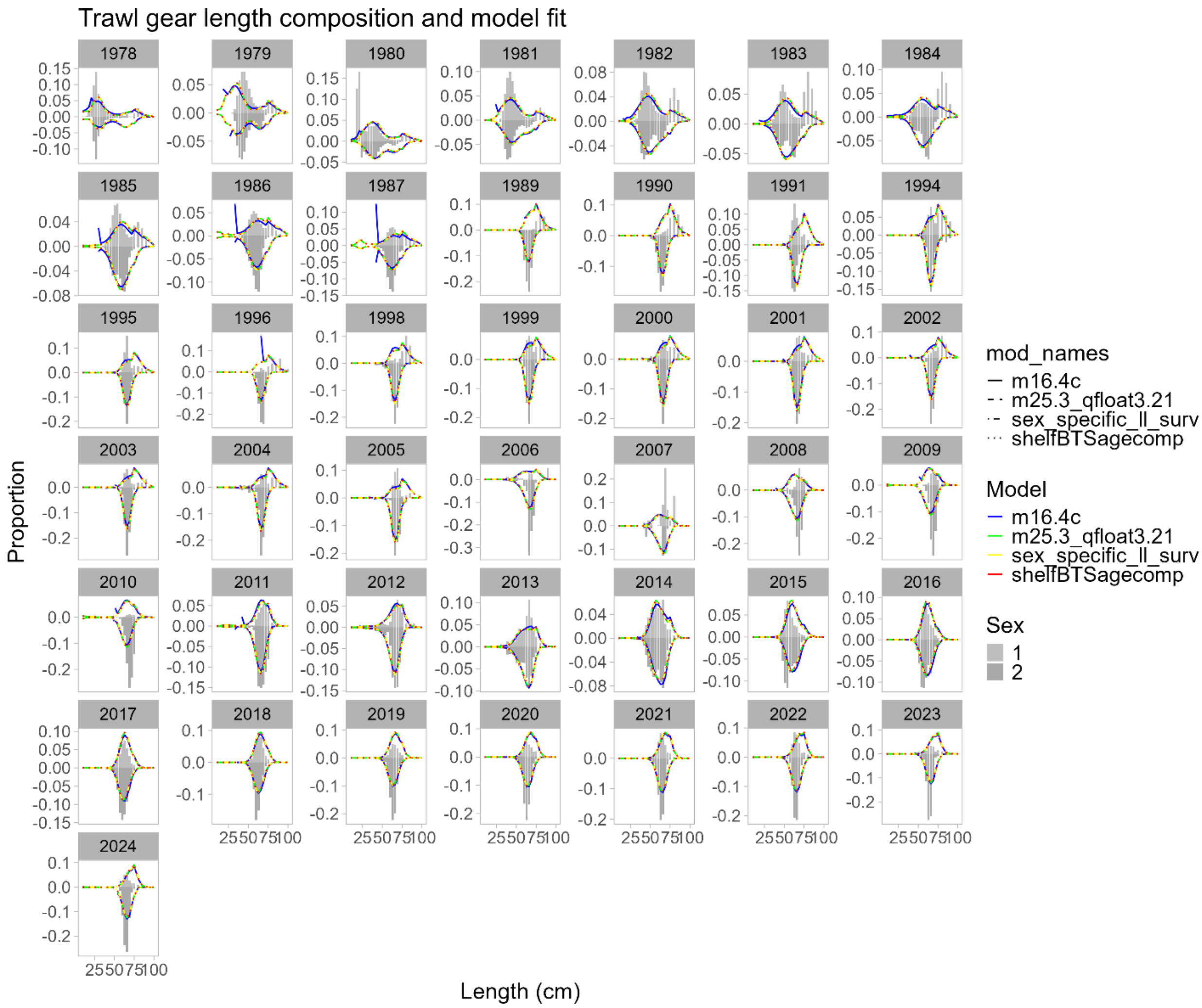


Figure B. 2. Fits to the trawl gear fishery length composition data for sensitivity runs.

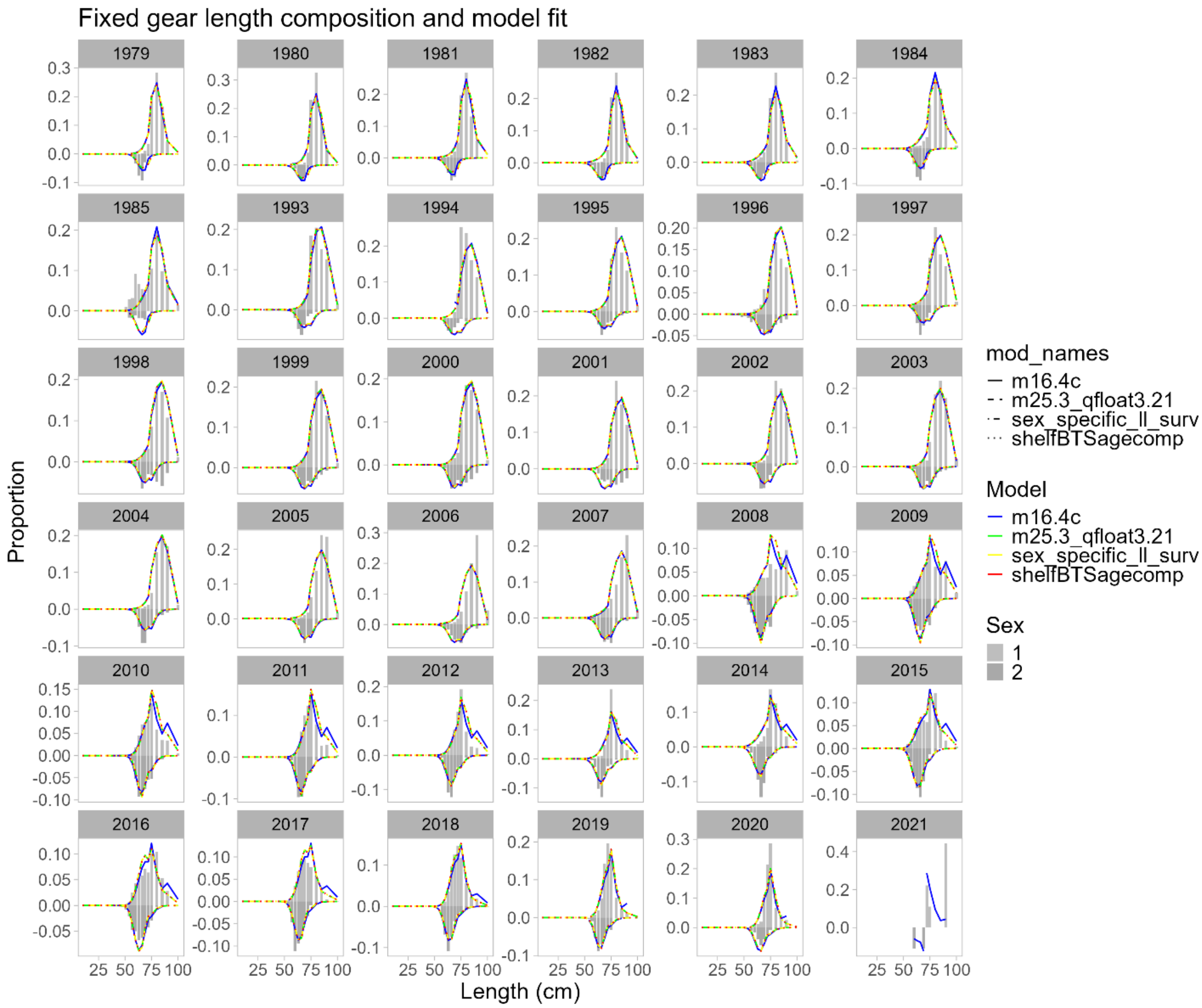


Figure B. 3. Fits to the fixed gear fishery length composition data for sensitivity runs.

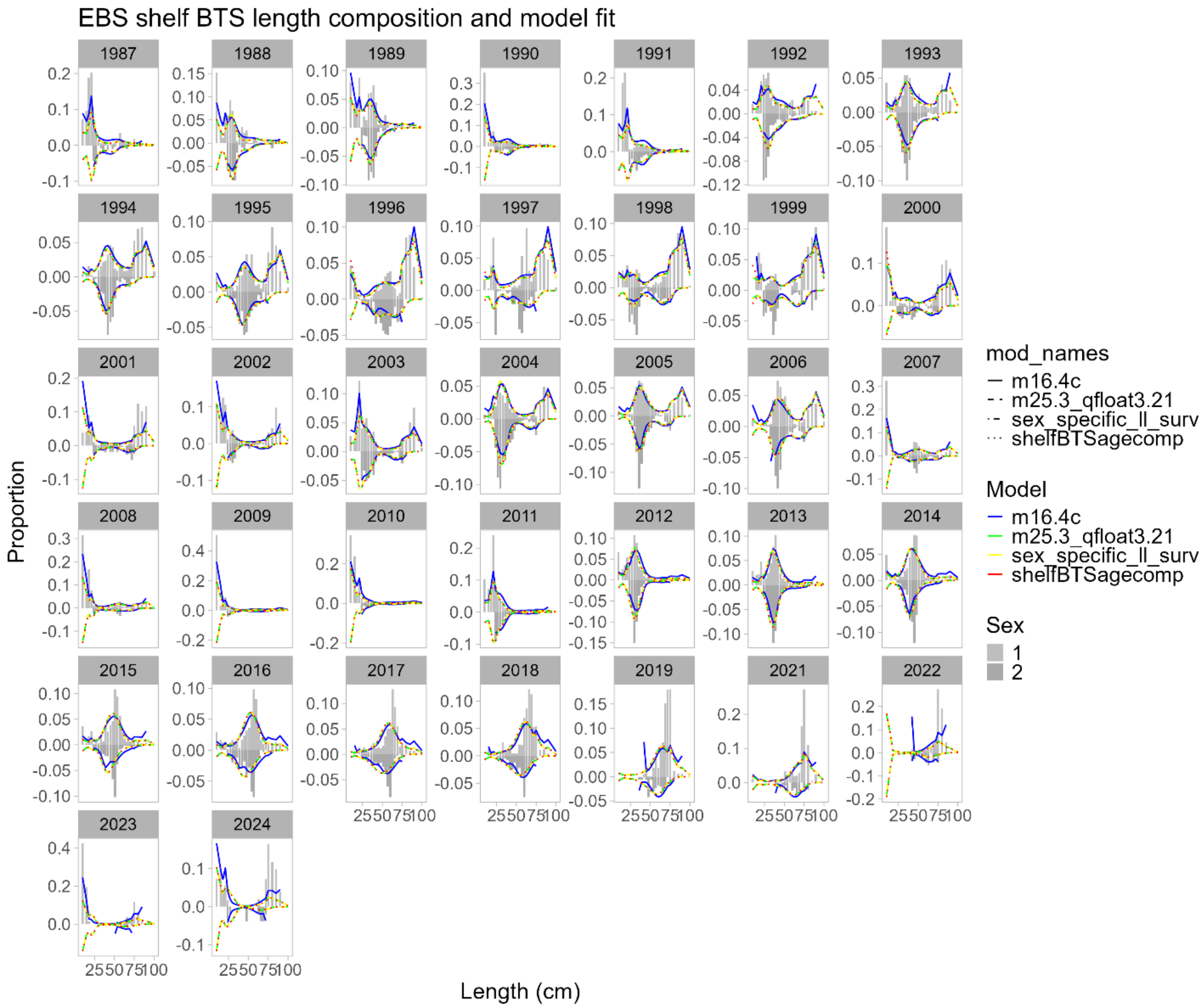


Figure B. 4. Fits to the EBS shelf bottom trawl survey length composition data for sensitivity runs.

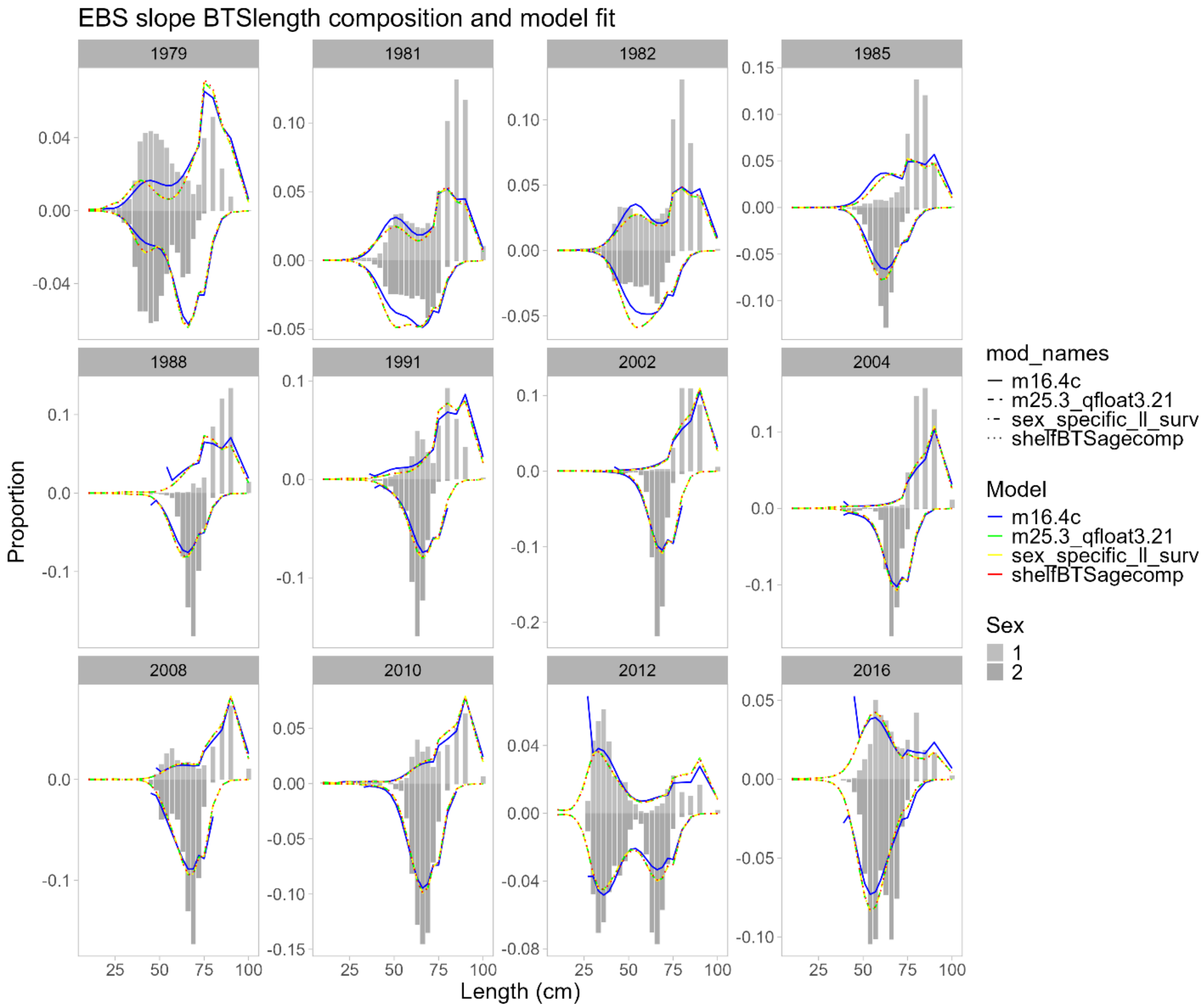


Figure B. 5. Fits to the EBS slope bottom trawl survey length composition data for sensitivity runs.

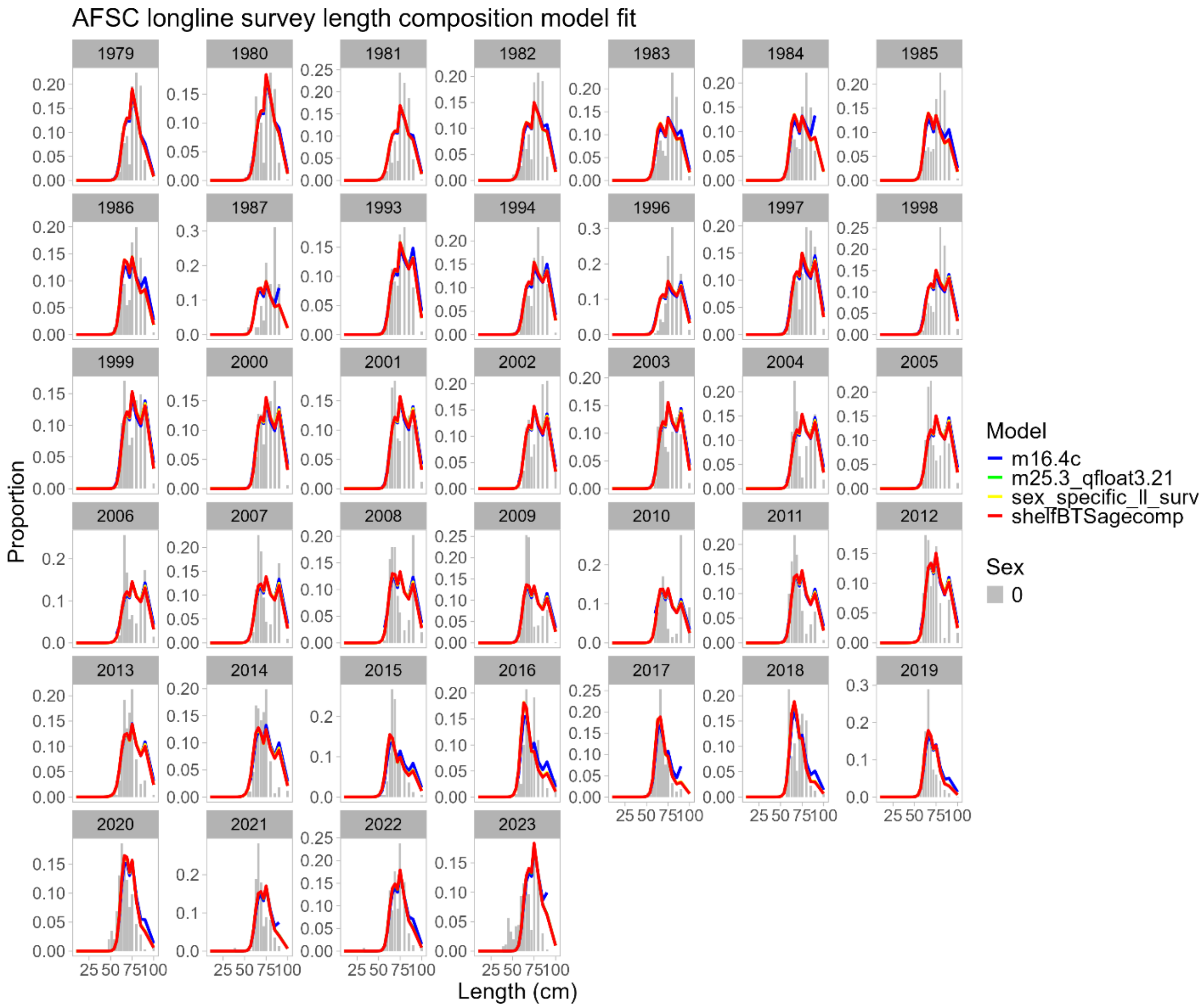


Figure B. 6. Fits to the longline survey length composition data for sensitivity runs for the combined sex data.

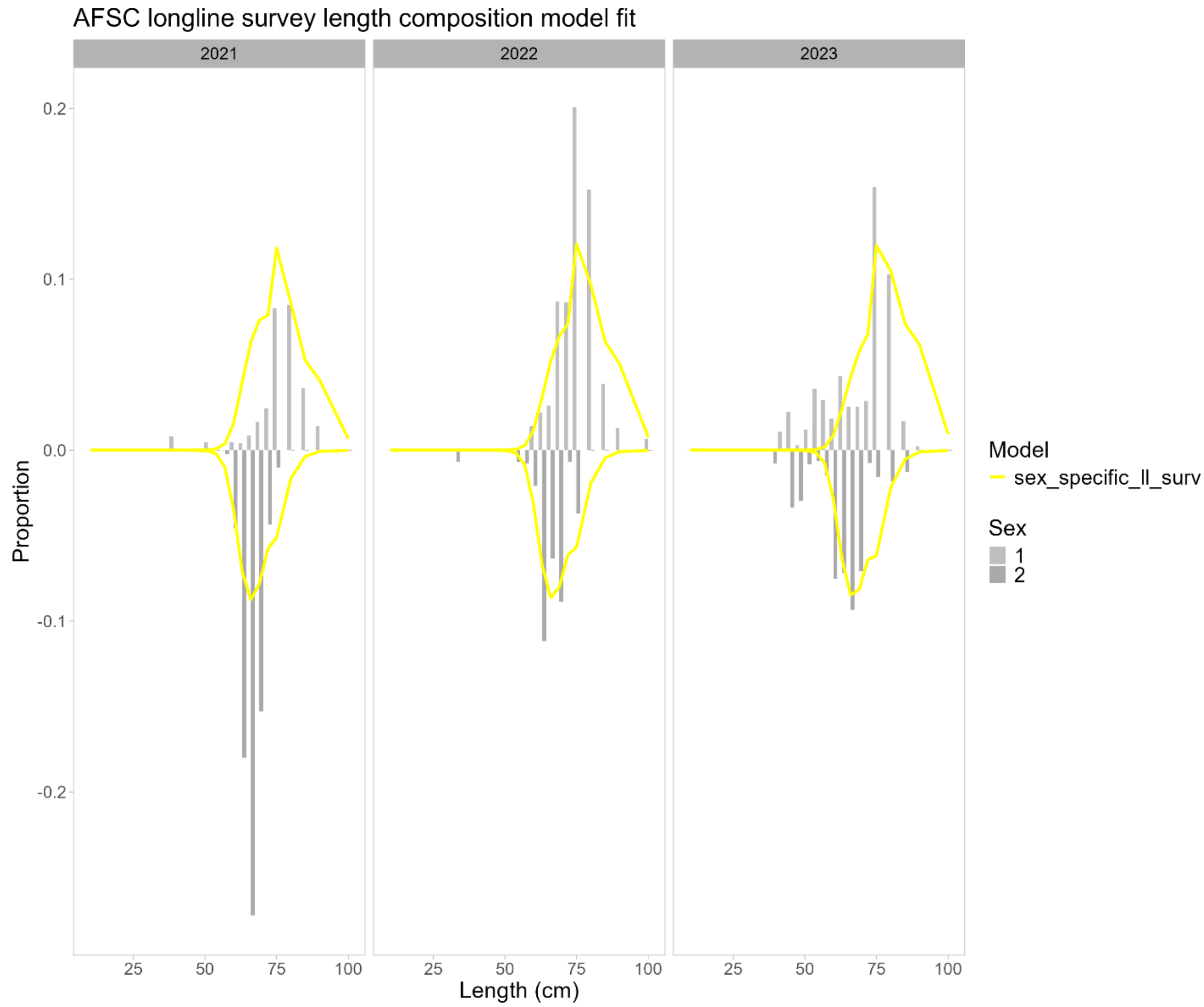


Figure B. 7. Fits to the longline survey length composition data for sensitivity runs for the sex-specific data.

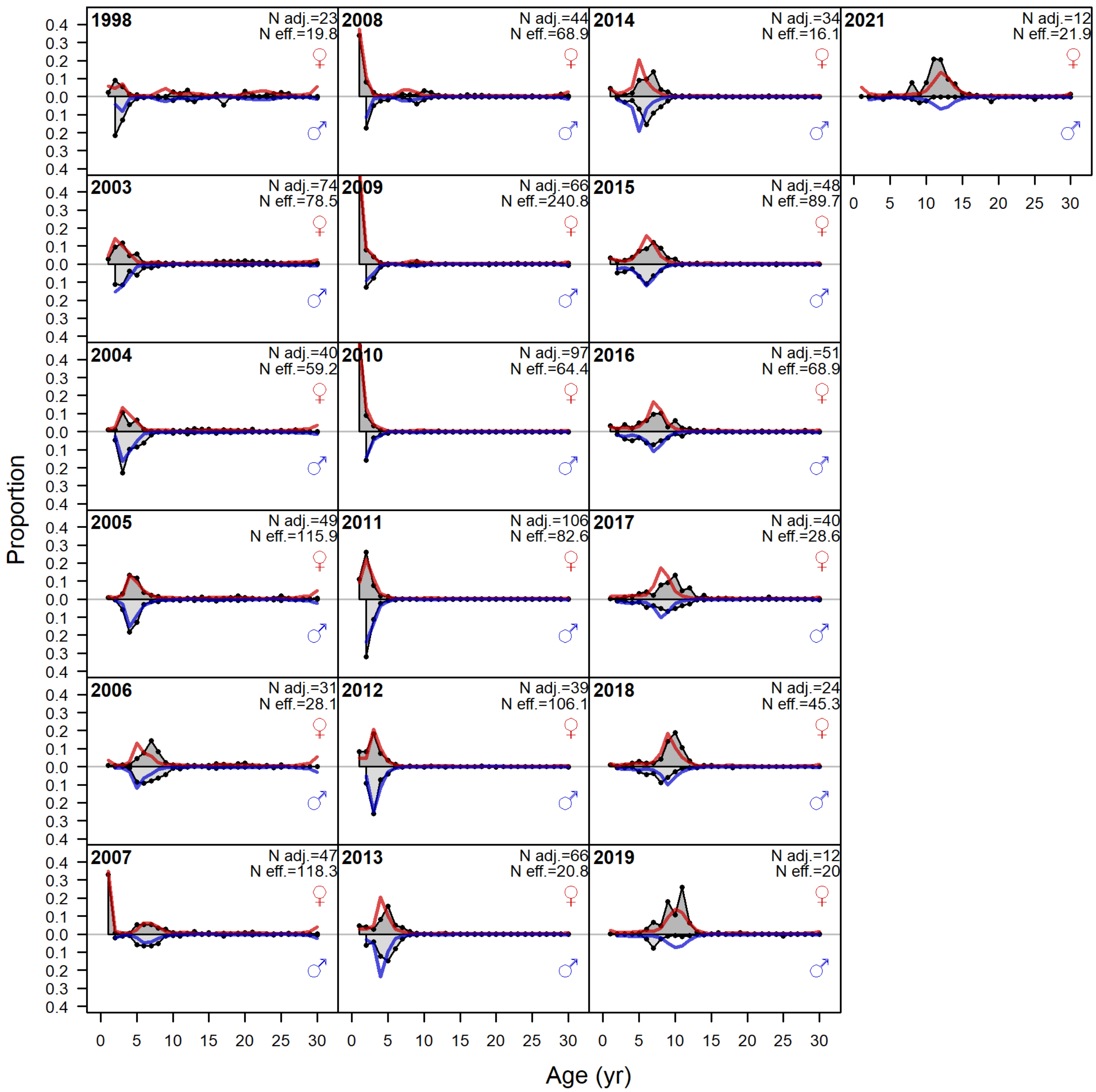


Figure B. 8. EBS shelf bottom trawl survey age composition data and the model fit to the data.

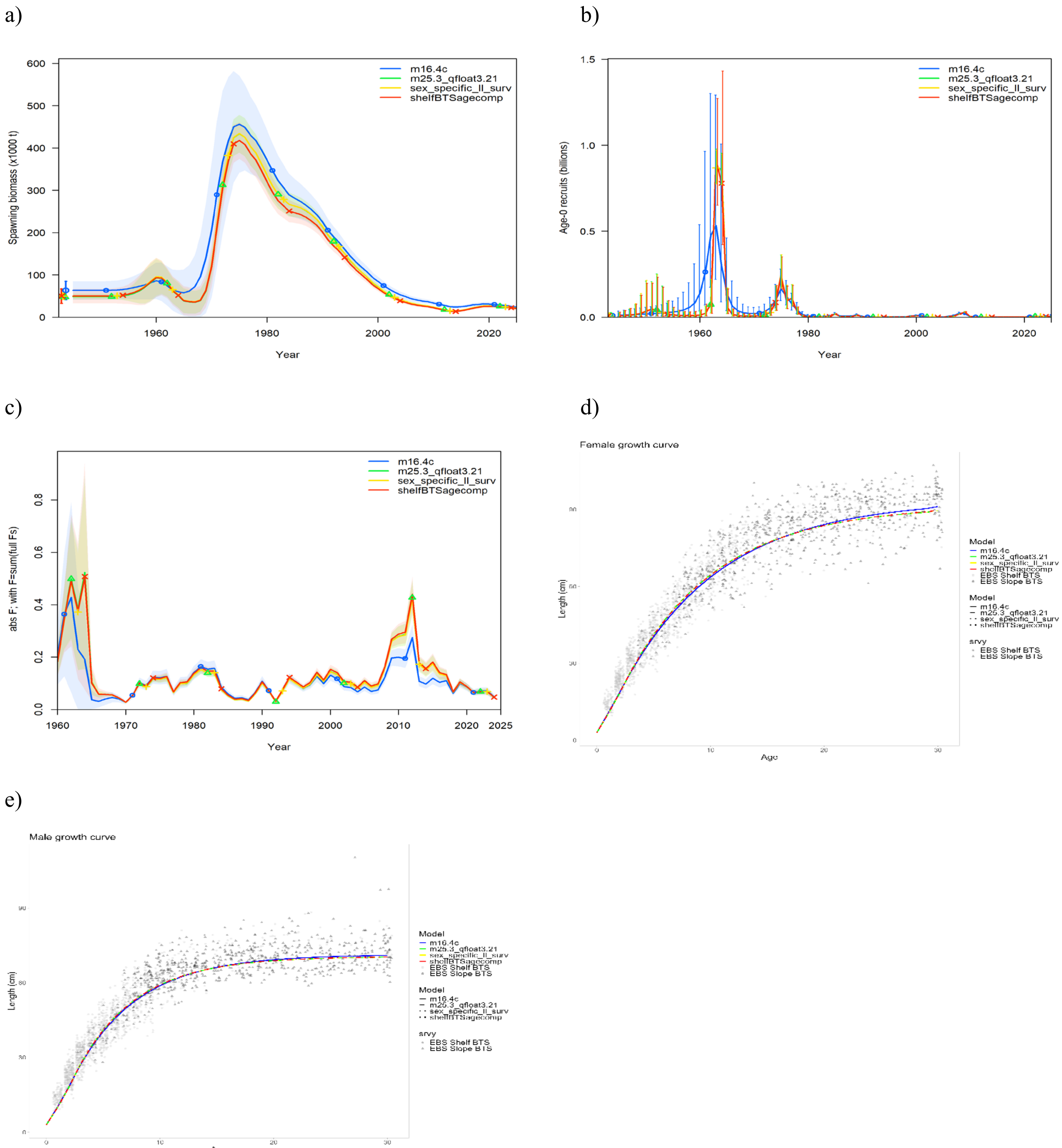


Figure B. 9. a) Spawning stock biomass, b) recruitment, c) fishing mortality, d) female growth curve, and e) male growth curve from model 25.3 sensitivities including longline survey sex-specific length composition data and shelf survey age composition data.