

Assessment Model Development and Stock Projections for Cape Cod-Gulf of Maine Yellowtail Flounder stock

Working Paper for ToRs 4-6

Larry Alade¹

¹Northeast Fisheries Science Center, Woods Hole, MA

1.0 Abstract

A state-space assessment model (WHAM) was developed for the Cape Cod-Gulf of Maine (CCGOM) yellowtail flounder stock over the period 1985-2022, using aggregate data from commercial landings and discards, as well as four fishery-independent surveys: the Northeast Fisheries Science Center (NEFSC) Spring and Fall Bottom Trawl Surveys, and two inshore state surveys (Massachusetts Department of Marine Fisheries (MADMF) Fall Inshore Bottom Trawl Survey and the Maine-New Hampshire (MENH) Fall Inshore Trawl Survey). Overall, the CCGOM yellowtail flounder spawning stock biomass has varied overtime, with recent years (2020-2022) showing low recruitment levels, but some recovery in SSB compared to historical lows. Projections indicate that maintaining fishing pressure at the FMSY proxy (F40%) would lead to increases in catch while gradually reducing SSB toward the SSB proxy reference point.

2.0 Introduction

The yellowtail flounder stock in the CCGOM region (Figure 2.0.1) was previously assessed using a Virtual Population Analysis (VPA). CCGOM yellowtail flounder was last assessed in 2022 during a management track assessment. The stock exhibited significant retrospective patterns, highlighting the limitations of the VPA model, particularly in its ability to effectively handle data variability and uncertainty. In an effort to address these issues, we developed an age-structured model for CCGOM yellowtail flounder within the WHAM (Woods Hole Assessment Model) framework.

WHAM is a state-space stock assessment model with a structure similar to the ASAP (Age-Structured Assessment Program) model. Like ASAP, WHAM fits aggregate indices and catch data along with their associated age compositions. However, WHAM offers enhanced flexibility by incorporating process error in several ways: during the transition of

abundance from age a in year y to age $a+1$ in year $y+1$ (representing survival), in natural mortality (M), stock-recruitment relationships, selectivity, and catchability (q). These process errors can be linked to environmental covariates (e.g., for natural mortality or recruitment), modeled as independent and identically distributed (iid) errors, or as auto-regressive processes (AR1) with correlations either across ages/parameters or years, or in a two-dimensional form (2DAR1) across both.

Random effects in WHAM help capture variability in survival (e.g., natural mortality) but can also account for factors like immigration/emigration, misreported catches, or misspecifications in selectivity. Aggregate indices and catch data in WHAM are assumed to follow a lognormal error distribution, with a fixed input standard deviation, although there is an option to estimate a scalar for observation error.

WHAM provides six different modeling options for age composition data: multinomial, Dirichlet-multinomial, Dirichlet, Multivariate Tweedie, logistic normal with zero values treated as missing, and logistic normal which pools zero values. The latter three options are considered ‘self-weighting’ because they estimate a dispersion parameter that adjusts the input effective sample size. Furthermore, the logistic normal and Dirichlet distributions can handle both positive and negative correlations, a more realistic reflection of observed data correlations (Francis, 2014).

WHAM can be accessed through the [NOAA Fisheries Integrated Toolbox](#) or from its [GitHub repository](#), where a suite of vignettes showcases the model's capabilities and provides practical examples for users. This shift to WHAM for CCGOM yellowtail flounder assessment represents a significant step towards a more robust and flexible modeling framework, capable of addressing retrospective issues and allow for explorations of environmental processes. In this working paper, model inputs, model selection process, and model results are summarized. Additionally, biological reference point determination and projection methodologies are discussed.

3.0 Data

3.1 Fishery Data

The CCGOM yellowtail flounder fishery historically supported substantial catches, though over time there has been a general decline, especially after 2000 (Figure 3.1.1). Discards account for only a small fraction of the total catch of the CCGOM yellowtail flounder stock, with the majority of catches attributed to the bottom trawl gear (Figure 3.1.2 and Figure 3.1.3).

During the late 1980s and early 1990s, catches were relatively stable, followed by a significant peak in 1990. After this peak, catches declined sharply throughout the mid-1990s. Although there was a modest recovery in the late 1990s into the early 2000s, the fishery saw a steep and prolonged decline in total catch from the mid-2000s onward. Since the early 2000s, catch levels have remained significantly reduced, with consistently low levels recorded through the 2010s and by 2020. A slight increase was noted in 2021, but overall, the stock remains at historically low levels.

In the WHAM framework, both commercial landings and discards are combined into a single fleet (Figure 3.1.1). Historically, the fishery targeted younger fish (ages 2-4), but recent trends show a shift toward older age classes (ages 4-6; Figure 3.1.4). For more detailed data on landings, discards, and fleet dynamics, refer to the Yellowtail Flounder Research Track Working Group Report ToR 2 Chapter.

3.2 Survey Data

In the development of the WHAM for the CCGOM yellowtail flounder stock, six fishery-independent surveys were initially considered based on their inclusion in previous assessments. These surveys included the NEFSC spring and fall bottom trawl surveys, along with the inshore MADMF spring and fall, and inshore MENH spring and fall surveys. These datasets were thoroughly examined during model development.

It is important to highlight that in 2009, the NEFSC transitioned from using the Albatross to the Bigelow vessel for bottom trawl surveys which impacted in not only a vessel change, but in the net and tow time (Additional details can be found in the Table 3.1.1 of the assessment report). To account for potential differences between vessels, length-based calibration factors were developed and applied to the Bigelow time series (Miller et al. 2010; Miller et al. 2013). Given the substantial length of the Bigelow dataset (over a decade), the Working Group also explored the possibility of using the Bigelow time series independently for both spring and fall, thus expanding the number of survey indices from six to eight for certain model configurations.

Ultimately, for the CCGOM candidate model, the decision was made to use the uncalibrated NEFSC survey indices. A comprehensive description of each of the surveys can be found in the Yellowtail Flounder Research Track Working Group Report ToR 3 Chapter.

4.0 Assessment Development

This Term of Reference (ToR) was addressed by updating the existing model (Virtual Population Assessment, VPA) to incorporate new data treatments and assess their impact on population vital rates (i.e. Spawning Stock Biomass (SSB) and fishing mortality (F)). A bridge was then established between the VPA and a statistical catch-at-age model (Age Structured Assessment Program, ASAP; Legault and Restrepo 1998), serving as a transition step toward developing a state-space model (Woods Hole Assessment Model, WHAM; Stock and Miller 2020) as the proposed base model for the CCGOM yellowtail flounder stock.

4.1 Virtual population Analysis (VPA)

The VPA software (ADAPT, Gavaris 1988), can be obtained from the [NOAA Fisheries Integrated Toolbox](#). The currently accepted VPA model for CCGOM yellowtail fits to total catch at age (assumed to be known without error) and is tuned to 29 age-specific indices of abundance: NEFSC-Spring for ages 1-6+; NEFSC-Fall for ages 1-5; MADMF-Spring for ages 1-6+, MADMF-Fall for ages 1-5, MENH-Spring for ages 2-5 and MENH-Fall for ages 2-4. The VPA assumes that indices occur at the beginning of the year, and therefore “ages” in the NEFSC-Fall survey are relabeled such that age a in year y is input as age $a+1$ in year $y+1$. The model is tuned by including the observed index in year $T+1$, where year T is the terminal year of catch data. The objective function is the sum of squared log scale residuals for all 29 indices, with equal weight given to each index.

The last management track assessment was conducted in 2022 with catch data through 2021 and indices through 2021 for the fall indices and 2022 for the spring indices (for T+1 calculation). The VPA was not updated with additional years of data, and all comparisons were made with the VPA from the 2022 management track assessment.

The Working Group implemented several key data adjustments to enhance the inputs for the Virtual Population Analysis (VPA). First, survey indices from the NEFSC and MADMF were updated from absolute swept-area biomass estimates to relative indices, a shift that better accommodates survey variability and improves the consistency of relative abundance estimates over time. Additionally, natural mortality (M) rates were updated from 0.2 to 0.4 based on recent life history studies, providing a more up to date representation of the species' mortality dynamics (for details, see the Yellowtail Flounder Research Track Working Group Report ToR 1 chapter). Maturity ogives were also refined; instead of using a time-series average, annual observed proportions at age were applied, allowing for temporal variation in age-based maturity. Furthermore, updates to the length-weight relationship enabled revisions to catch-at-age data from 1994 to 2022,

aligning growth assumptions with current data and providing a more accurate reflection of stock composition.

VPA analyses were conducted to assess the impact of different data revisions on the CCGOM yellowtail flounder stock, comparing each run to the 2022 management track (MT2022). The “New index” run, which updated survey indices from absolute biomass to relative measures, showed only minor changes in both F and SSB relative to MT2022. The “M04” run, which increased natural mortality from 0.2 to 0.4, generally resulted in lower F values and higher SSB estimates, indicating that a higher natural mortality rate reduces the estimated fishing pressure. The “Mat” run, incorporating annualized maturity, produced modest increases in SSB, particularly in recent years, without significantly altering F trends. The “Combined” run, which incorporated all data revisions, resulted in the highest SSB levels and the lowest F values across the time series, highlighting the cumulative impact of these adjustments (Figures 4.1.1 and 4.1.2).

4.2 Building the bridge from VPA to WHAM

Building a bridge from VPA to WHAM, a brief preliminary setup was first conducted in an Age-Structured Assessment Program, ASAP (Legault and Restrepo 1998) to facilitate the seamless importation of model data inputs into the WHAM framework. The Working Group (WG) determined that the primary focus should be directed towards developing a robust WHAM model rather than expending additional resources refining the ASAP model. ASAP was used as an intermediate tool, providing a streamlined transition into the WHAM modeling environment. This approach allows for efficient leveraging of data while maximizing compatibility with WHAM’s advanced features, such as flexible random effects structures and improved handling of environmental covariates.

4.3 ASAP “Pit Stop”

The Age Structured Assessment Program, ASAP (Legault and Restrepo 1998) software, available through the NOAA Fisheries Integrated Toolbox, is an age-structured population model that estimates stock sizes using observed catch, catch-at-age data, and abundance indices. It assumes fishing mortality is separable into year and age components but allows flexibility with fleet-specific computations and selectivity changes over blocks of years. The model fits data using a negative log-likelihood objective function, incorporating multinomial error for age composition data and lognormal error for other components such as total catch, survey indices, and recruitment deviations. Further details are available in the technical manual (Legault 2008). The treatment of surveys in ASAP and WHAM

differs from VPA; in ASAP and WHAM, the spring and fall surveys are retained in the year they were conducted and tuned to abundance levels adjusted by partial-year mortality corresponding to the survey timing. However, VPA generally assumes that each survey provides an index of abundance relative to the entire calendar year. This lack of seasonal adjustment can lead to potential biases if mortality fluctuates significantly throughout the year.

To bridge build from VPA to ASAP for model comparison, we configured ASAP using data directly imported from VPA, modifying settings to approximate VPA's assumptions. For instance, a CV of 0.01 was fixed for aggregate catch, and the effective sample size for catch-at-age was set to 150 to force ASAP to nearly match the VPA assumption that catch is known without error. Indices were input similarly, with a consistent CV of 0.3 applied across indices in an attempt to emulate the VPA treatment of indices to be equally weighted in the objective function. No stock-recruitment function was estimated, with steepness fixed at 1 and recruitment freely estimated.

Unlike VPA, ASAP uses a separability assumption for fishing mortality and selectivity, so a single selectivity block for all model years was fit, assuming full selection for ages 4-6 based on VPA results. The ASAP_BridgeRun1 model suffered from convergence problems. Given the poor model diagnostics in ASAP and the decision to prioritize the transition to WHAM, no further ASAP runs were conducted. Instead, subsequent model development focused entirely on WHAM, where additional explorations were carried out to refine the assessment framework for CCGOM yellowtail flounder to better address the complexities of the stock.

4.4. WHAM Model Explorations

4.4.1 Initial Parameterization (aka - Fixed Effect Formulation) - Run0

Due to WHAM's ability to emulate ASAP, the initial setup was a non-state space ASAP-like model. This setup included the number of age classes (1-6+), starting values for effective sample sizes (150), and CV on the combined fleet landings (0.01). The only structural difference with this exploration is the way that recruitment is estimated. In ASAP, a mean recruitment was estimated with annual deviations from that mean that are controlled by user-specified fixed CV on those deviations. In WHAM configured as ASAP, recruitment in each year is estimated as a fixed parameter. In this case, the model passed both first-order convergence criteria (with a maximum gradient of 1.45e-11) and second-order convergence criteria, as the Hessian matrix was invertible. Following this initial setup, over 800 model variations were tested and compared, exploring different assumptions regarding fleet and survey selectivities, recruitment, age compositions, life history

parameters, and environmental influences. These decisions, detailed below, were not made in a linear sequence; rather, earlier choices were frequently revisited and adjusted as new insights emerged from subsequent model changes.

4.4.2 NEFSC survey (“To split or Not to split”) – Run 1

In 2009, the NEFSC bottom trawl survey underwent a significant transition, shifting from the Albatross to the Bigelow as the primary vessel, along with changes to the survey protocols (Miller et al. 2010). This prompted a thorough examination of whether to split the NEFSC survey time series in 2009 to account for potential shifts in catchability and data comparability. A comparison of model runs using the calibrated NEFSC survey indices versus those with a split time series revealed that the model with split indices produced superior diagnostics, including better fits and reduced retrospective bias.

Further, the Working Group (WG) identified some inconsistencies with the existing NEFSC bottom trawl survey strata definitions. Specifically, strata 57 and 62 had not been sampled previously, while strata 58 and 63, which had recorded catches of yellowtail flounder, were excluded in the previous definition. After reviewing this, the WG determined that strata 57 and 62 should be excluded, and strata 58 and 63 should be incorporated into the survey design (Figures 4.4.2.1 and 4.4.2.2). A quick exploratory model was run to assess the impact of this revised strata configuration. Results showed negligible differences compared to the previous strata definition, suggesting the changes had minimal influence on model outcomes. Consequently, the WG decided to adopt the model configuration that used split NEFSC survey indices for both the fall and spring, along with the revised strata definitions.

4.4.3 Fishery Selectivity - Run 5

The initial fleet selectivity was modeled using a logistic function, with selectivity increasing asymptotically across ages 1-6+, assuming near-zero selectivity at age-1 and full selectivity at older ages. Age-specific selectivity models were also evaluated, with full selectivity fixed at age-4 and all other ages freely estimated. To capture variability in fishery selectivity, several random effect structures were tested in WHAM, including 1) independent and identically distributed (iid) random effects, 2) autoregressive (ar1) effects between logistic parameters and ar1_a effects on ages for age-specific formulations 3) autoregressive processes across years (ar1y), and 4) two-dimensional autoregressive processes across both parameters and years (2dar1).

The model using logistic selectivity with 2dar1 random effects, in combination with NEFSC-uncalibrated survey indices, produced the best overall fit and retrospective patterns compared to both the base model without random effects and alternative model formulations using the NEFSC uncalibrated survey indices. Consequently, logistic fleet selectivity with 2dar1 random effects was carried forward as the candidate model configuration (Figures: 4.4.3.1 and 4.4.3.2).

4.4.4 Survey Selectivity - Run 5

In the initial model, each of the six surveys—NEFSC fall and spring, as well as MADMF and MENH fall and spring inshore surveys—were modeled with both logistic and age-specific selectivities, similar to what was done for the combined fleet. No random effects were applied to the survey selectivities, reflecting the consistency in survey design and methods. The Working Group (WG) also determined that introducing random effects could undermine the reliability of the surveys as stable indices of relative abundance.

A comparison between age-specific and logistic functional forms were evaluated for each of the surveys. The logistic selectivity models resulted in better overall fits and improved retrospective patterns (Mohn's rho) across most of the surveys. For the spring surveys, selectivity showed a gradual increase, starting near zero for age-1 and reaching full selectivity by ages 4-6+. In contrast, the fall surveys exhibited more rapid increases in selectivity, with age-1 starting at a slightly higher selection and achieving full selectivity by ages 2-6+. This pattern was incorporated into the candidate model as it provided a better representation of the observed data trends across the different seasons and survey platforms (Figure 4.4.4.1).

4.4.5 Numbers-at-age - Run 5 vs Run 14 vs Run 71

In the WHAM framework, incorporating random effects into numbers-at-age (NAA) was a key part of improving model performance. Various configurations of random effects were tested, including independent and identically distributed (iid), autoregressive processes across years (ar1_y), ages (ar1_a), and across both (2dar1). To ensure accurate recruitment dynamics, recruitment was decoupled from NAA, allowing age-1 fish to change independently from older ages (2-6+).

Out of the 12 tested configurations, only a subset met the convergence criteria necessary for further evaluation. Among these, the autoregressive across ages (ar1_a) model yielded the best overall results, demonstrating the lowest AIC, stronger model fit, and improved retrospective performance (Mohn's rho). Given these diagnostics, the ar1_a formulation

was selected for the candidate model. It should be noted that despite the improvement in Mohn's rho, retrospective runs showed emerging issues with scaling in the retrospective peels, which warranted further investigation in subsequent runs to ensure robustness and consistency across retrospective periods (Figure 4.4.5.1).

Additionally, further runs were conducted to explore the influence of random effects on both fleet selectivity and numbers-at-age. The candidate model from this expanded combinatorial approach showed further improvement in model diagnostics when assuming iid random effects on NAA, while maintaining the 2dar1 random effect on fleet selectivity. It should be noted that with the iid random effect on numbers-at-age, the need to decouple recruitment from older ages was no longer necessary because the iid random effect assumes that each age class has independent variation from others.

4.4.6 Recruitment assumption and Environmental Covariates - Run 87 and Ecov Runs

Prior to testing the effects of environmental covariates on recruitment, preliminary analyses were conducted to determine the most suitable recruitment assumption for the candidate model. Recruitment was modeled either as random about the mean or using a Beverton-Holt stock-recruitment relationship with iid process error. While the Beverton-Holt model passed initial convergence criteria, model diagnostics, including Mohn's rho and residual patterns, deteriorated significantly. Additionally, the model performed poorly in fitting the stock-recruitment relationship.

Subsequently, environmental covariates such as the Atlantic Multidecadal Oscillation (AMO) and spring bottom temperature were tested for their potential effects on recruitment. These covariates were modeled as either a random walk or an autoregressive process, under the assumption that recruitment would be influenced by environmental conditions from the previous year. However, none of the environmental covariate models led to improved residuals or retrospective Mohn's rho statistics.

The model assuming recruitment as random about the mean without environmental covariates yielded the lowest AIC, further supporting its selection as the candidate model. This formulation provided the most consistent fit and better overall model diagnostics. Consequently, the final candidate model excluded environmental covariates and assumed an autoregressive (iid) process for recruitment.

4.4.7 Natural Mortality and Environmental Covariates

Natural mortality (M) was modeled as a constant across ages and years, with a fixed value of $M = 0.4$, derived from longevity estimates (see Yellowtail Flounder Research Track Working Group Report ToR 1). Various configurations were explored to refine this assumption, including random effects models such as autoregressive processes across years (ar1_y) and both ages and years (2dar1). However, these alternatives produced worse model fits and retrospective patterns. Additionally, age-specific M , derived from weight-at-age, was tested alongside these random effects but yielded no significant improvement over the base assumption.

Following the guidance from the ToR 1 recommendations for CCGOM yellowtail flounder, environmental covariates such as the Atlantic Multidecadal Oscillation (AMO) and spring bottom temperature were incorporated to assess their potential influence on natural mortality. These covariates were tested using both linear and polynomial models, with random walk and autoregressive (ar1) processes considered for each. Despite thorough exploration, none of the environmental covariate models led to improved model performance in terms of fit or retrospective diagnostics.

Given the inferior results from the environmental covariate models and the tested random effects, the decision was made to retain the $M = 0.4$ assumption for the candidate model.

4.4.8 Fishery and Survey Age Composition - Run 101

In WHAM, the default age composition structure is multinomial, consistent with the setup used in ASAP. However, in this analysis, alternative age composition structures—such as logistic-normal—were also tested for both the combined fishery and eight survey indices. A key advantage of the logistic-normal distribution is its flexibility in handling zero observations. Specifically, it can either treat zeros as missing data (miss0) or pool data from surrounding age classes to generate values for those zeros (pool0). Additionally, the logistic-normal-miss0 structure can incorporate an autoregressive process (ar1) to account for correlations between neighboring age classes, thereby smoothing transitions and providing more realistic age compositions (logistic-normal-ar1-miss0).

In total, 27 combinations were tested, covering three age composition types (multinomial, logistic-normal-miss0, logistic-normal-pool0) across nine data blocks (one combined fleet and eight survey indices). The multivariate Tweedie and Dirichlet-multinomial structures were not explored due to the high computational demands of the Tweedie model and the

subjectivity involved in specifying effective sample sizes in the Dirichlet-Multinomial approach (Thorson and Punt, 2014; Stock and Miller, 2021).

The logistic-normal-miss0 configuration resulted in most reasonable fits to the data, and Mohn's Rho retrospective. Incorporating the ar1 process further improved diagnostics, particularly by capturing smoother transitions between age classes, except in the case of the MADMF fall survey, where the simpler logistic-normal-miss0 performed better. As a result, the candidate model adopted the logistic-normal-ar1-miss0 structure for the combined fleet and most surveys, with the exception of the MADMF fall survey, which retained the logistic-normal-miss0 configuration.

4.4.9 Survey Inclusion/Exclusion - Run 115-369 (Run 304 or Run 406)

As mentioned previously, the candidate model up to this point of model development exhibited scaling issues in the retrospective peels, signaling challenges in the model's ability to resolve population scale. This raised the hypothesis that the scaling of retrospective patterns could be potentially linked to choice of survey inclusion in the model, as each survey's geographic and temporal coverage provides distinct views of population abundance. Eight surveys were initially considered, including uncalibrated NEFSC spring and fall bottom trawl surveys (pre- and post-2009) and state inshore surveys from MADMF and MENH.

To investigate this, 255 model runs were conducted with various survey configurations by systematically excluding up to seven surveys (from a total of eight, due to the NEFSC Bigelow and Albatross split) to identify which surveys might contribute to retrospective scaling inconsistencies.

Dropping the spring state surveys (MADMF and MENH) resulted in significant improvements to the retrospective peels (Figures 4.4.9.1), particularly in reducing the magnitude of retrospective bias while maintaining acceptable model fit and Mohn's Rho diagnostics. Thus, the decision to exclude these surveys was based on a refined hypothesis that the spatial footprint and timing of the spring surveys may have overlapped with spawning and migration, affecting fish availability to the state surveys, thus potentially resulting in distorting the scale of the population.

4.4.10 Fleet Selectivity – Revisited

While model 304 initially seemed reasonable, further evaluation of reference points revealed unusual results that warranted closer scrutiny. Specifically, the model produced unrealistically high F_{MSY} proxy values that exceeded the fishing mortality levels observed throughout the history of the fishery. While the calculations were mathematically correct, the unusually high F_{MSY} raised concerns. A deeper examination of the inputs to the reference points—selectivity, weights-at-age, maturity, and natural mortality (M)—suggested that the issue likely stemmed from the fishery selectivity specification.

In model 304, the estimated fishery selectivity showed a declining trend for younger fish, particularly for ages 4 and 5, which have historically been fully selected. By contrast, only the 6+ age group was fully selected in recent years. This pattern partially explains the abnormally high F reference points, implying that the actual fishing pressure on the younger fish was much lower than the FMSY proxy due to the reduced selectivity patterns at these ages.

It was initially suspected that the variation in selectivity might be driven by the 2dar1 random effects used in the model's selectivity specification. Consequently, a reevaluation of fleet selectivity was conducted, focusing on testing different blocking and random effect combinations. Specifically, selectivity blocks were introduced in 1994 (when changes in reporting regulations were enacted) and in 2010 (when sector management was implemented). Various random effects were explored, including independent and identically distributed (iid), autoregressive across ages (ar1), autoregressive across years (ar1_y), and 2dar1, as well as turning off the random effects entirely.

Each selectivity configuration was tested with and without the lognormal adjustment in process and observation errors. Of the 364 factorial model runs conducted, turning off random effects on fleet selectivity and imposing a two-block selectivity structure starting in 1994 resulted in a more improved model diagnostics over model 304. Specifically, this configuration led to better retrospective Mohn's rho statistics, reduced bias in the self-test, and improved stability from the jitter analysis. As a result, model 452 became the new base model, offering superior diagnostics over model 304, particularly in retrospective patterns, bias assessments, and estimated model CVs, while maintaining similar performance in fitting the indices and catch data.

4.4.11 Candidate Model Setup (Run 452)

The final model formulation, referred to as “Run 452,” and found in Table 4.4.11.1 incorporates a two-block logistic fleet selectivity, with blocks spanning 1985-1993 and 1994-2022, and no random effects applied to the fleet selectivity. This model integrates six key fishery-independent surveys: the NEFSC spring and fall bottom trawl surveys (both uncalibrated), split between the Albatross series (1985-2008) and Bigelow series (2009-2022), as well as the inshore MADMF and MENH fall bottom trawl surveys covering their respective time frames. A logistic functional form is applied consistently across these survey indices without random effects. The fleet age composition follows a logistic-normal-miss0 distribution, while the surveys are modeled using a logistic-normal-ar1-miss0 structure. Additionally, the numbers-at-age (NAA) include autoregressive random effects (ar1_a), and recruitment is not decoupled from ages 2-6+, maintaining a connection between recruitment and these older age classes.

5.0 Model Results

5.1 Candidate Model Diagnostics

The candidate model demonstrated strong performance, passing both first-order convergence criteria (with a maximum gradient of 2.58e-10) and second-order convergence criteria, as the Hessian matrix was invertible. A jitter analysis further confirmed model stability, with the model converging on a global solution at a 100% rate (Figure 5.1.1), indicating robustness and reliability in the optimization process.

Given the relatively low coefficient of variation (CV) assumed for the aggregate fleet (commercial landings and discards) at 0.05, the model exhibited a very good fit to the data, showing minimal patterning (Figure 5.1.2). The model’s fit to the NEFSC Spring, NEFSC Fall, MADMF Fall, and MENH Fall bottom trawl surveys was similarly strong, although some residual patterning persisted throughout the time series (Figures 5.1.3–5.1.8). Notably, there were only a few outlier years across the eight surveys where model estimates fell outside the confidence bounds.

The OSA residual diagnostics for the aggregate fleet and the six survey indices (NEFSC Spring, NEFSC Fall, MADMF Fall, MENH Fall), were generally well-behaved and displayed approximate normality in their distribution (Figures 5.1.9–5.1.15). While some tailing (i.e., instances where residual patterns at the distribution tails fell outside the normality cone) was observed, there was no evidence of significant or consistent patterning over time. These results suggest a good model fit overall. Furthermore, OSA residual diagnostics for

the age composition data exhibited even more normally distributed residuals across the fleet and the survey indices (Figures 5.1.16–5.1.22). Bubble plots of OSA quantile residuals (Figures 5.1.23–5.1.29) provided further visual confirmation of the model's adequate fit and diagnostic robustness.

In terms of retrospective analysis, the model showed a slight bias in estimating spawning stock biomass (SSB), with Mohn's $\rho = -0.063$ (Figures 5.1.30, 5.1.31), indicating a minor underestimation of SSB. Conversely, there was a very slight tendency to overestimate fishing mortality (F) with Mohn's $\rho = 0.054$ (Figures 5.1.32, 5.1.33), indicating a minor overestimation of F . Additionally, while not a formal criterion, the retrospective analysis revealed a slight tendency to overestimate recruitment (Mohn's $\rho = 0.17$; Figures 5.1.34, 5.1.35). To further support the characterization of this minor retrospective bias, ρ -adjusted SSB and F values were compared to the 90% confidence intervals from the terminal year. All estimates were well within the confidence bound of the terminal year estimates.

The Akaike Information Criterion (AIC) was applied throughout the model selection process. However, it should be noted that AIC values, when presented out of context, are not inherently informative for model comparison. Despite this, the AIC value for the candidate model was -1982.2. This value, while useful in relative comparisons, must be interpreted in conjunction with other models, as discussed in the Assessment Model Development section.

To further evaluate model performance, self-tests were conducted, focusing on the percent bias for key population metrics. The results indicated a mean percent bias for fishing mortality (F), recruitment (R), and SSB of 9.71%, -8.7%, and -17.66%, respectively. The model once again showed strong stability, achieving a 100% convergence rate (Figure 5.1.36).

5.2 Candidate Model Estimates

Predictions from the candidate model show that fishing mortality (F) has significantly declined over time, starting from $F = 1.79$ in 1985 to $F = 0.06$ in 2022, representing a 97% decrease. Notable fluctuations occurred in the 1990s, with F peaking at $F = 2.39$ in 1997, followed by a steady decrease through the 2000s and 2010s (Figure 5.2.1).

Spawning Stock Biomass (SSB) has shown substantial variability across the time series. From $SSB = 1165$ mt in 1985, SSB increased to a peak of $SSB = 8559$ mt in 2017, marking an over 7-fold increase from its 1985 value. However, SSB reached its lowest point in 1987 at $SSB = 1092$ mt, followed by a gradual recovery, with the most recent value in 2022 recorded

at SSB = 8645 mt, representing almost an 8-fold increase compared to the 1987 low (Figure 5.2.1).

Recruitment (R) has also demonstrated considerable variability. In 1985, recruitment was R = 69,533 individuals, the highest value in the time series. This was followed by a sharp decline, reaching R = 6,796 individuals in 2001, an approximate 10-fold decrease. More recently, recruitment has fluctuated, increasing to R = 42,839 individuals in 2021 before declining to R = 12,692 individuals in 2022, a 70% decrease compared to the previous year (Figure 5.2.2).

The coefficient of variation (CV) for fishing mortality (F) from 1985 to 2022 remained below 0.3 for most of the time series, ranging from a low of 0.11 in 1991 to a peak of 0.38 in 2020, indicating increasing uncertainty in more recent years. The terminal year (2022) recorded a CV of 0.25, reflecting some reduction in uncertainty compared to 2020 but still above the earlier average (Figure 5.2.3).

In comparison, the CVs for Spawning Stock Biomass (SSB) were generally lower than those for F, staying within the range of 0.16 to 0.25 for much of the time series. However, notable increases in uncertainty were observed in the last few years, with SSB CV peaking at 0.36 in 2020 and remaining above 0.22 in 2022 (Figure 5.2.3).

Recruitment (R) showed the highest levels of uncertainty across the time series, with CV values fluctuating between 0.27 and 0.44. The terminal years (2020–2022) exhibited the most uncertainty, with CVs exceeding 0.40, peaking at 0.44 in 2020, and ending at 0.37 in 2022, indicating significant variability in recruitment estimates during this period (Figure 5.2.3).

Additionally presented are the estimated aggregate fleet and survey selectivities (Figure 5.2.4), as well as the estimated catchabilities of the surveys (Figures 5.2.5).

The Numbers-at-Age (NAA) time series from 1985 to 2022 has been consistently dominated by age-1 individuals, who represent the largest portion of the stock throughout the time series. For example, in 1985, age-1 individuals accounted for 69,533 out of a total of 83,049 fish, which is approximately 83.7% of the population that year. In 2022, the number of age-1 individuals dropped to 12,692, representing about 45.4% of the total population of 27,732 individuals. This reflects a decline in dominance by the youngest age group over time.

While age-1 continues to be the dominant age group, there has been a notable increase in the proportion of older age groups, particularly age-6+ individuals, in recent years. In 1985, age-6+ fish accounted for only 244 individuals, or 0.3% of the total population. By 2022, the

number of age-6+ individuals had grown to 3,460, representing approximately 12.5% of the total population, a nearly 14-fold increase in both numbers and proportion.

This shift is particularly evident in the last five years (2017–2022), where the contribution of age-6+ individuals steadily increased, reaching its highest in 2022. For perspective, the proportion of age-6+ fish has risen from 0.6% in 2017 to 12.5% in 2022. Meanwhile, age-5 individuals contributed 10.3% in 2022, while the age-4 cohort accounted for 26.9%.

This broadening of the age structure, with a higher proportion of older individuals surviving, suggests an important shift in population dynamics, with the stock becoming more evenly distributed across age groups in recent years (Figures 5.2.6 and 5.2.7).

This proportional increase in the plus group is clearly reflected in the SSB-at-age over time. Historically, the Spawning Stock Biomass (SSB) was primarily composed of age-2 and age-3 individuals. For example, in 1990, age-2 fish accounted for 35% of the total SSB, while age-3 fish made up 57%, together representing the majority of the biomass.

In contrast, during the last decade, there has been a notable shift in the age structure of SSB. By 2022, the proportion of older fish, particularly those age-6+, had increased substantially. Age-6+ individuals, which accounted for only 4.5% of the SSB in 1990, now represent approximately 13.6% of the total SSB. This marks a significant change in the population structure over time, indicating an increased contribution of older, more mature fish to the biomass.

Meanwhile, the proportion of age-2 and age-3 individuals has diminished somewhat. In 2022, age-2 and age-3 fish together comprised approximately 49.6% of the SSB, a reduction from their historical dominance (Figures 5.2.8 and 5.2.9).

6.0 Biological Reference Points

6.1 Fishing Mortality

A spawner-per-recruit analysis was conducted to determine the FMSY proxy reference point, F40%. For this calculation, an appropriate averaging period for weight-at-age (WAA) and maturity was examined, while natural mortality and fleet selectivity, being constant across the candidate model's time series, required no additional analysis. A moving-window analysis was performed on WAA and maturity (Figures 6.1.2, 6.1.3), where weight or maturity at a given age (e.g., age-5) in year y was predicted using an average from

years $y-1$ to $y-i$, with i varying from 2 to 10. Root mean squared error (RMSE) was calculated by summing residuals across all ages (1-6+) for each i value. This process was repeated to match "near-term" projections by predicting an average of years $y-3$ to y from an average of years $y-4$ to $y-(i+3)$. The two-year window yielded the lowest RMSE, suggesting that recent data (two years) are more predictive of near-term WAA trends, likely due to a declining trend in WAA for older ages over the past one or two decades and increased maturity for younger age fish (Figures 6.1.4, 6.1.5).

Based on this analysis, two-year averages of WAA and fleet selectivity were used to calculate the FMSY proxy reference point, F40%, while natural mortality and maturity remained constant over time, making year-averaging unnecessary. This F40% represents the fishing mortality rate that maintains 40% of the unfished spawning potential, calculated at 1.64 (Table 6.1.1). As noted earlier in Section 4.4.10, model m304 required extensive re-evaluation of fleet selectivity due to the resulting unrealistically high F40% estimates estimated at 3.0. While the F40% for the candidate model m452 is considered somewhat high, the Working Group (WG) found it reasonably acceptable for the research track compared to m304.

Several factors contribute to the high F40% reference point. Higher natural mortality and earlier maturity mean fish are removed by natural causes or reproduce at younger ages, enabling the population to endure increased fishing pressure while still preserving 40% of its spawning potential. This raises the F40% reference point, as the stock can withstand more fishing without significantly lowering SSB. To further explore these high reference points, an additional sensitivity analysis was conducted with natural mortality (M) adjusted from 0.4 to 0.2, which resulted in a notably lower F40% reference point of 0.47. This analysis highlights how assumptions regarding natural mortality could affect fishing reference points; higher M allows for greater fishing pressure at F40%, while lower M requires a more conservative F to achieve sustainability.

6.2 Spawning Stock Biomass

Based on the moving-window analysis results, SSB40% was calculated using the model's mean recruitment across 1985-2022, along with the two-year averages for maturity and WAA (natural mortality and fleet selectivity were constant and required no further analysis for this calculation). Due to the lack of a stock-recruit relationship, absence of strong trends in recruitment patterns and the precision of recruitment estimates based on time series mean ($CV=0.31$; Figure 5.2.3), the WG chose to use the full recruitment time series for reference point estimation. This approach aligns with practices used for other groundfish stocks in the region (NEFSC 2022b). The resulting $SSB_{F40\%}$ estimate is 4,870 mt. The

decision to use the entire time series of recruitment for estimating SSB_{MSY} proxy reference reflects an assumption that recruitment fluctuations are representative of the stock's long-term productivity and are not influenced by some persistent trends. For the terminal year (2022), SSB/SSB_{40%} = 8645/4870 = 1.77 and F/F_{40%} = 0.06/1.64 = 0.04 (Figure 6.2.1; Table 6.1.1). It should be noted that the primary goal of this research track is to develop assessment and projection methods for future management track assessments. Stock status recommendations are not part of this research track's Terms of Reference, and results will not directly inform management. Instead, findings from this assessment will support a management track assessment scheduled for September 2025, which will incorporate data through 2023 to inform management decisions for 2025-2026.

7.0 Projections

7.1 *Projections Settings*

The WHAM framework offers integrated projection capabilities. For the candidate model, these projections incorporate parameter uncertainty and carry forward random effects in numbers-at-age (NAA) using the ar1_a structure. Short-term projections were performed with the assumption that the 2023 catch would be the same as in 2022, which is standard practice for bridge-year projections. From 2024 to 2026, fishing was set at a rate of F40% = 1.64. To align with the reference points detailed in Section 6.0, fleet selectivity and natural mortality rates were kept constant during the projection period, while projected weights-at-age (WAA) and maturity were averaged from the final two years of data. Similar to the approach used for reference points, average recruitment over the model's full time series (1985–2022) was assumed. A complete overview of settings for these short-term projections is provided in Table 7.1.1. It's important to note that these projections are provisional, as they will need to be updated with 2023–2024 data in the 2025 management track process. Therefore, the projections presented in Section 7.2 are not intended for direct application in management decisions.

7.2 *Projections Results*

Short-term projections for the candidate model, spanning 2023 to 2026, suggest an increase in both catch and Spawning Stock Biomass (SSB). Catch is anticipated to climb from 303 mt in 2023 to a peak of 5,076 tons in 2024, before stabilizing around 2,440 tons by 2026. SSB, on the other hand, is projected to decrease from 8,382 tons in 2023 to 4,768

tons in 2025, with a slight stabilization at 4,784 tons in 2026. Despite these projected increases, both catch and SSB remain lower than historical levels (Table 7.2.1; Figure 7.2.1).

8.0 Discussion

The development of a WHAM-based assessment model for CCGOM yellowtail flounder incorporated several significant methodological updates, each contributing to refining stock estimates. First, splitting the NEFSC survey in 2009—reflecting the transition from the Albatross to the Bigelow survey vessel—was essential to account for potential shifts in catchability due to changes in gear and protocol. Splitting the survey series resulted in better model diagnostics, particularly in retrospective performance, as it helped to mitigate retrospective bias by capturing potential discontinuities in the time series. The decision to split was further validated by slight improvements in the model’s ability to accurately track stock dynamics across survey changes, supporting the inclusion of such adjustments for long-term consistency in data interpretation.

Retrospective analyses indicated scaling issues, particularly with SSB and F, which were addressed through a combined adjustment of survey inclusion/exclusion and selectivity blocks. Scaling in retrospective peels posed challenges, especially in the terminal years, but improvements were achieved by systematically exploring different survey combinations and selectivity assumptions. This iterative approach reduced the scaling problem and helped align retrospective patterns, demonstrating that selective inclusion of surveys and revised selectivity structures can mitigate scaling inconsistencies.

An in-depth examination of fleet selectivity revealed that alternative selectivity structures had a significant impact on model outputs. In the initial model, selectivity estimates for younger ages showed a declining trend, resulting in high F_{MSY} proxy values. To address this, various selectivity blocks and random effects were tested, including independent and identically distributed (iid), autoregressive across ages (ar1), autoregressive across years (ar1_y), and a two-dimensional autoregressive process (2dar1). The preferred candidate model, which removed random effects on fleet selectivity and implemented two selectivity blocks starting in 1994, produced improved retrospective statistics, reduced bias in self-tests, and increased stability in jitter analyses. This approach aligned selectivity assumptions more closely with expected fishing patterns, showing higher selectivity for ages 4–6 in the fishery.

In addition, a sensitivity analysis was conducted to explore the effects of alternative natural mortality assumptions. By lowering the natural mortality (M) from 0.4 to 0.2, the F40% reference point was notably reduced from 1.64 to 0.47, illustrating the sensitivity of fishing

mortality reference points to assumptions about M. This analysis suggests that under lower natural mortality, the population would require more conservative fishing limits to sustain 40% of the unfished spawning potential. This sensitivity analysis underscores the importance of carefully considering natural mortality assumptions, as they directly affect management targets and stock sustainability. The decision to set M at 0.4 in this research track is strongly supported by life-history and longevity-based natural mortality analyses specific to this stock.

One notable difference of the current model compared to the other yellowtail flounder stocks is its lack of environmental covariates, such as the Atlantic Multidecadal Oscillation (AMO) and bottom water temperatures. While preliminary tests indicated that these variables did not significantly enhance model fits or diagnostics, their influence on recruitment dynamics and catchability could be explored further in future assessments. Integrating environmental data could enhance understanding of recruitment variability and potentially improve forecast accuracy under changing climatic conditions.

Comparing the performance of WHAM (Run 452) to the traditional VPA approach revealed several benefits. WHAM's state-space framework allows for the integration of random effects, offering a more robust approach to handling uncertainties in recruitment and selectivity. Model Run 452 produced more stable estimates of SSB and F, with reduced retrospective bias and improved diagnostic metrics compared to VPA. These strengths make WHAM a preferable choice for long-term assessment, particularly when dealing with complex stock dynamics and variable age compositions.

For CCGOM yellowtail flounder, the lognormal adjustment was turned off. WHAM has the capability to adjust estimates of recruitment and numbers-at-age through a process known as lognormal adjustment (formerly referred to as "bias correction"). For details, see Stock and Miller (2021). In the CCGOM yellowtail flounder candidate model, this adjustment is disabled. This decision was based on model diagnostics, treating lognormal adjustment as an adjustable parameter within the model framework. Currently, there is no established guidance on whether the lognormal adjustment should be turned on or off, nor on whether this decision should be stock-specific, although ongoing work may provide direction. This note serves as a reminder for future assessments of CCGOM yellowtail flounder to reconsider this assumption regarding lognormal adjustment if more definitive guidance becomes available.

Future assessment could explore variability in survey catchability, particularly in the context of inshore migration patterns that may affect stock availability to surveys. An investigation into seasonal migration hypotheses could help clarify shifts in survey catchability and potentially inform adaptive management practices. Additionally,

understanding these migration patterns could refine spatial coverage assumptions, offering a more dynamic view of catchability across the range of CCGOM yellowtail flounder.

References

- Gavaris, S. (1988). *An adaptive framework for the estimation of population size*. Canadian Atlantic Fisheries Scientific Advisory Committee Research Document 88/29.
- Francis, R.I.C.C. (2014). "Replacing the multinomial in stock assessment models: A first step." *Fisheries Research*, 151, 70-84.
- Legault, C.M., & Restrepo, V.R. (1998). "A flexible forward age-structured assessment program." *ICCAT Collected Volume of Scientific Papers*, 49(3), 246-253.
- Miller, T.J., Das, C., Politis, P.J., Miller, A.S., Legault, C.M., Brown, R.W., & Wright, W.R. (2010). *Estimating Albatross IV to Henry B. Bigelow calibration factors*. NOAA Technical Memorandum NMFS-NE-206. Northeast Fisheries Science Center.
- Miller, T.J., Legault, C.M., and Brown, R.W. (2013). *Calibration factors for yellowtail flounder in the transition from the Albatross IV to the Henry B. Bigelow*. Northeast Fisheries Science Center Reference Document. NOAA Fisheries.
- Miller, T.J., & Stock, C.A. (2020). *A state-space framework for incorporating covariates and random effects in age-structured stock assessment models: The WHAM model*. ICES Journal of Marine Science, 77(7-8), 2540–2550. <https://doi.org/10.1093/icesjms/fsaa072>
- Thorson, J.T., & Punt, A.E. (2014). *Status of the Vector Autoregressive Spatio-Temporal (VAST) Package for Modeling the Spatial-Temporal Distribution of Species Biomass and Abundance*. *Fisheries Research*, 169, 145–155. <https://doi.org/10.1016/j.fishres.2014.02.027>
- Stock, B.C., & Miller, T.J. (2021). *A State-Space Approach to Fisheries Stock Assessment Model Construction Using WHAM (Woods Hole Assessment Model)*. *Fisheries Research*, 240, 105959. <https://doi.org/10.1016/j.fishres.2021.105959>

Tables

Table 4.11.1. Candidate model (m452) settings.

Age Classes	1-6+
Fleet Selectivity	Logistic
Fleet Selectivity Random Effects	None
Fleet Age Composition	Logistic normal-miss0
Fleet CV	0.05
Survey Selectivities	Logistic
Survey Selectivity Random Effects	None
Survey Catchability Random Effects	None
Survey Catchability Environmental Effects	None
Survey Age Compositions	Logistic normal-ar1-miss0
Recruitment Assumptions	Not decoupled from ages 2-6+ Random about the mean (1985-2022)
Environmental Covariate	None
Natural Mortality Assumptions	M = 0.4 across all ages and years
Numbers-at-Age Random Effects	ar1_a

Table 6.1.1. Biological reference points estimated as described in section 6.0 and terminal year (2022) values of SSB and F.

$F_{40\%}$	1.64
F_{2022}	0.06
$SSB_{40\%}$	4,870 mt
SSB_{2022}	8,645 mt
MSY proxy	1,998 mt

Table 7.1.1. Settings used for projections of the candidate model m452.

Random Effects	ar1_a for NAA
Natural Mortality	Constant ($M = 0.4$)
Weight-at-Age	Terminal 2-year average
Maturity	Terminal 2-year average
Recruitment	Mean 1985 - 2022
Environmental Covariates	None

Table 7.2.1. Estimates and uncertainties (90% and 95% confidence intervals) of four years of projected Catch (mt), F, R (000s), and SSB (mt). Forecasts were done using bridge year (2023) catch equal to 2022 catch and then fishing at $F_{40\%} = 0.73$ in years 2024-2026.

Type	Year	Estimation	Low 90	High 90	Low 95	High 95
Catch	2023	303	303	303	303	303
Catch	2024	5,076	1,502	17,152	1,189	21,658
Catch	2025	2,476	594	10,304	452	13,540
Catch	2026	2,439	475	12,526	347	17,137
F	2023	0.056	0.022	0.141	0.019	0.168
F	2024	1.638	1.273	1.876	1.226	1.947
F	2025	1.638	1.273	1.876	1.226	1.947
F	2026	1.638	1.273	1.876	1.226	1.947
SSB	2023	8,382	3,456	20,327	2,917	24,087
SSB	2024	5,838	1,850	18,422	1,485	22,959
SSB	2025	4,768	1,162	19,571	886	25,652
SSB	2026	4,784	1,104	20,739	833	27,468
R	2023	25,286	7,721	82,808	6,152	103,937
R	2024	25,286	7,721	82,808	6,152	103,937

R	2025	25,286	7,721	82,808	6,152	103,937
R	2026	25,286	7,721	82,808	6,152	103,937

Figures

FLOUNDER, YELLOWTAIL - Cape Cod/Gulf of Maine (Sex: NONE)

Map of statistical areas comprising the stock area (1964 - 2024)

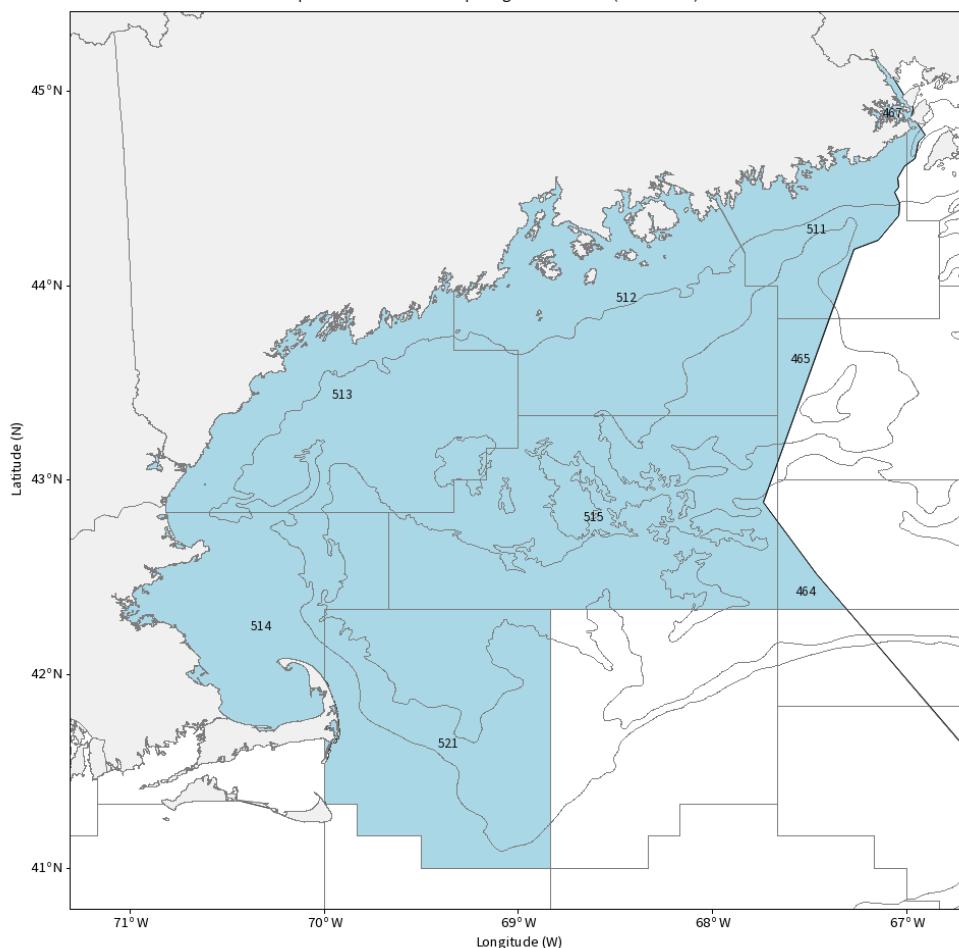


Figure 2.0.1: Statistical areas that comprise the Cape Cod/Gulf of Maine yellowtail flounder stock.

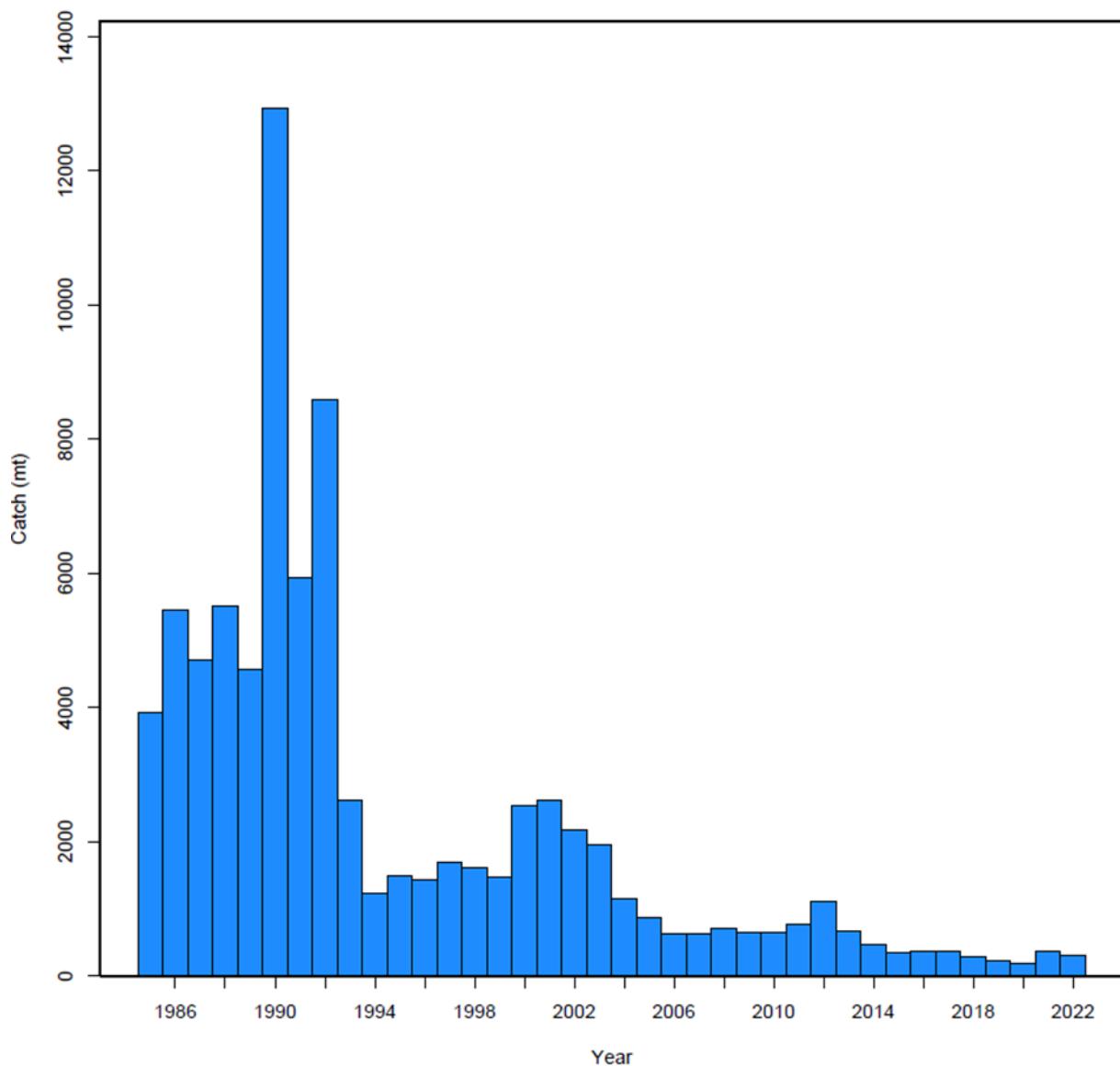


Figure 3.1.1: Total removals of yellowtail flounder from the Cape Cod/Gulf of Maine stock. This represents a combined landings and discards.

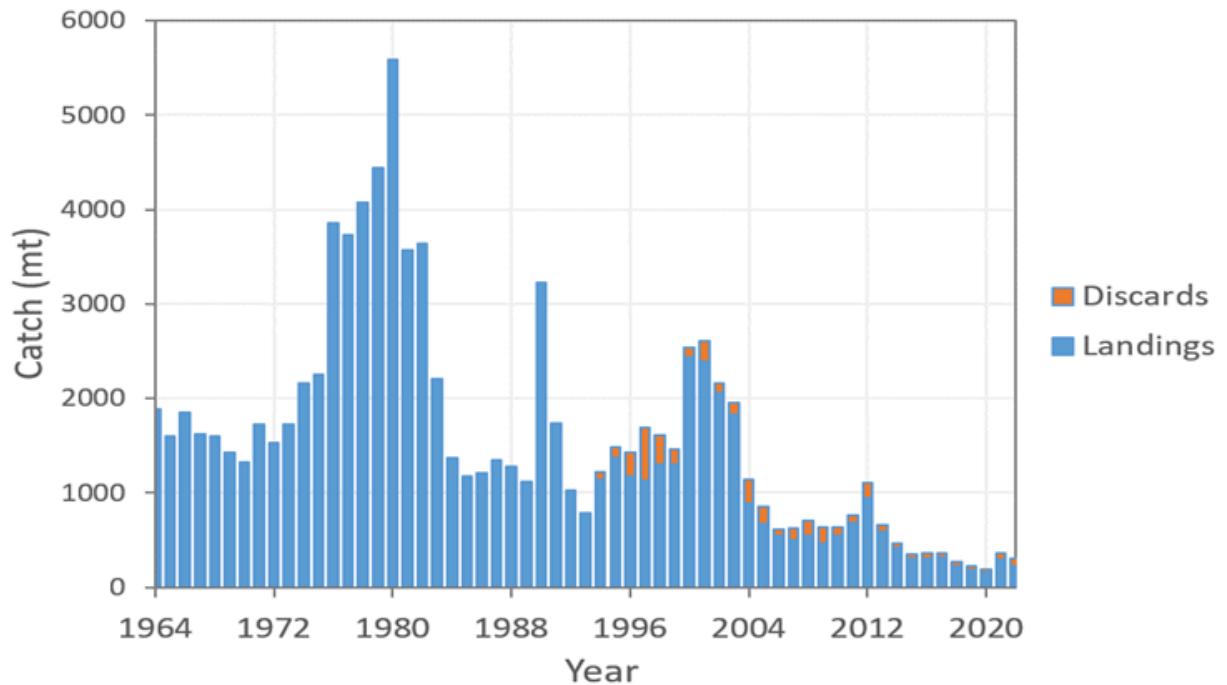


Figure 3.1.2: Total removals of yellowtail flounder from the Cape Cod / Gulf of Maine stock disaggregated by discards, and landings.

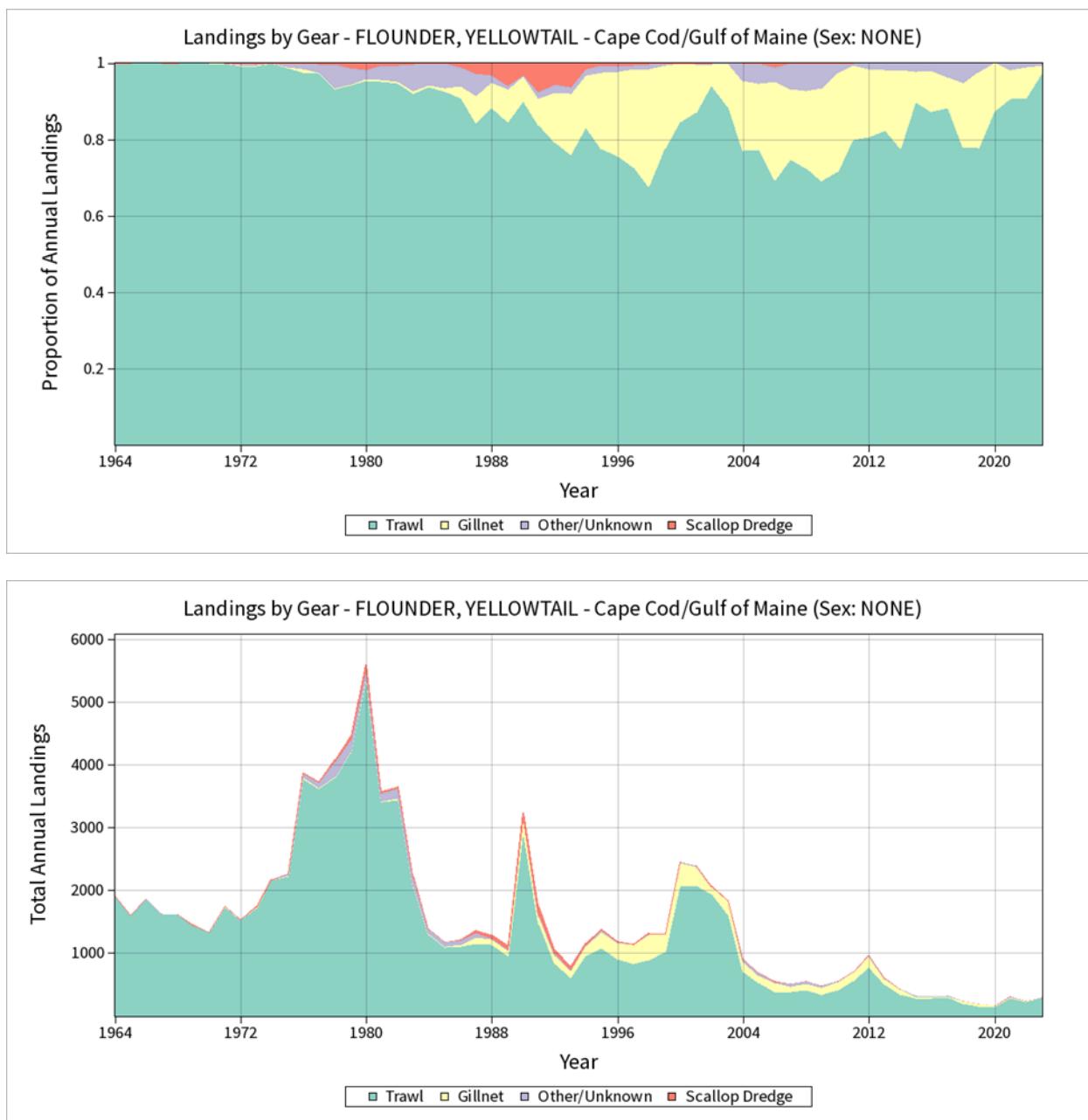


Figure 3.1.3: Proportion of total landings of yellowtail flounder from the Cape Cod/Gulf of Maine stock disaggregated by gear type (top) and total removals by gear type (bottom).

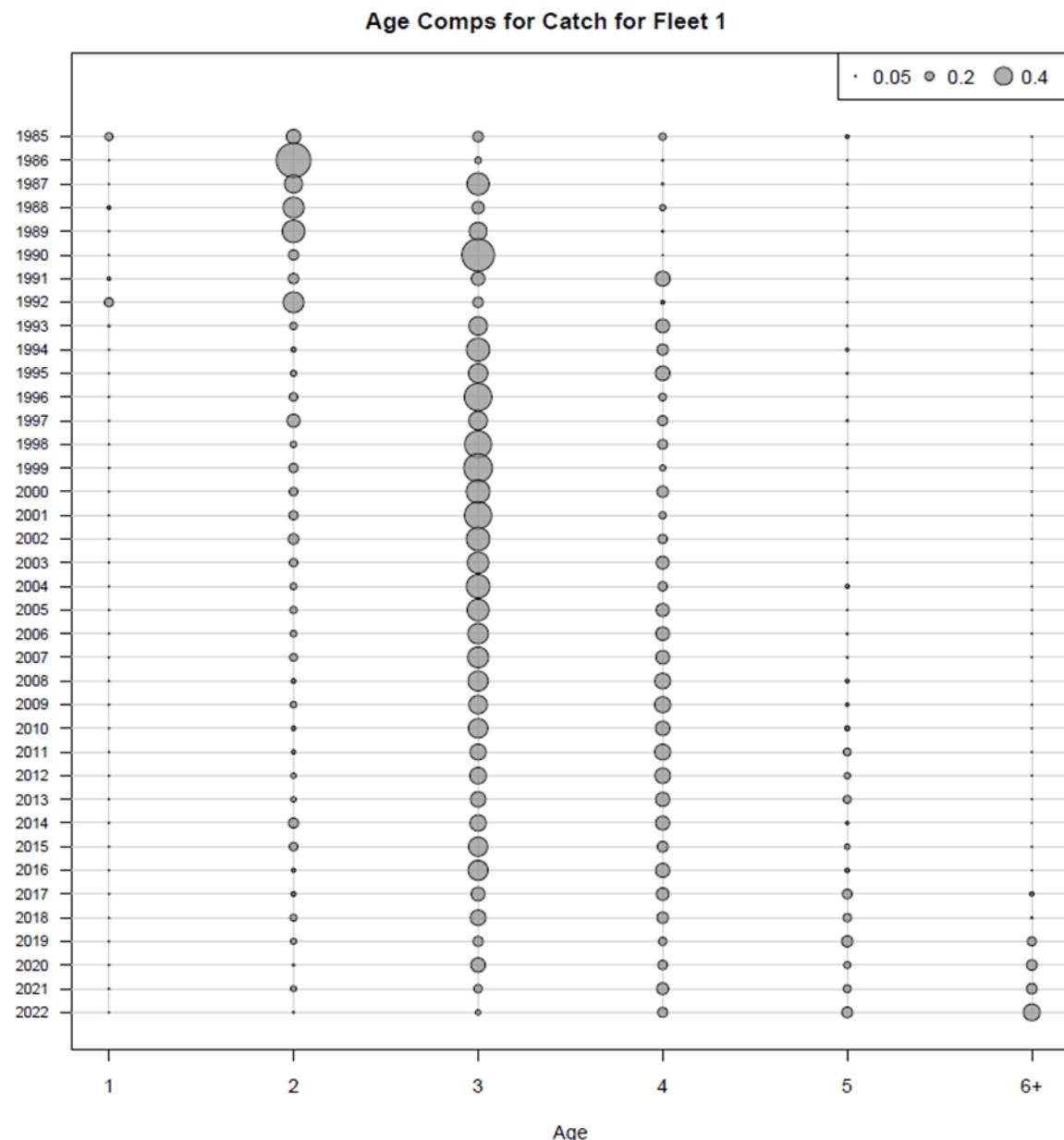


Figure 3.1.4: Age composition of the combined fishery catch.

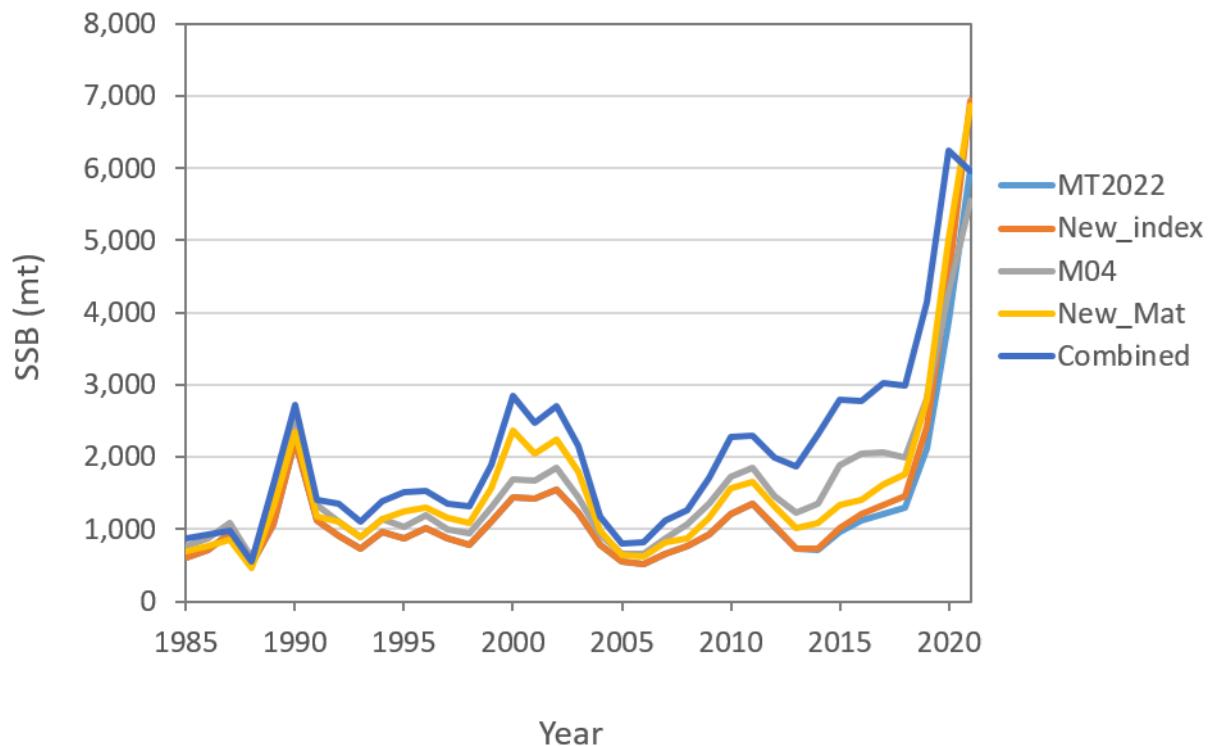


Figure 4.1.1: VPA Comparison of SSB estimates from different data revisions to the previous 2022 Management Track assessment.

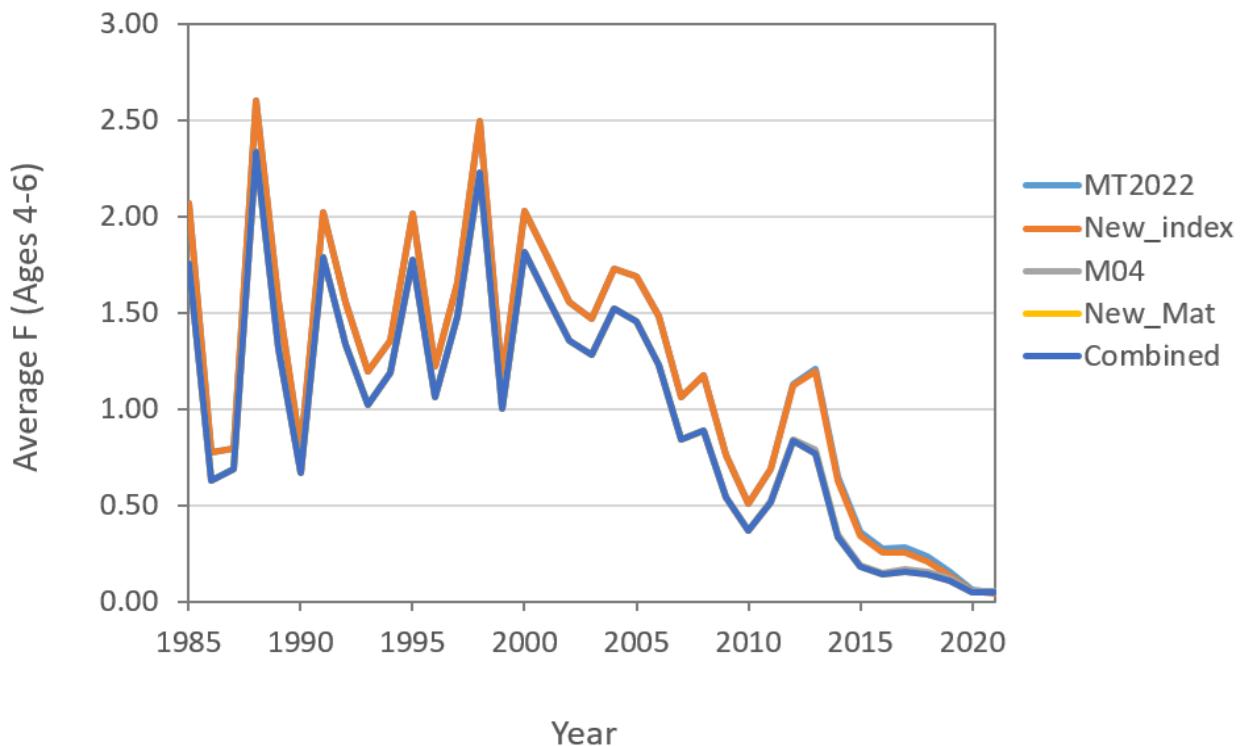


Figure 4.1.2: VPA Comparison of F estimates from different data revisions to the previous 2022 Management Track assessment.

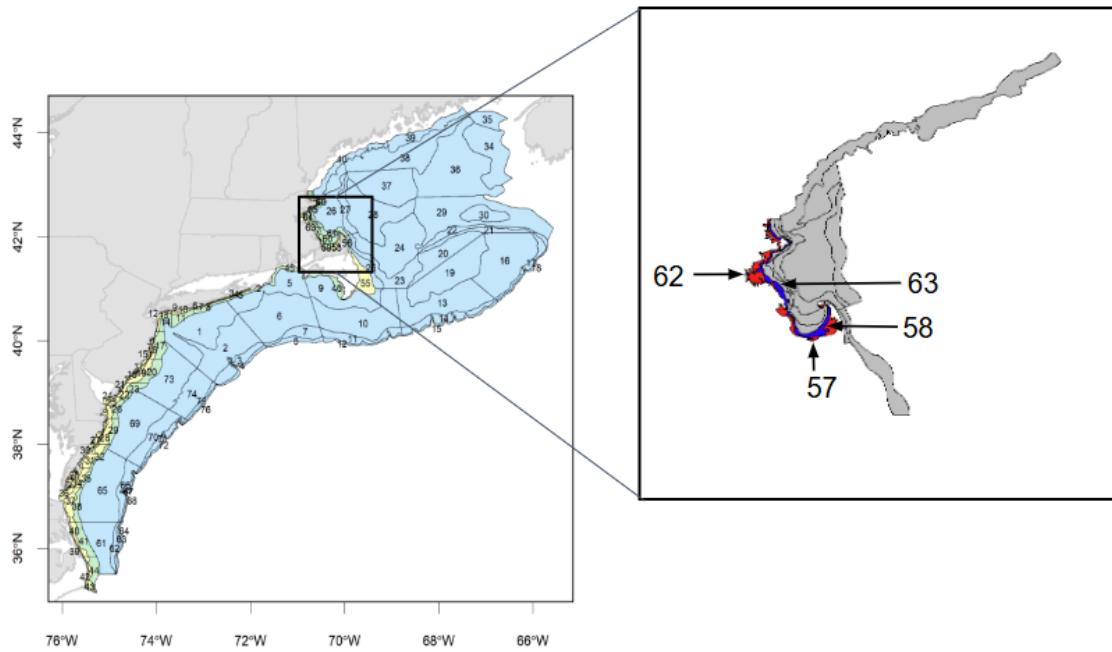


Figure 4.4.2.1: Map showing survey strata adjustments for CCGOM yellowtail flounder: Strata 57 and 62 (highlighted in red), previously unsampled, are now excluded from the survey design. Strata 58 and 63 (highlighted in blue), which recorded yellowtail flounder catches but were previously excluded, are now incorporated.

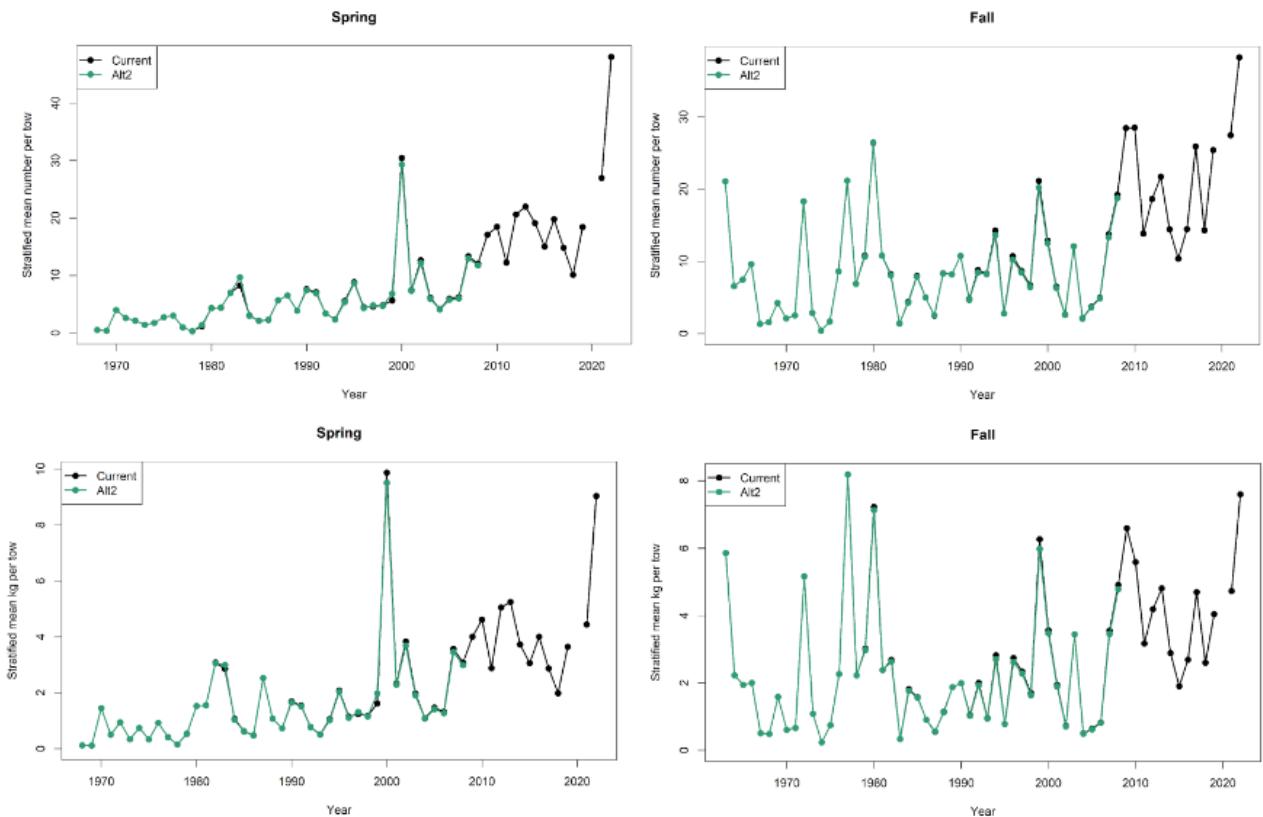


Figure 4.4.2.2: Comparison of survey trends for CCGOM yellowtail flounder between the current survey strata and the revised survey strata (Alt2). The top panels display numbers per tow for spring (left) and fall (right), while the bottom panels show weight per tow for spring (left) and fall (right).

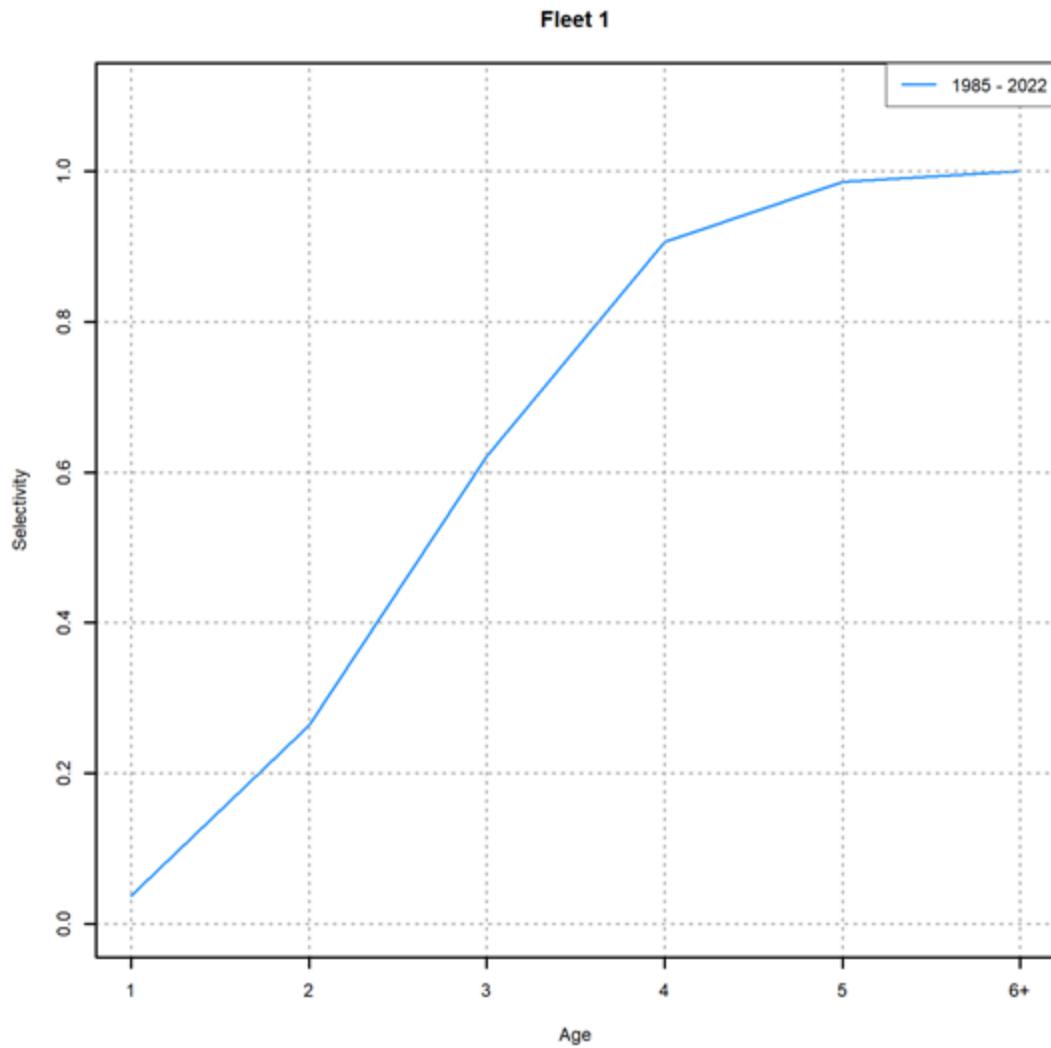


Figure 4.4.3.1: Combined fleet selectivity from model m5 (Not final candidate model) averaged across the series 1973-2022.

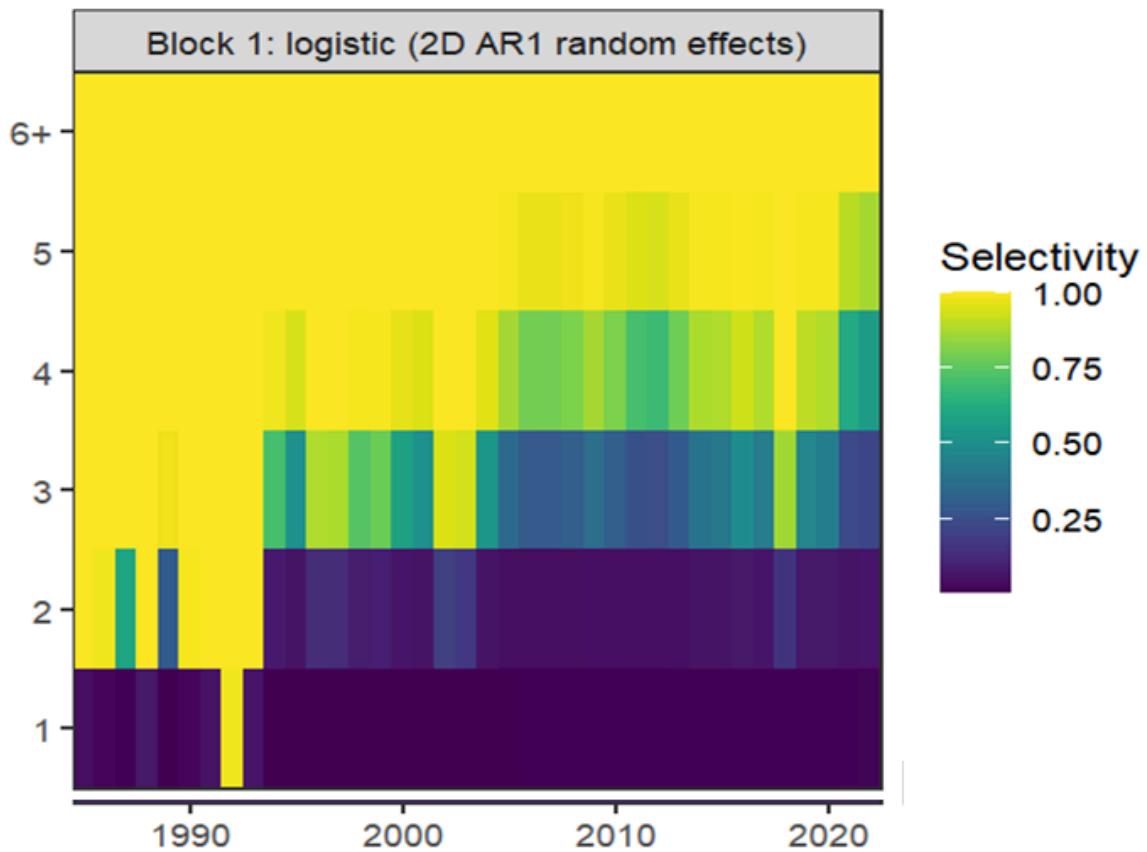


Figure 4.4.3.2: Fleet selectivity from 1985-2022 from model m5 (Not final candidate model) with 2dar1 random effects

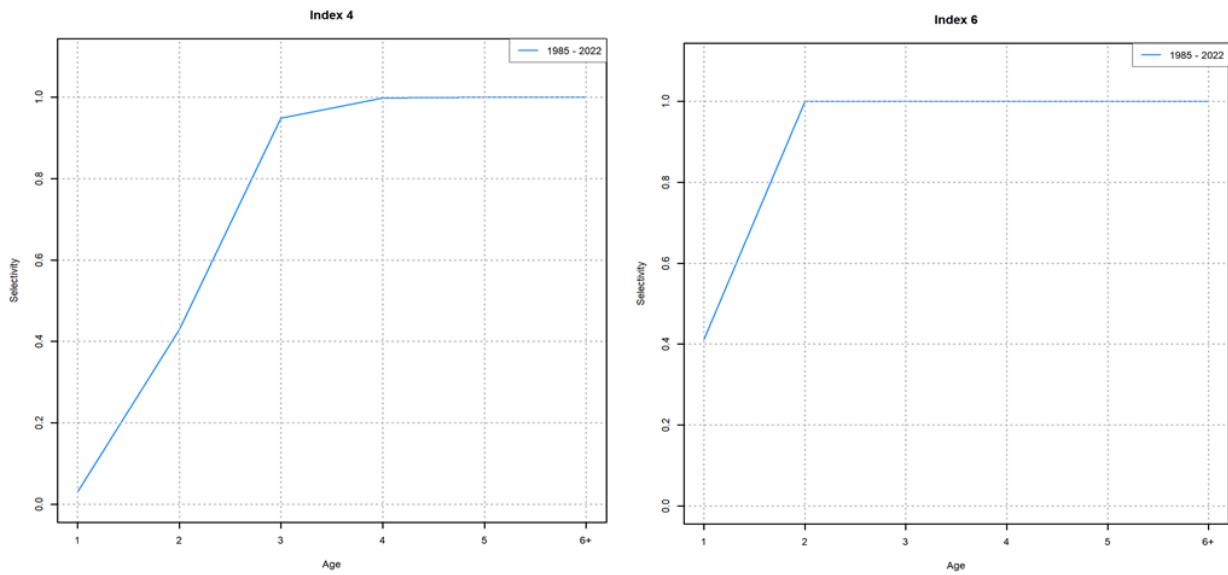


Figure 4.4.4.1: Ex. of estimated selectivity pattern from indices used in model m5. NEFSC Bigelow Fall (Left) and NEFSC Bigelow Spring (Right).

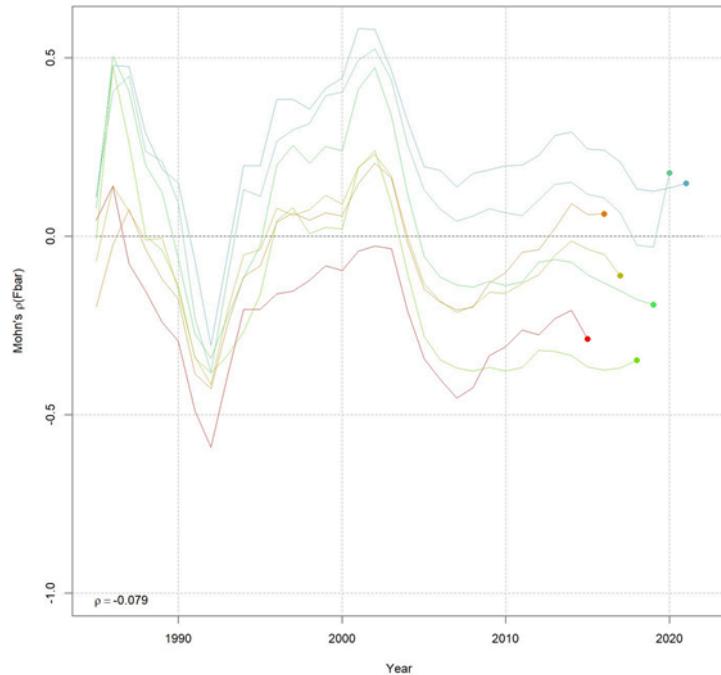
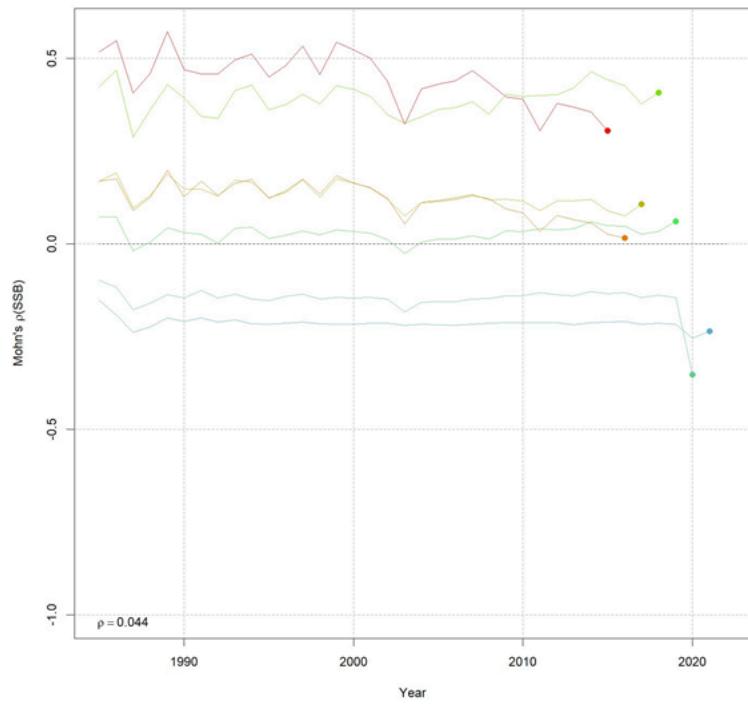


Figure 4.4.5.1: Retrospective pattern during model development for model m14. SSB(top) and F(bottom) from model m14.

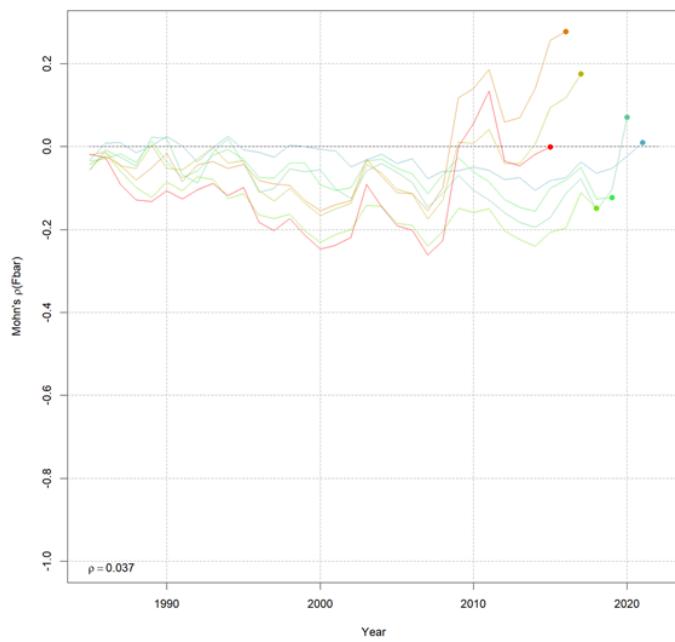
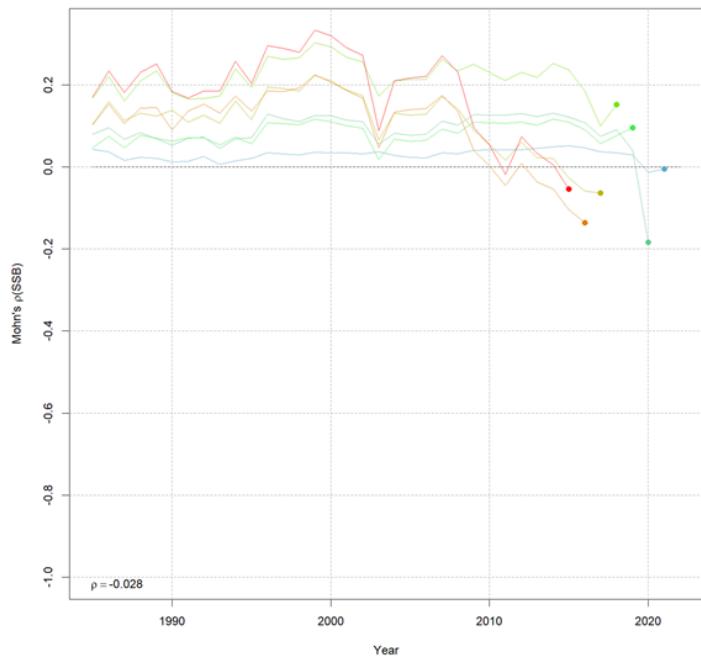


Figure 4.4.9.1: Retrospective pattern during model development for model m304. SSB(top) and F(bottom) from model m304.

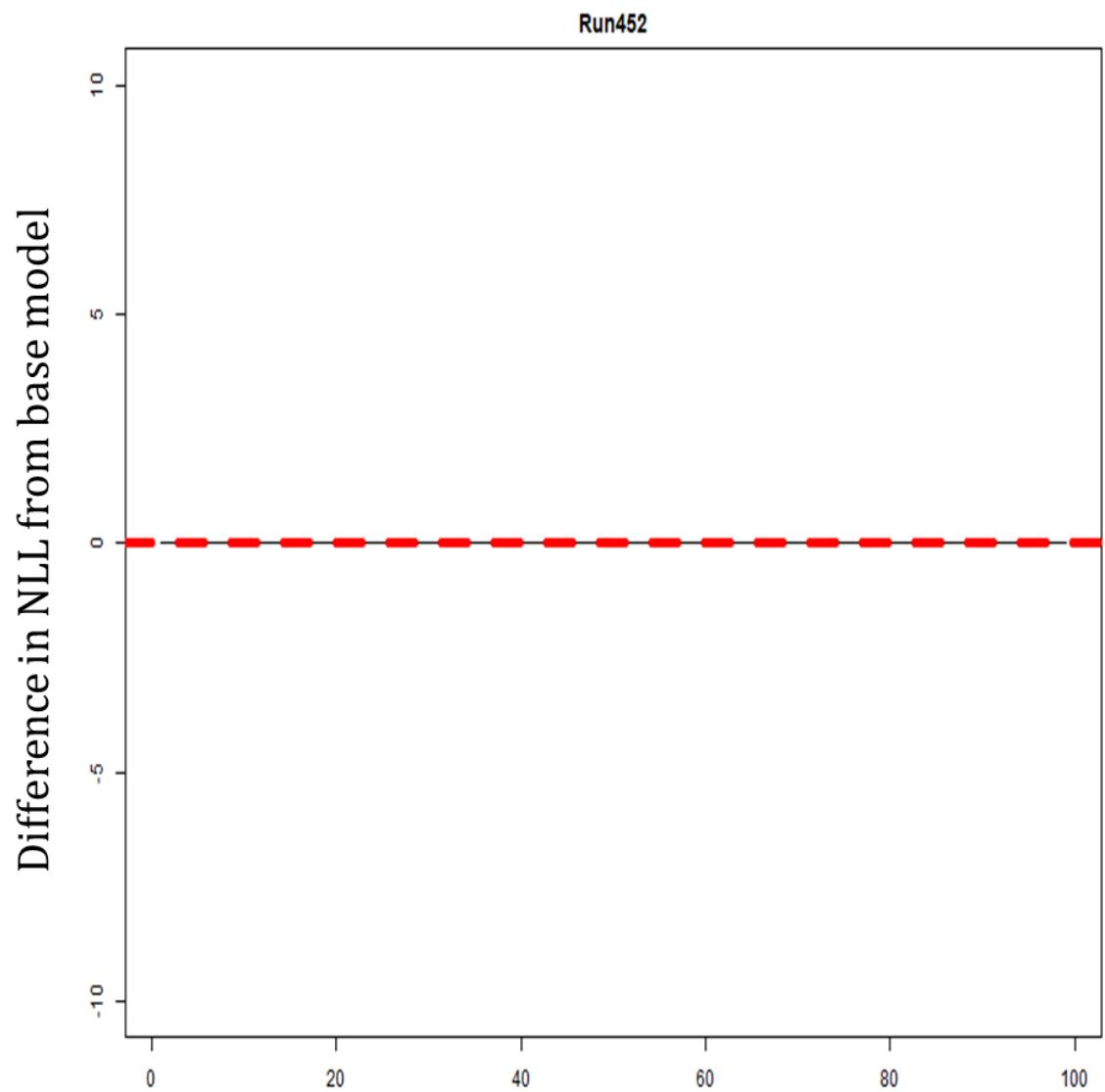


Figure 5.1.1: Jitter analysis results of the candidate model. Convergence rate was 100% and all but three runs (of 100) converged to the global minimum, indicating high stability.

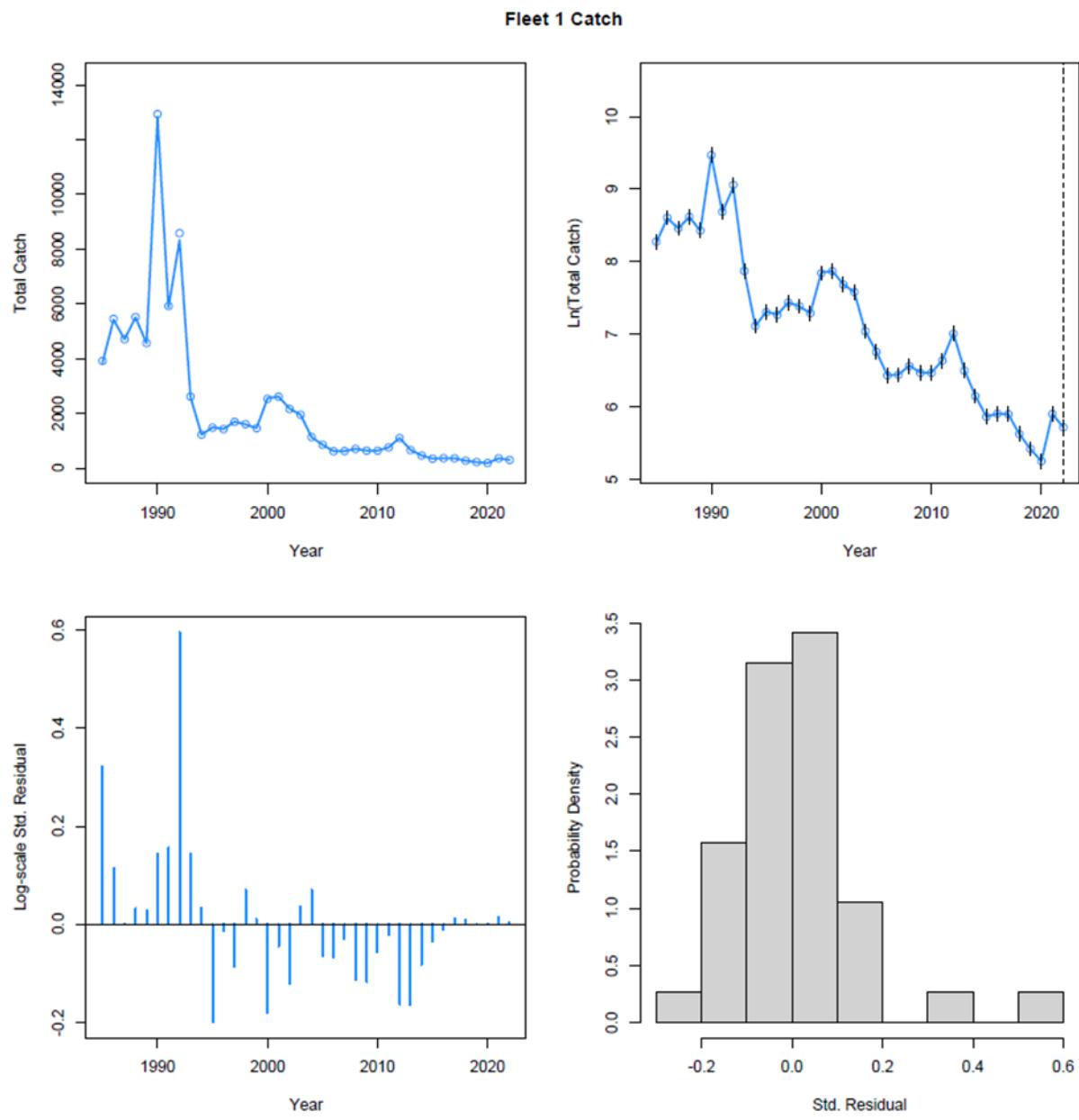


Figure 5.1.2: Candidate model m452 model fit to the aggregate fleet data

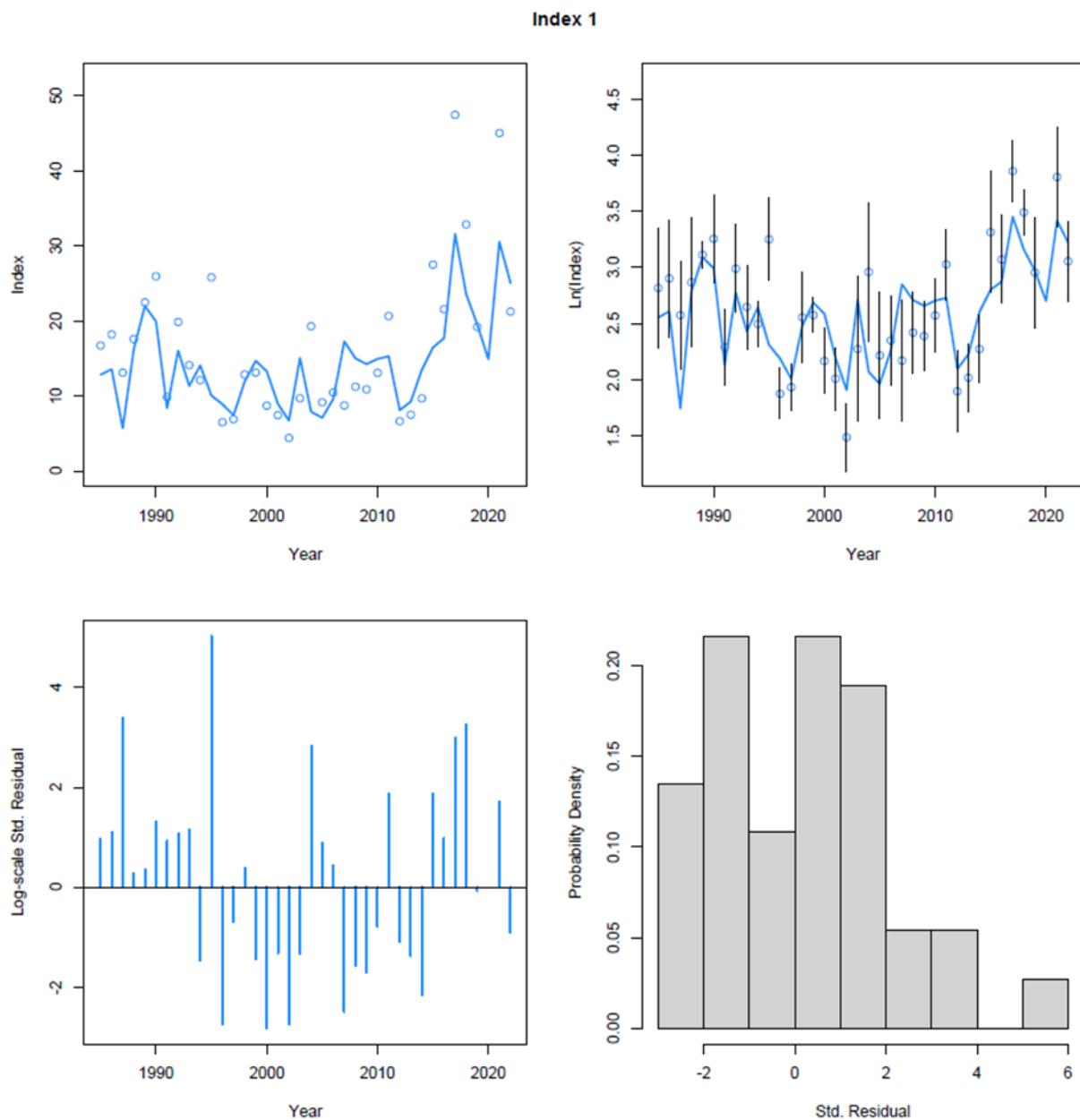


Figure 5.1.3: Candidate model m452 model fit to the aggregate MADMF fall bottom trawl survey index.

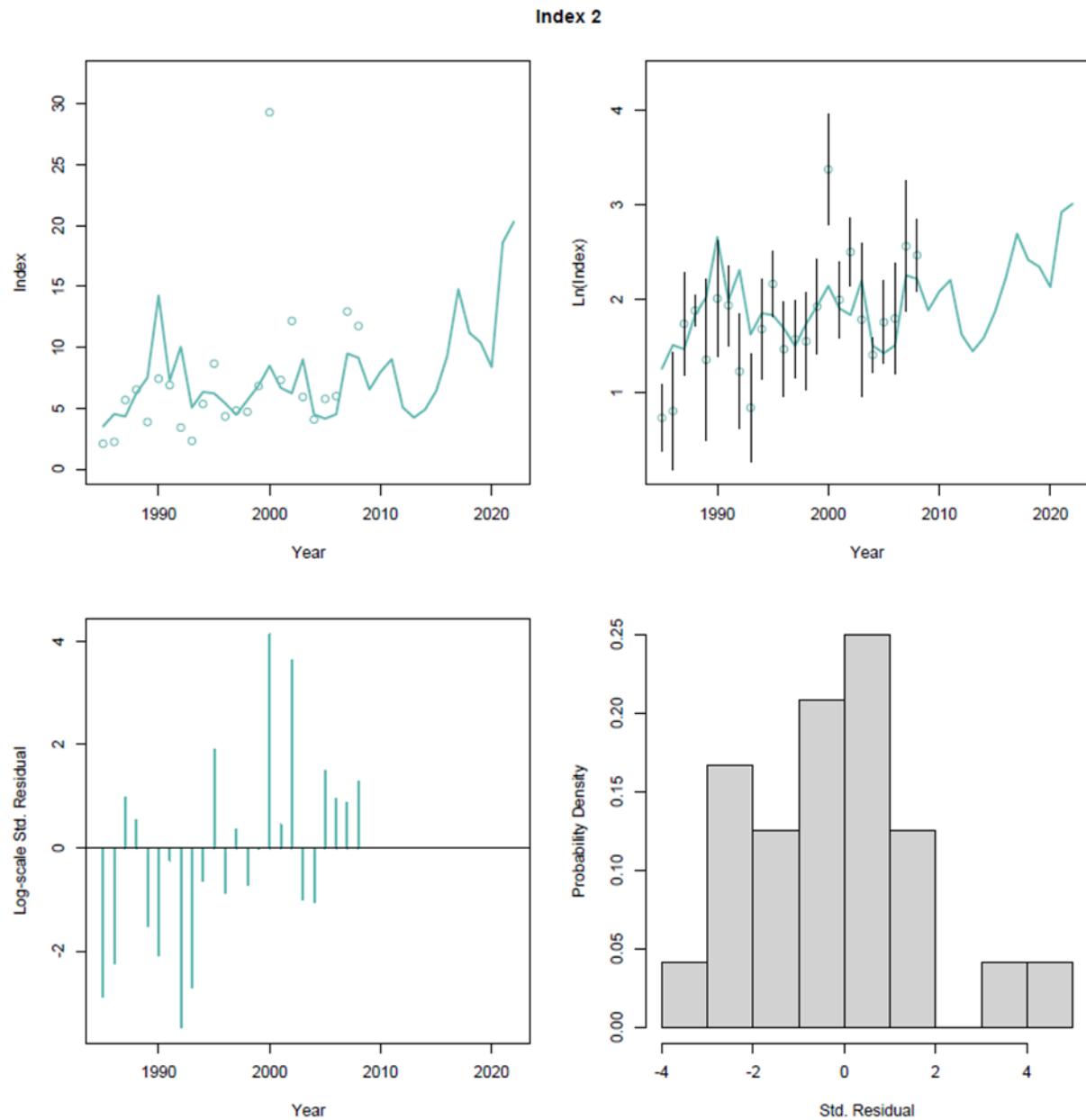


Figure 5.1.4: Candidate model m452 model fit to the aggregate NEFSC spring bottom trawl survey index for the Albatross vessel (1985-2008).

Index 3

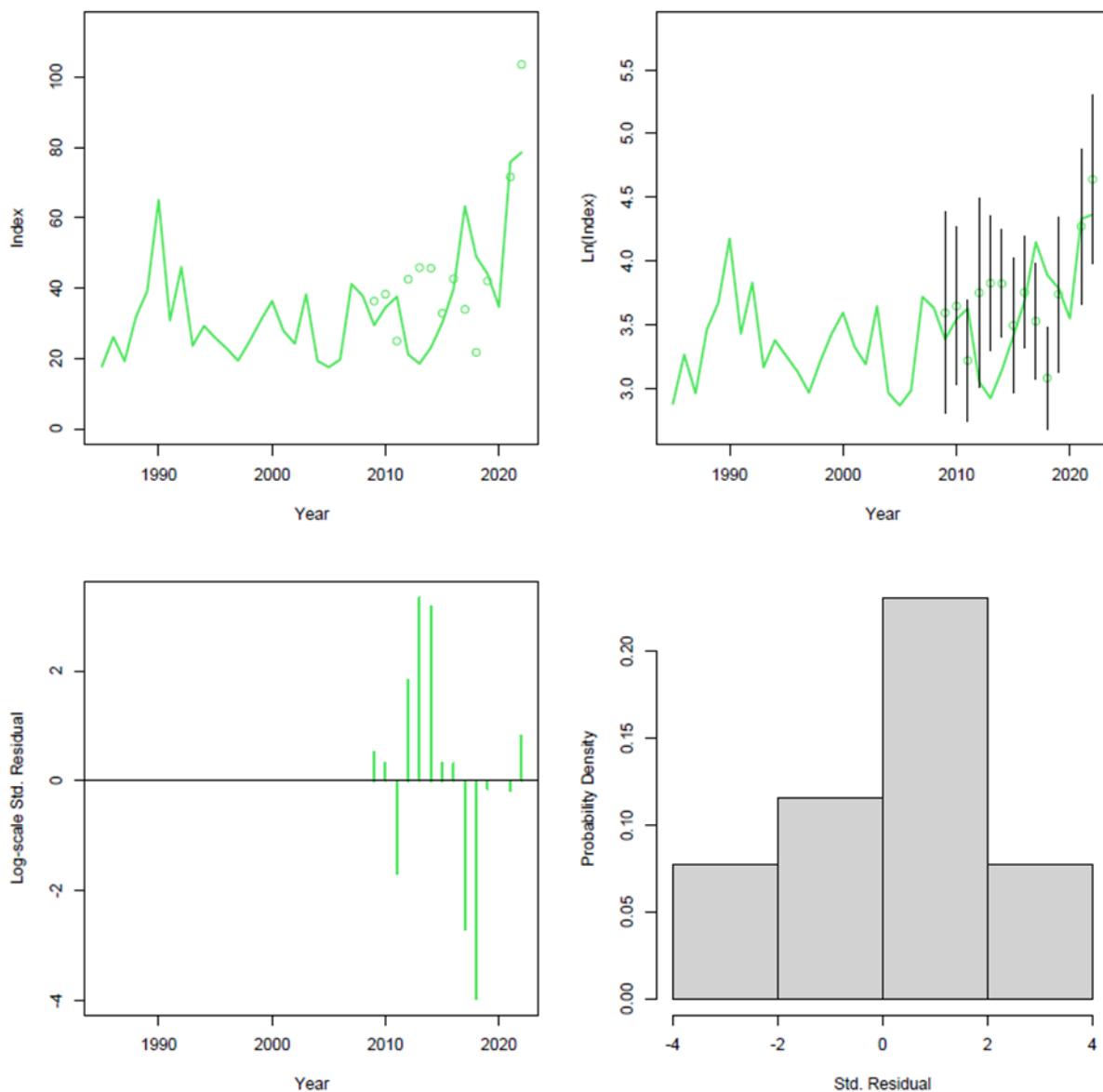


Figure 5.1.5: Candidate model m452 model fit to the aggregate NEFSC spring bottom trawl survey index for the Bigelow vessel (2009-2022).

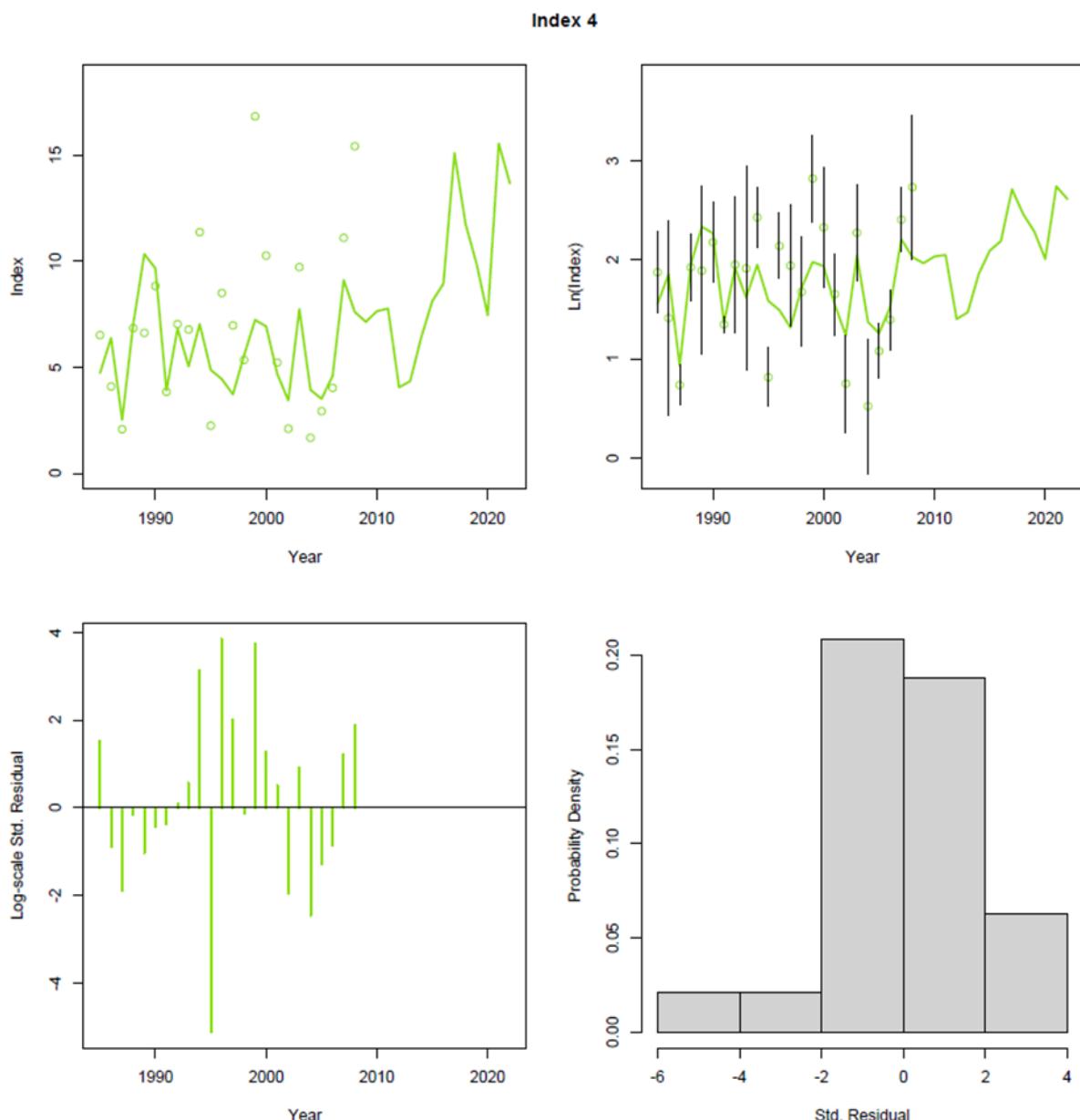


Figure 5.1.6: Candidate model m452 model fit to the aggregate NEFSC fall bottom trawl survey index for the Albatross vessel (1985-2008).

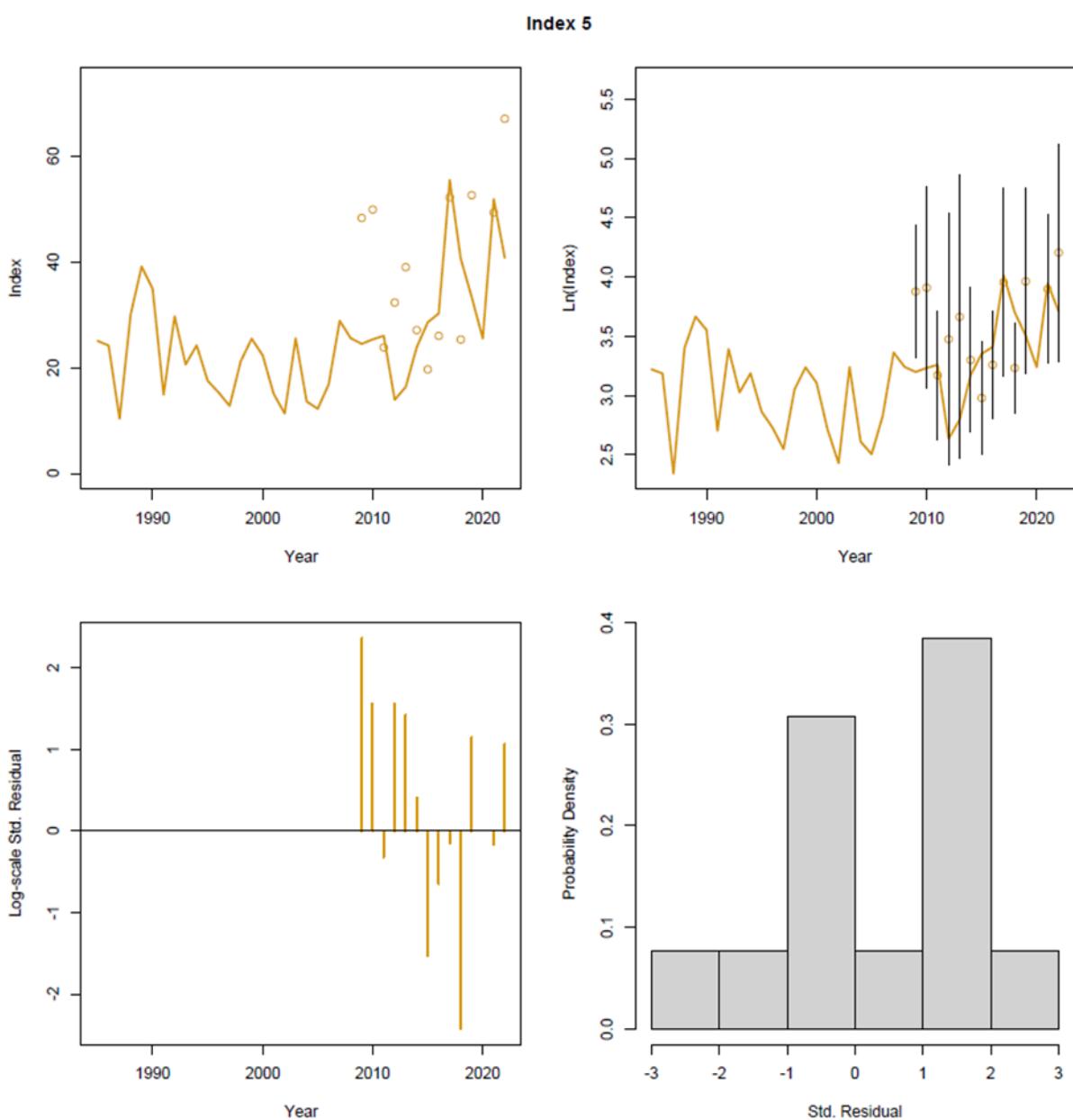


Figure 5.1.7: Candidate model m452 model fit to the aggregate NEFSC fall bottom trawl survey index for the Bigelow vessel (2009-2022).

Index 6

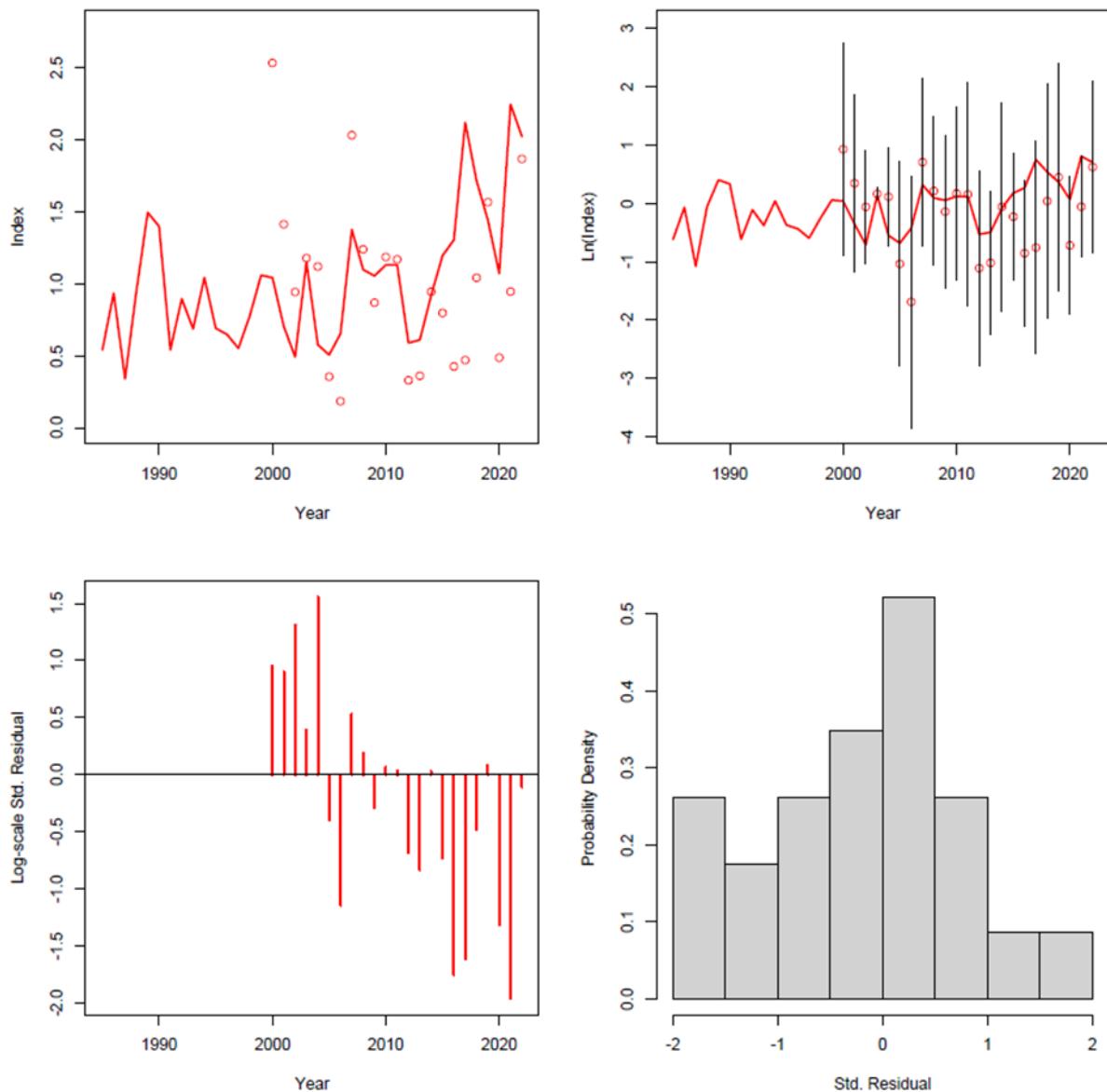


Figure 5.1.8: Candidate model m452 model fit to the aggregate Inshore MENH fall bottom trawl survey index.

OSA residual diagnostics: Fleet 1

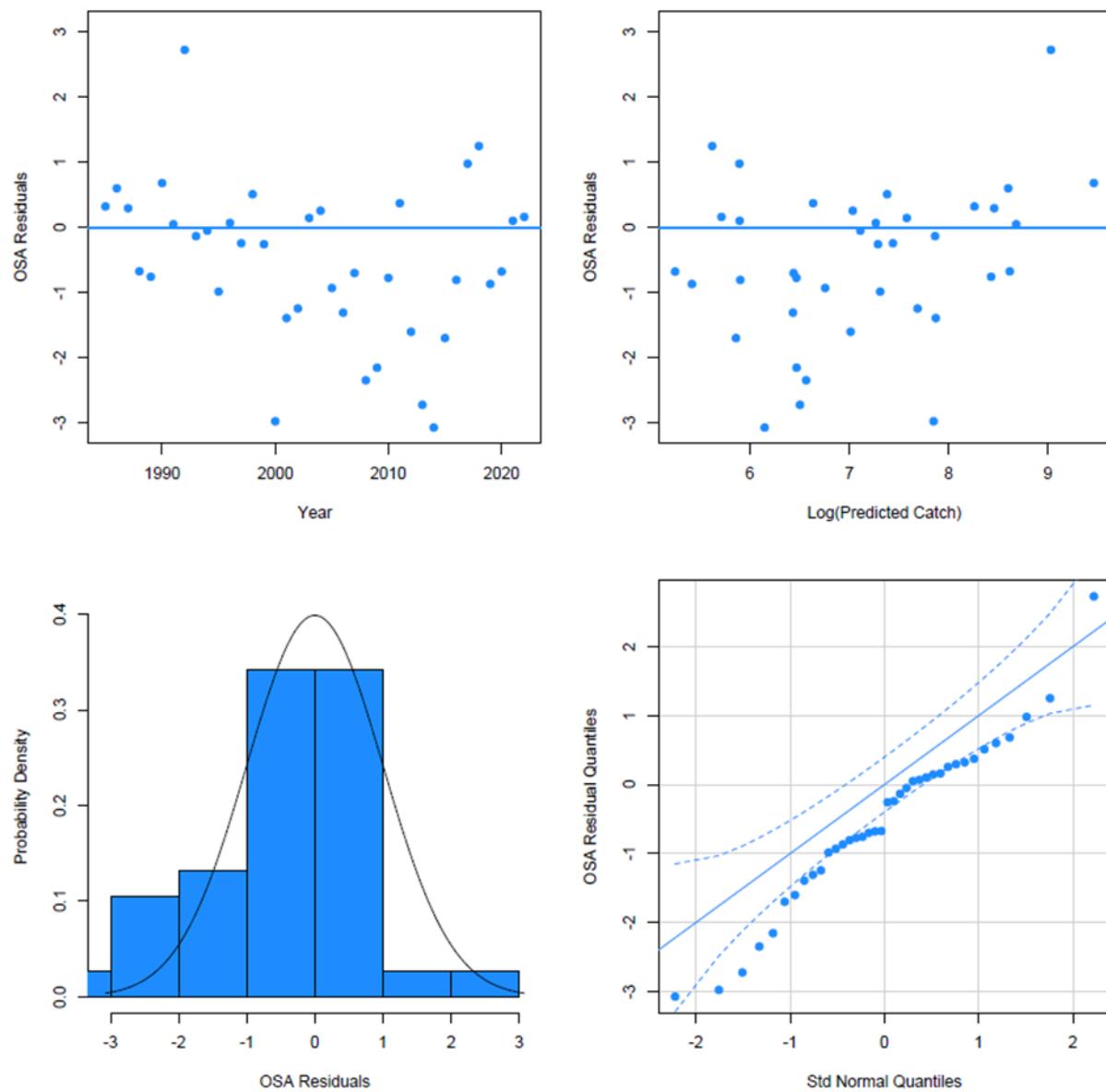


Figure 5.1.9: Candidate model m452 OSA residual diagnostic for the aggregate fleet.

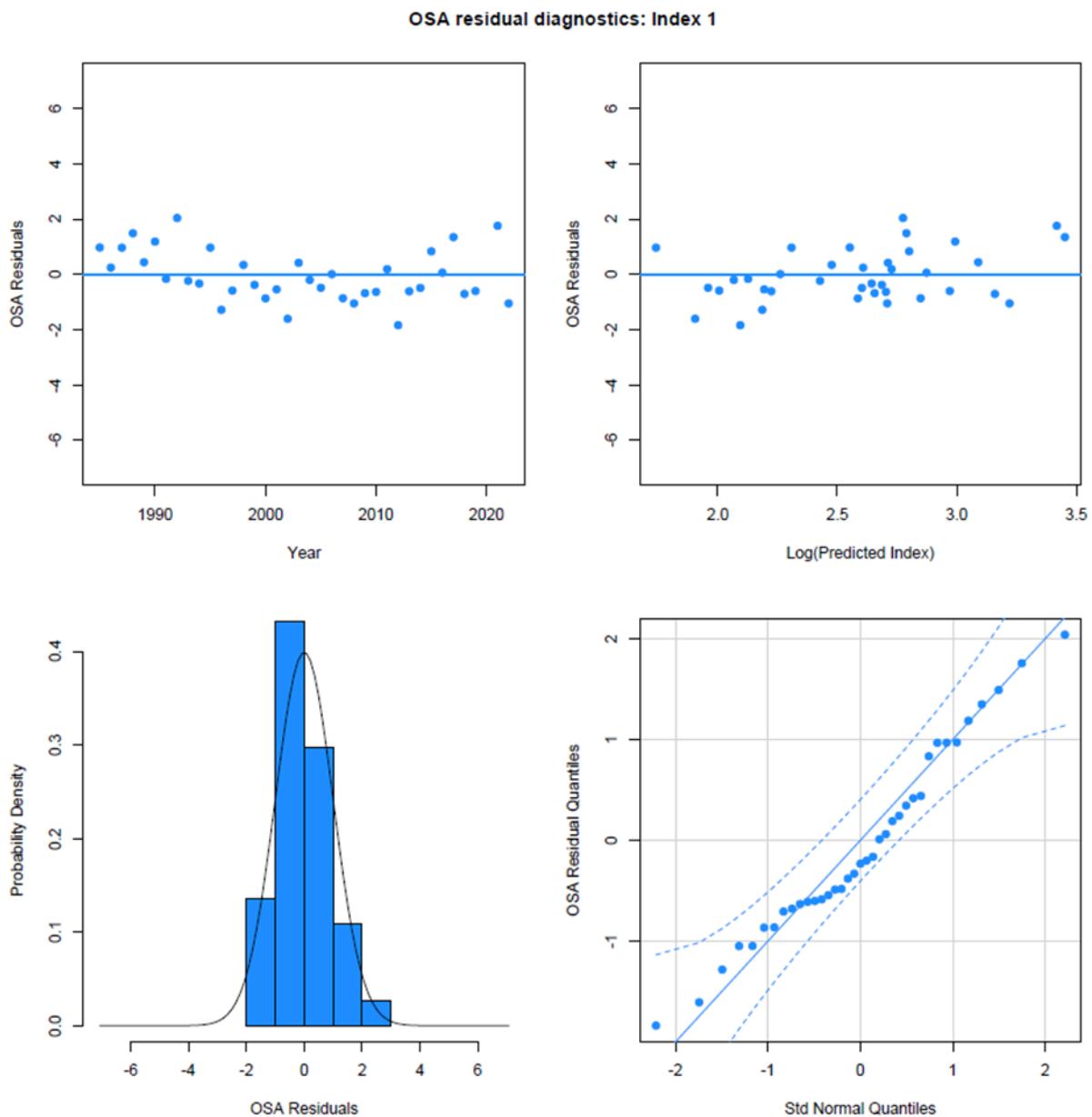


Figure 5.1.10: Candidate model m452 OSA residual diagnostic for the Inshore MADMF fall bottom trawl index.

OSA residual diagnostics: Index 2

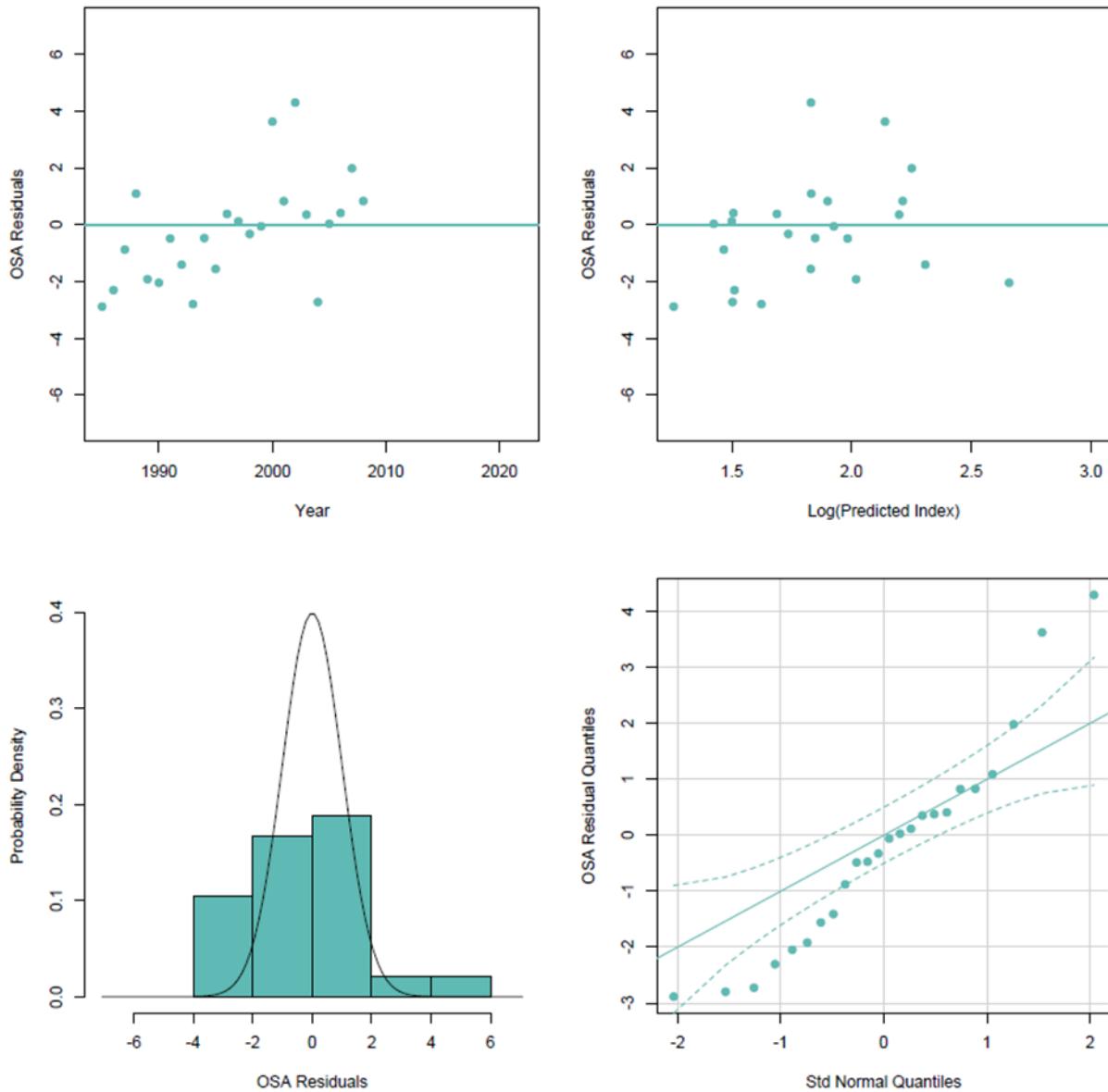


Figure 5.1.11: Candidate model m452 OSA residual diagnostic for the NEFSC Spring bottom trawl index for the Albatross vessel (1985-2008).

OSA residual diagnostics: Index 3

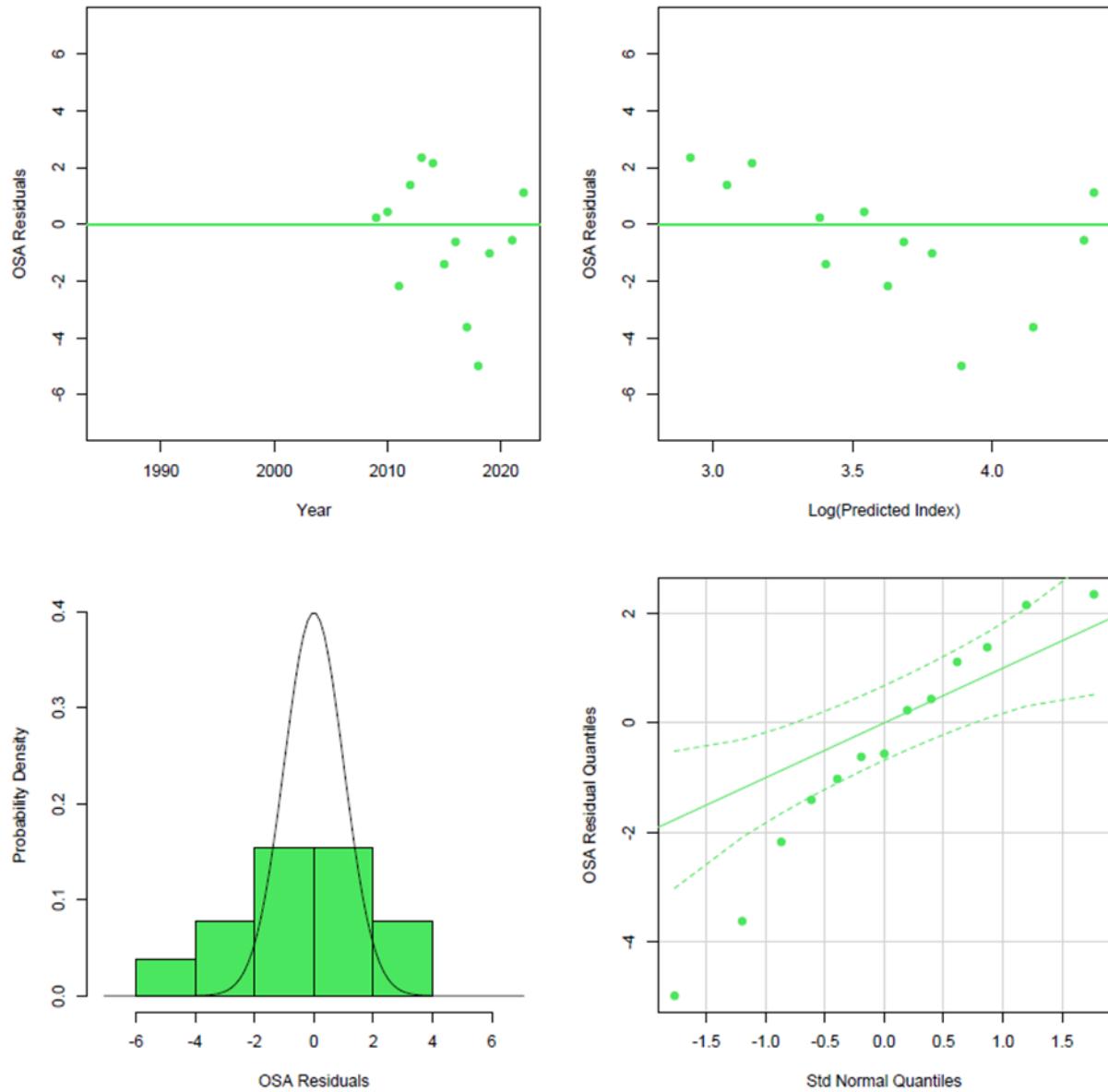


Figure 5.1.12: Candidate model m452 OSA residual diagnostic for the NEFSC Spring bottom trawl index for the Bigelow vessel (2009-2022).

OSA residual diagnostics: Index 4

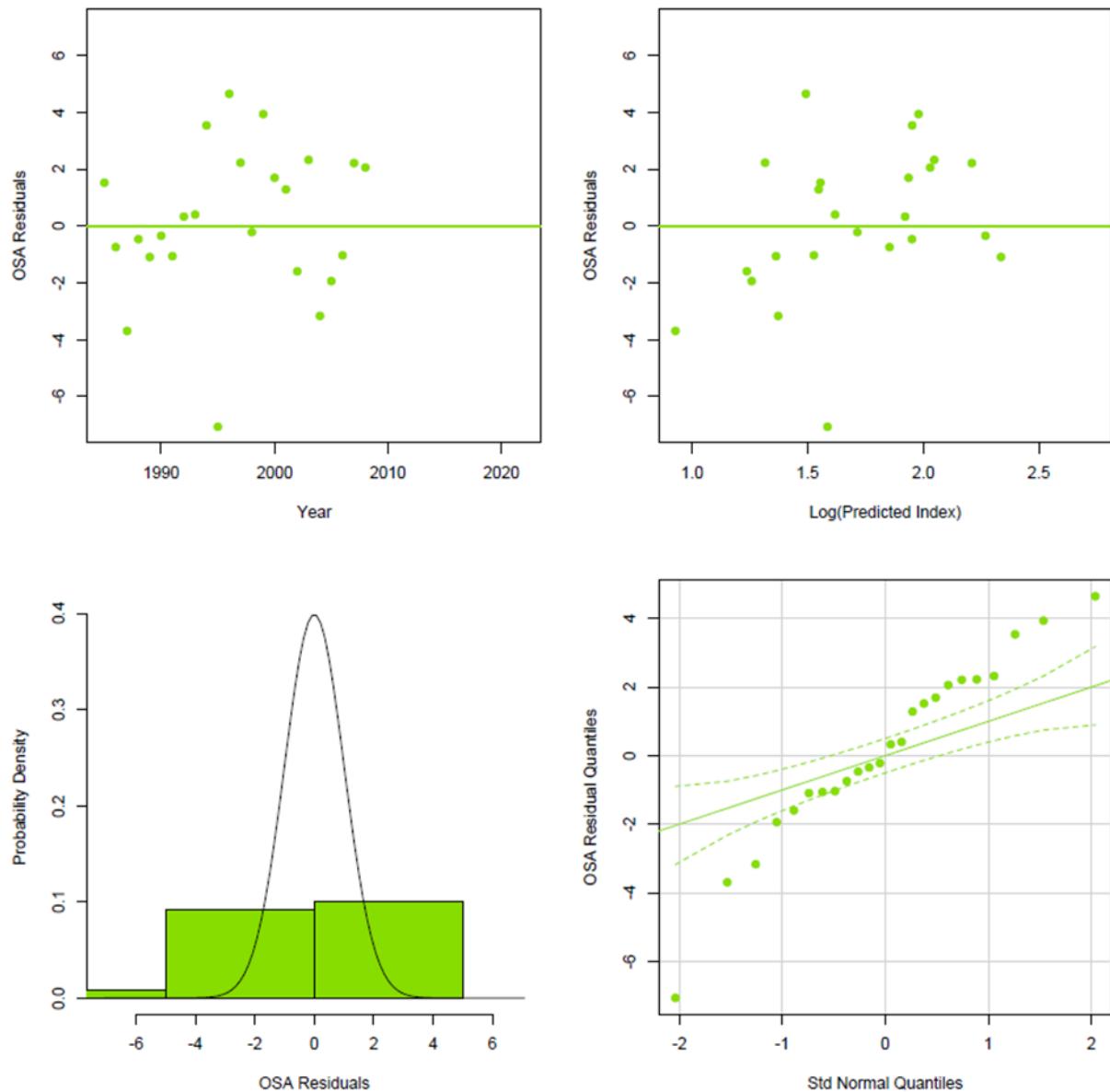


Figure 5.1.13: Candidate model m452 OSA residual diagnostic for the NEFSC Fall bottom trawl index for the Albatross vessel (1985-2008).

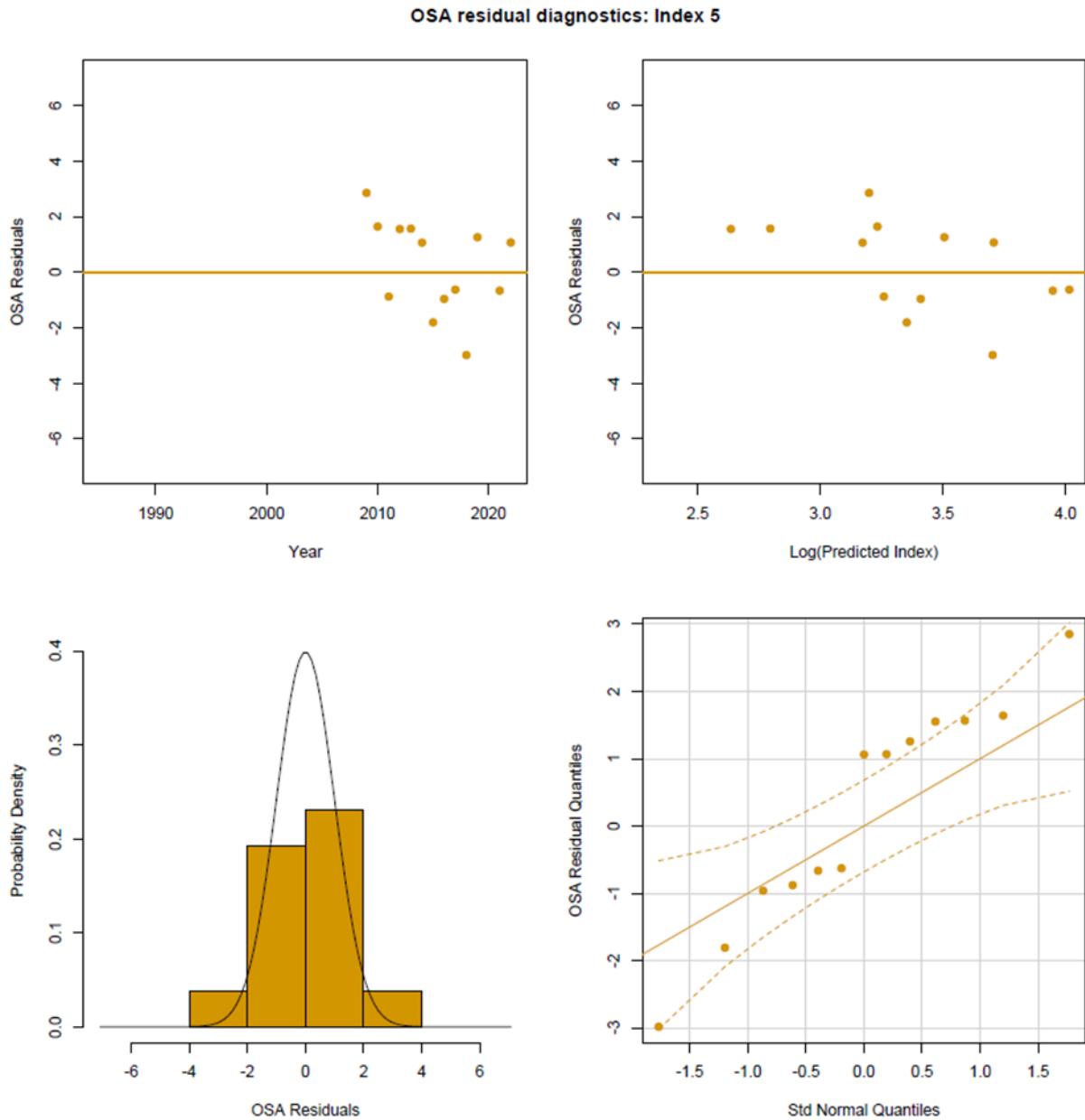


Figure 5.1.14: Candidate model m452 OSA residual diagnostic for the NEFSC Fall bottom trawl index for the Bigelow vessel (2009-2022).

OSA residual diagnostics: Index 6

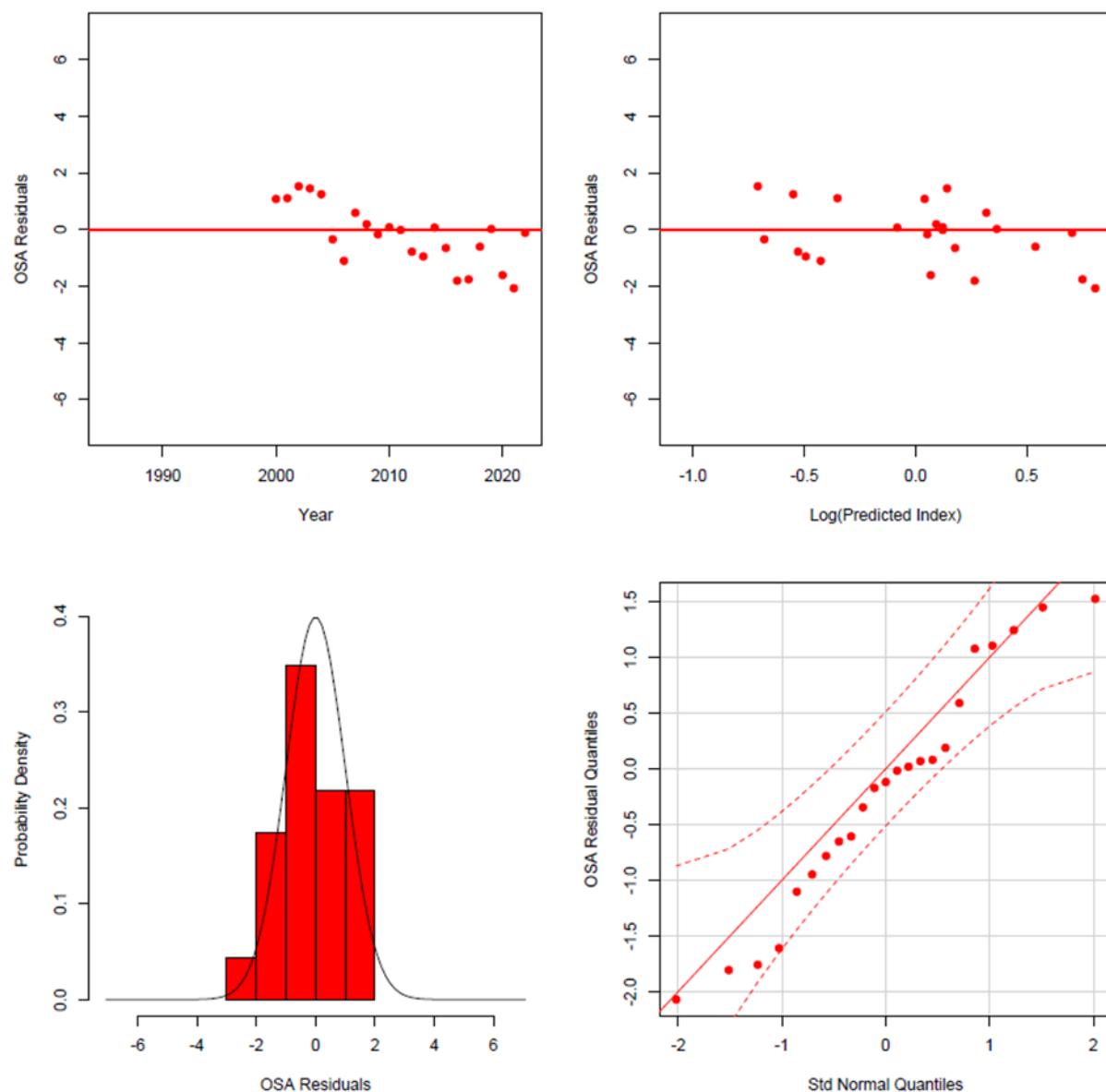


Figure 5.1.15: Candidate model m452 OSA residual diagnostic for the Inshore MENH Fall bottom trawl index.

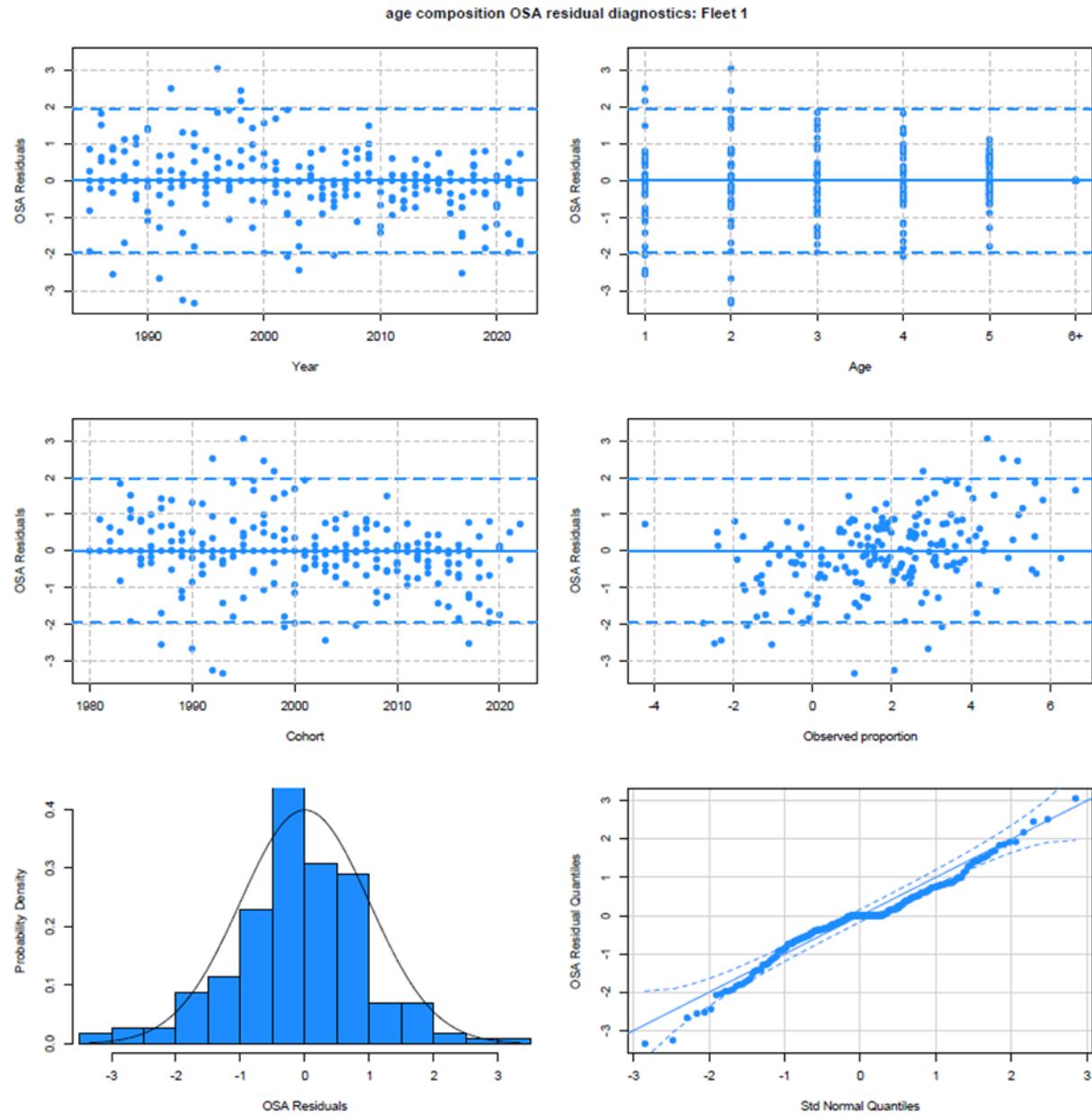


Figure 5.1.16: Candidate model m452 OSA residual diagnostic for the aggregate fleet age composition.

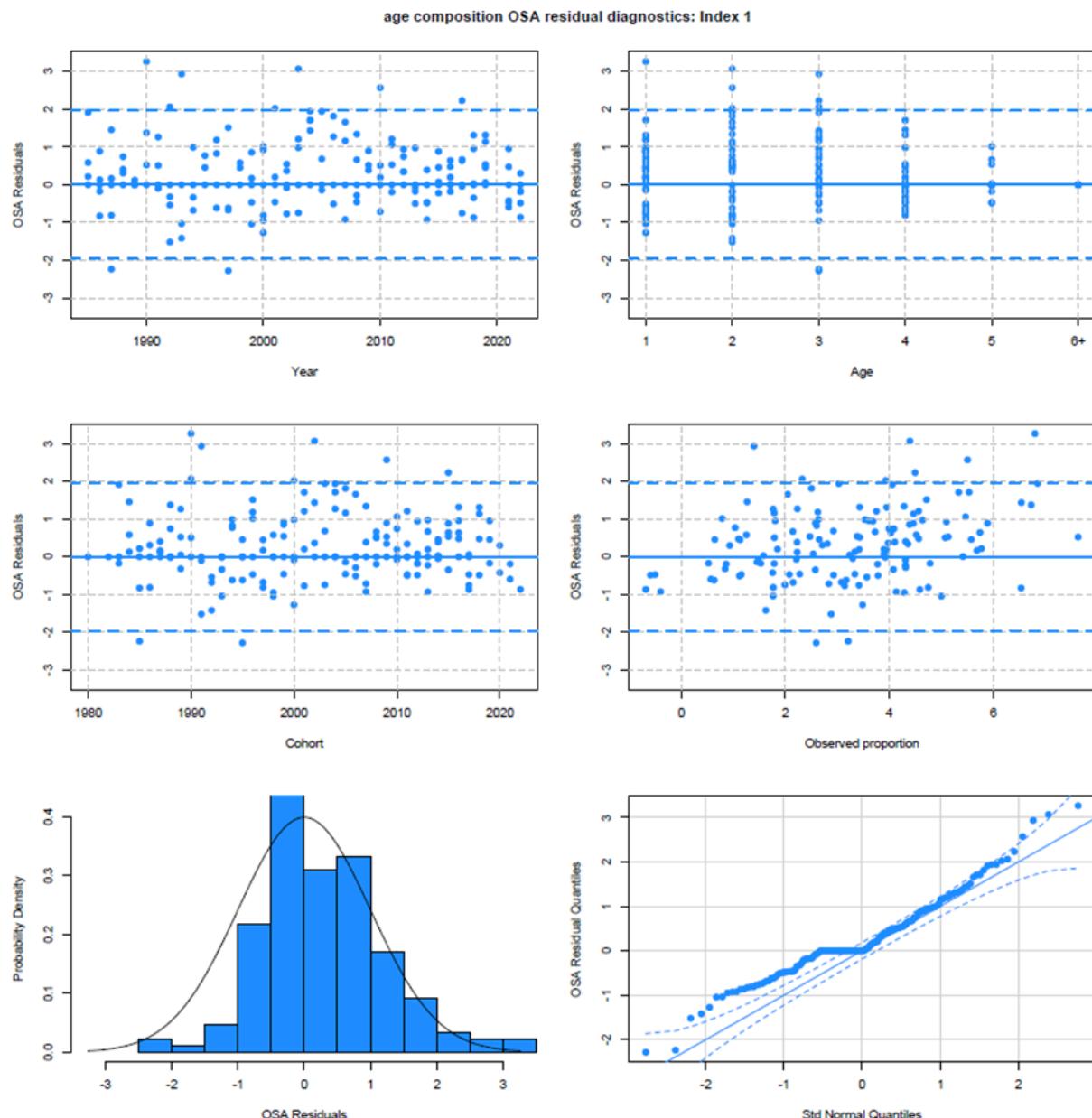


Figure 5.1.17: Candidate model m452 OSA residual diagnostic for the Inshore MADMF Fall bottom trawl index age composition.

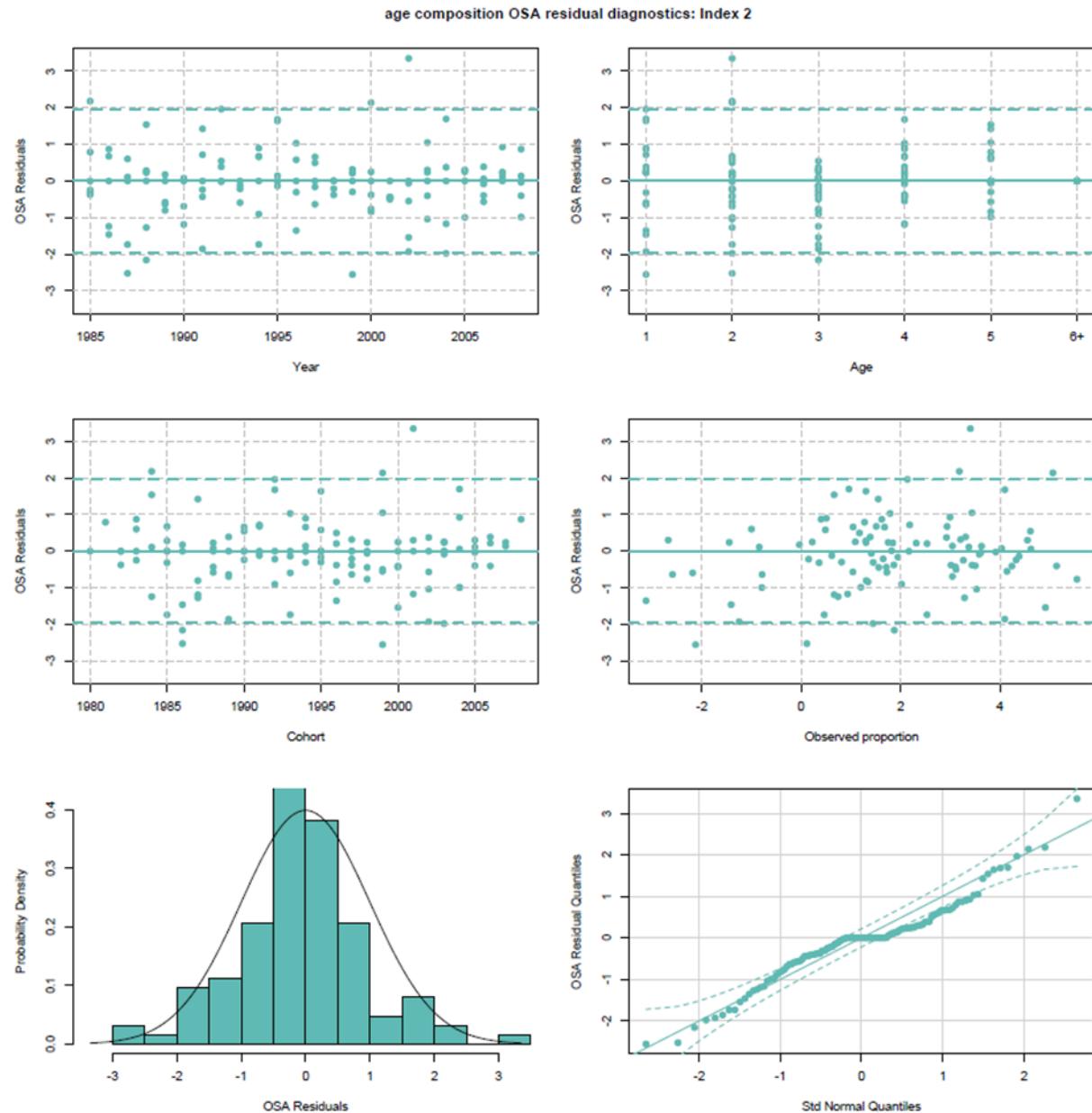


Figure 5.1.18: Candidate model m452 OSA residual diagnostic for the NEFSC Spring bottom trawl index age composition for the Albatross vessel (1985-2008).

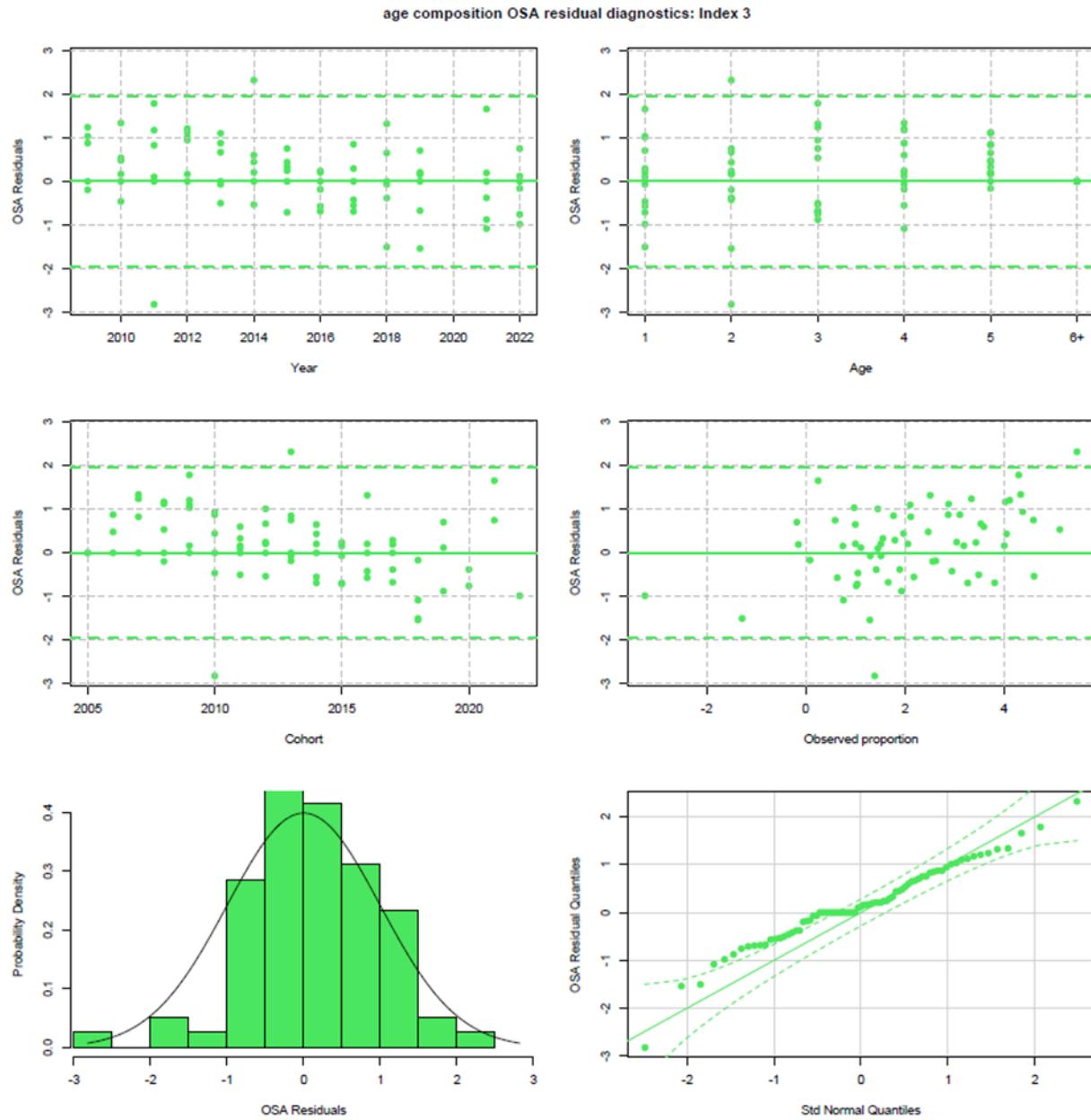


Figure 5.1.19: Candidate model m452 OSA residual diagnostic for the NEFSC Spring bottom trawl index age composition for the Bigelow vessel (2009-2022).

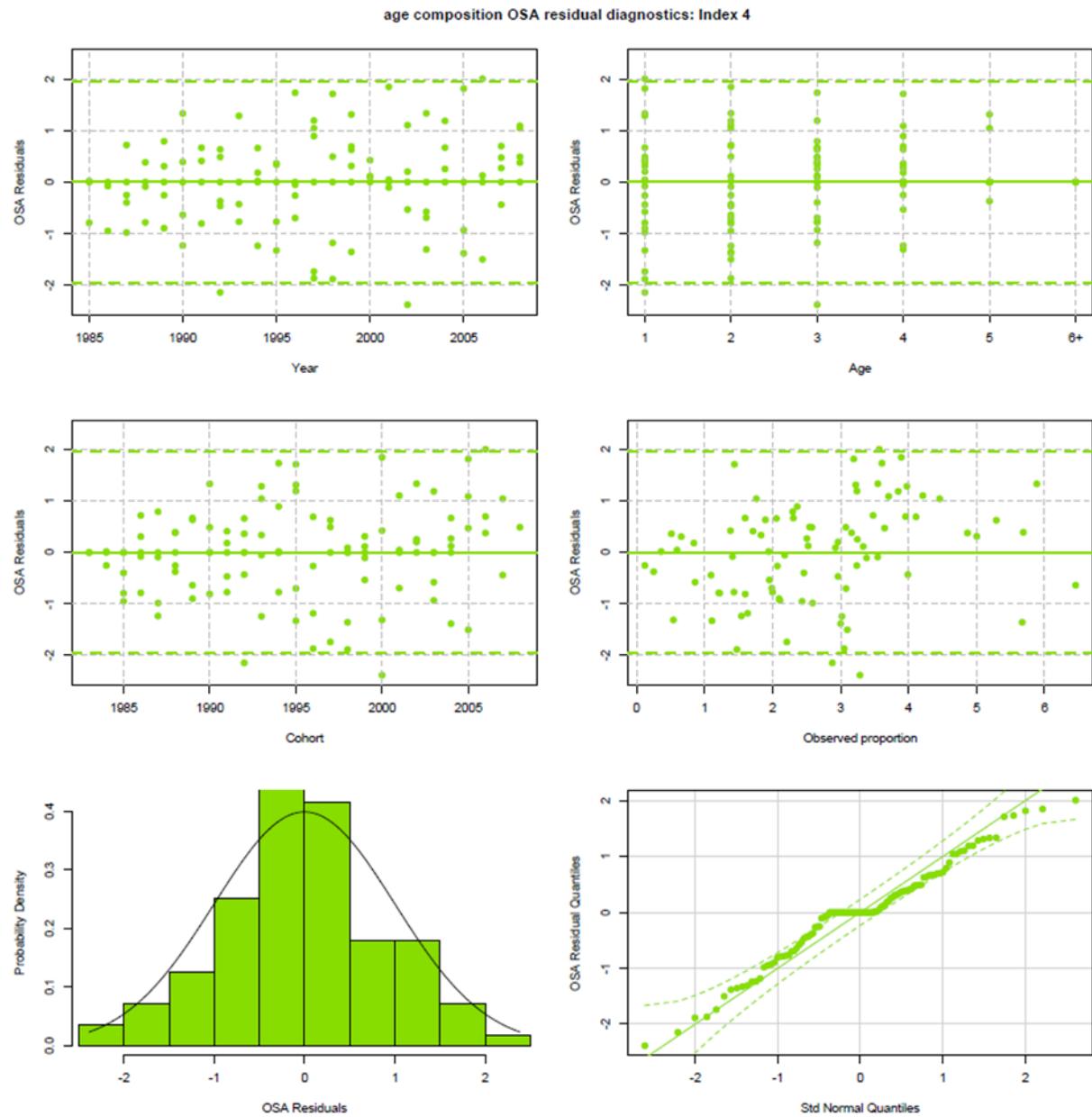


Figure 5.1.20: Candidate model m452 OSA residual diagnostic for the NEFSC Fall bottom trawl index age composition for the Albatross vessel (1985-2008).

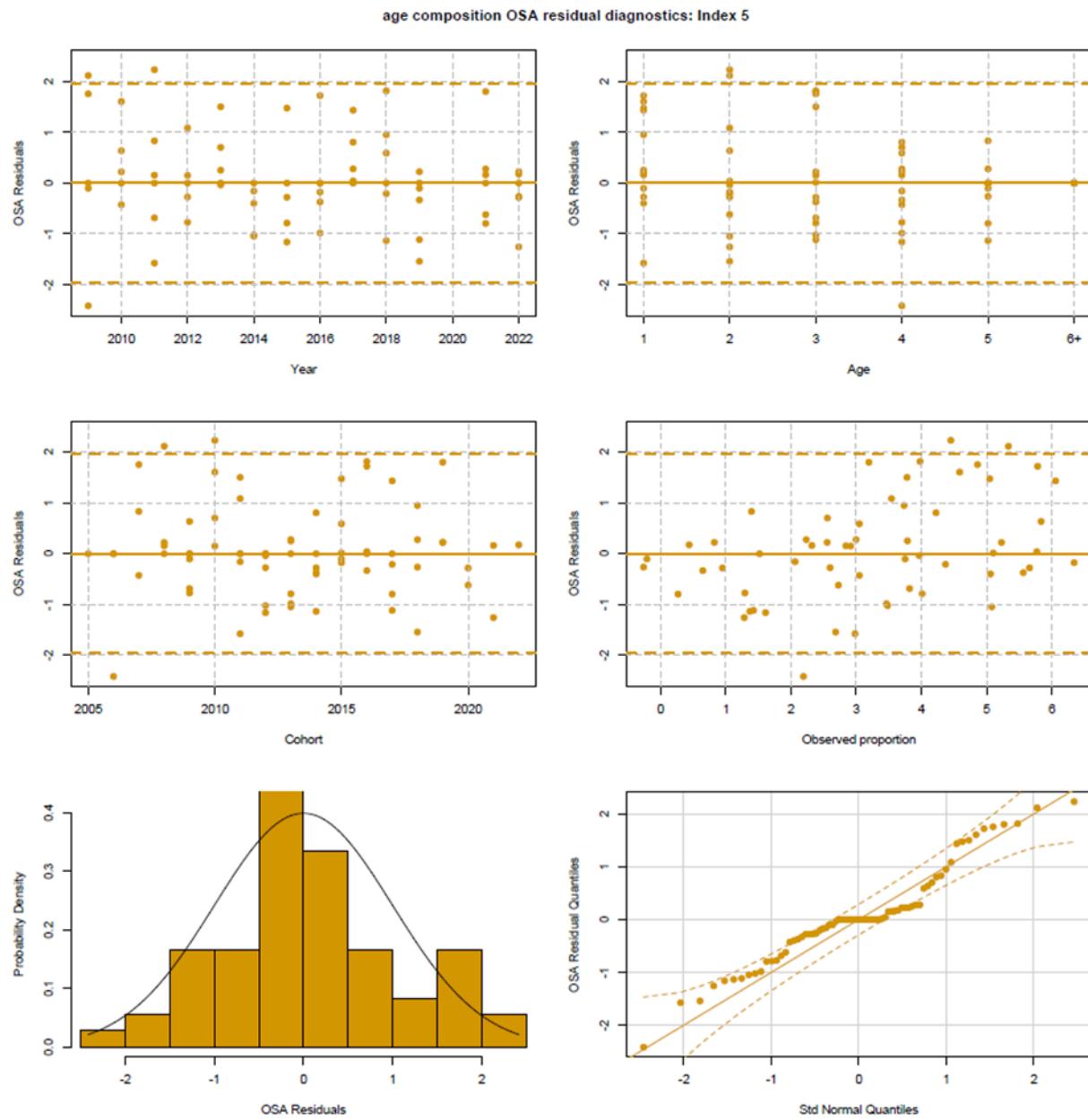


Figure 5.1.21: Candidate model m452 OSA residual diagnostic for the NEFSC Fall bottom trawl index age composition for the Bigelow vessel (2008-2022).

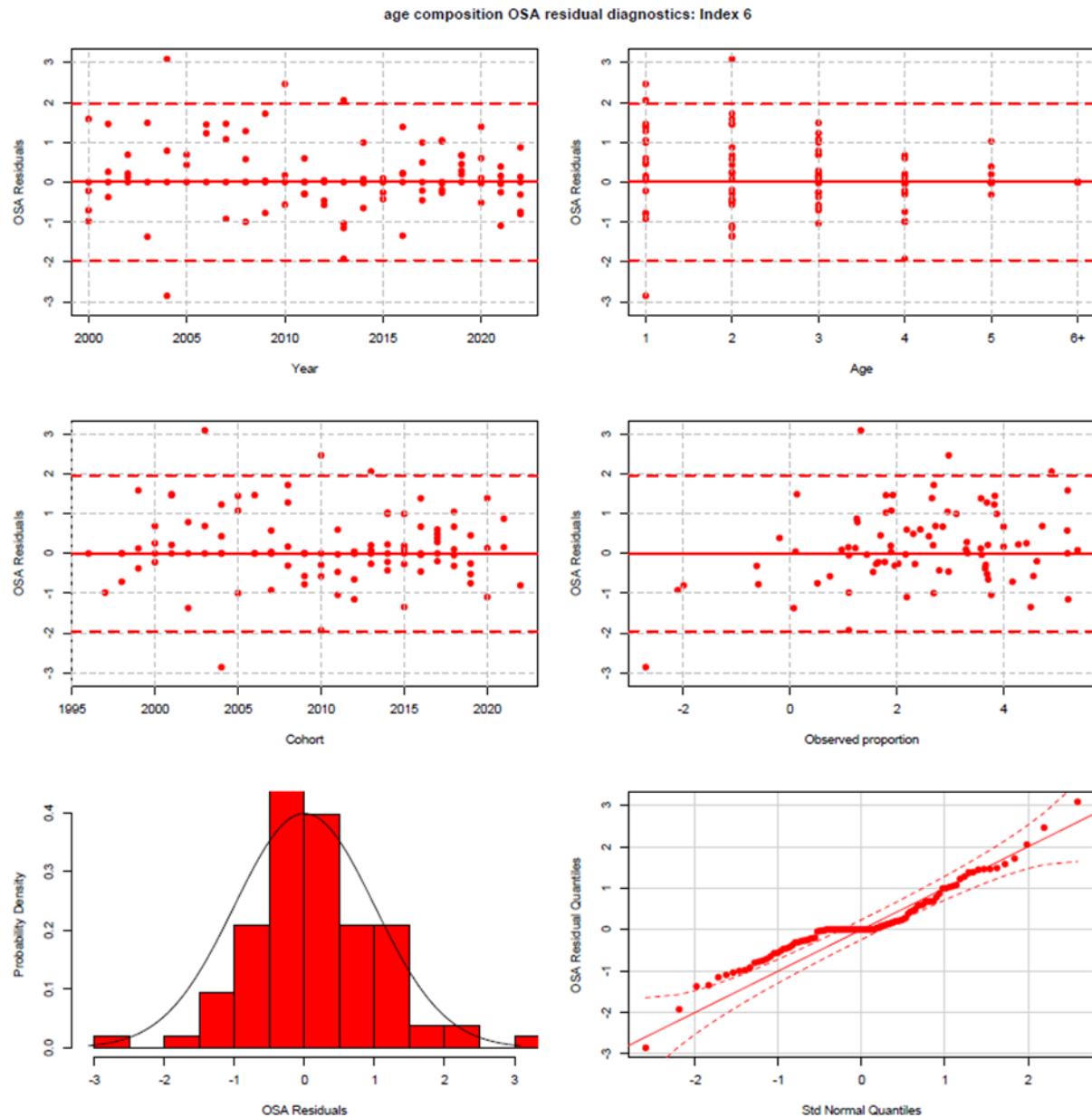


Figure 5.1.22: Candidate model m452 OSA residual diagnostic for the Inshore MENH Fall bottom trawl index age composition.

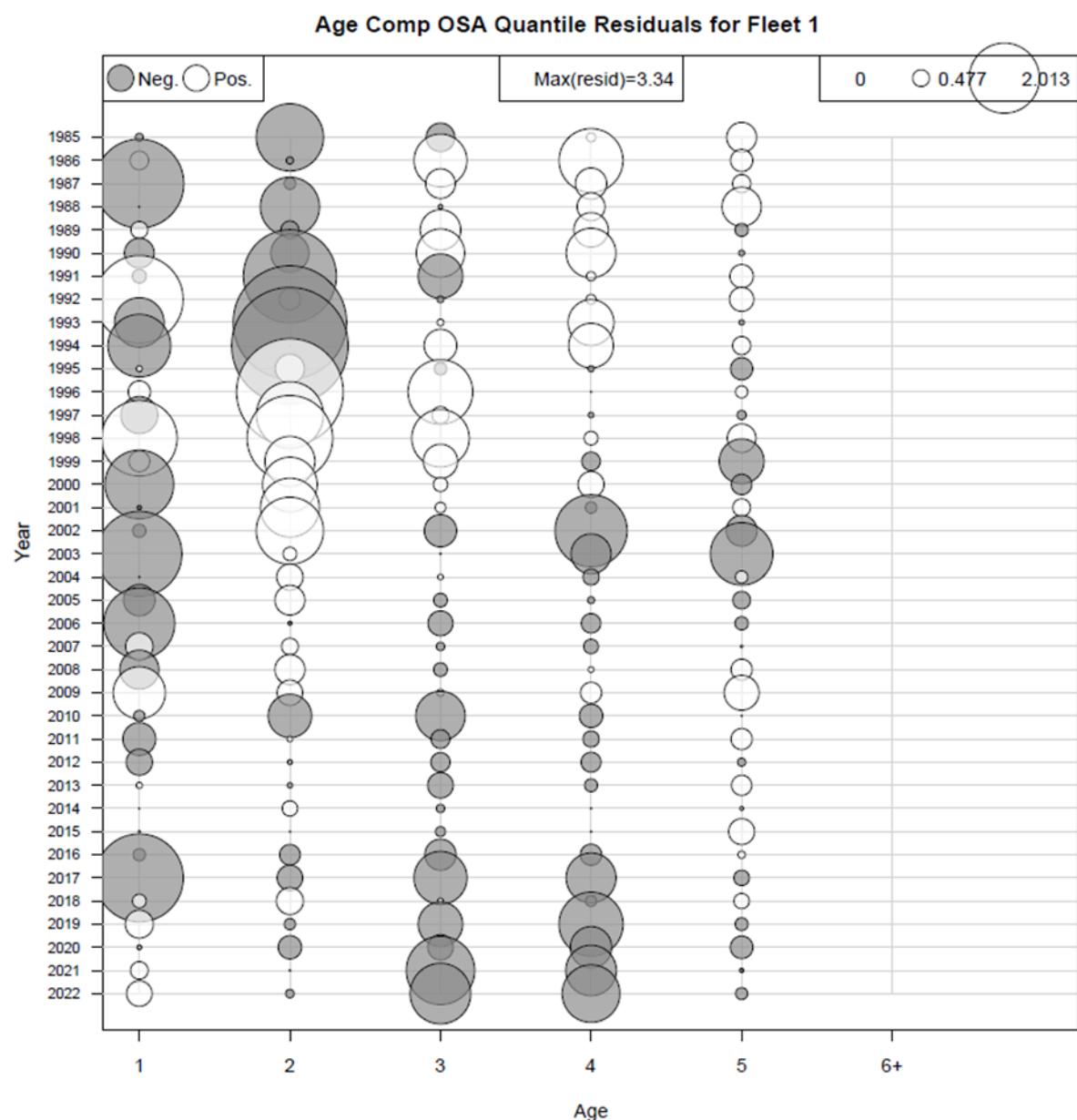


Figure 5.1.23: Candidate model m452 OSA residual bubble plot diagnostics for of the aggregate fleet.

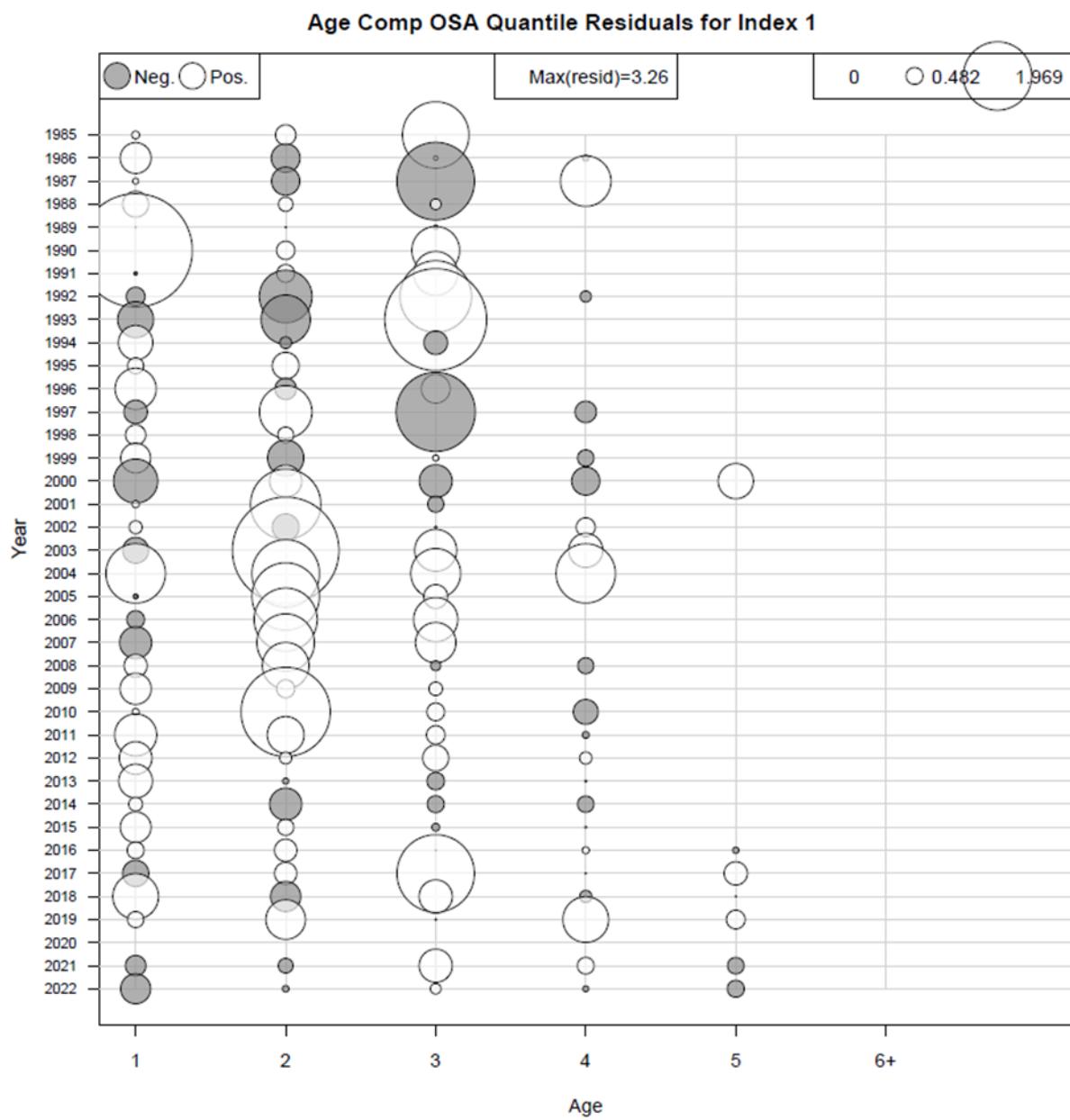


Figure 5.1.24: Bubble plot of OSA residual diagnostics for the Inshore MADMF Fall bottom trawl survey index age composition.

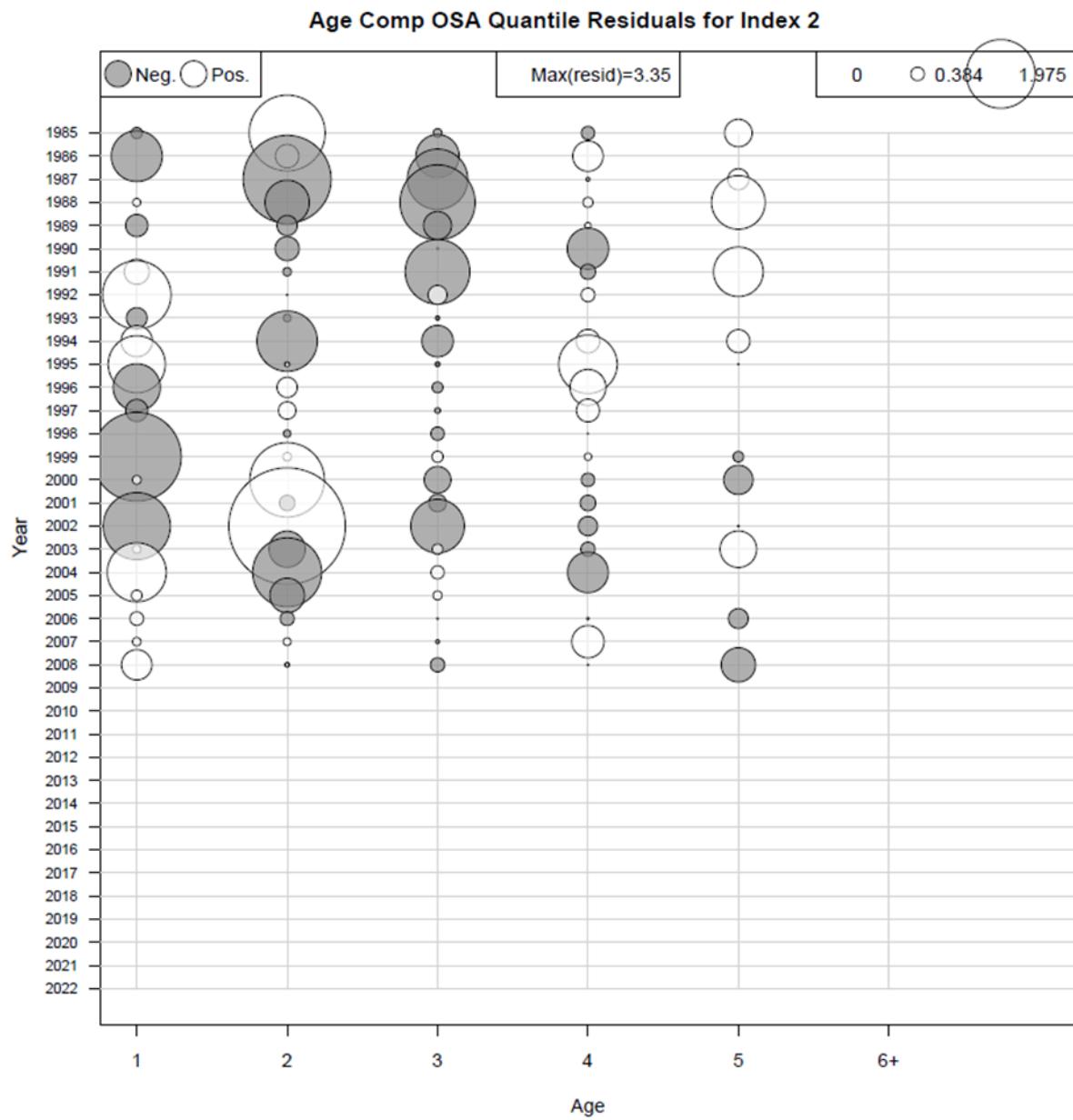


Figure 5.1.25: Bubble plot of OSA residual diagnostics for the NEFSC Spring bottom trawl survey index age composition from the Albatross vessel (1985-2008).

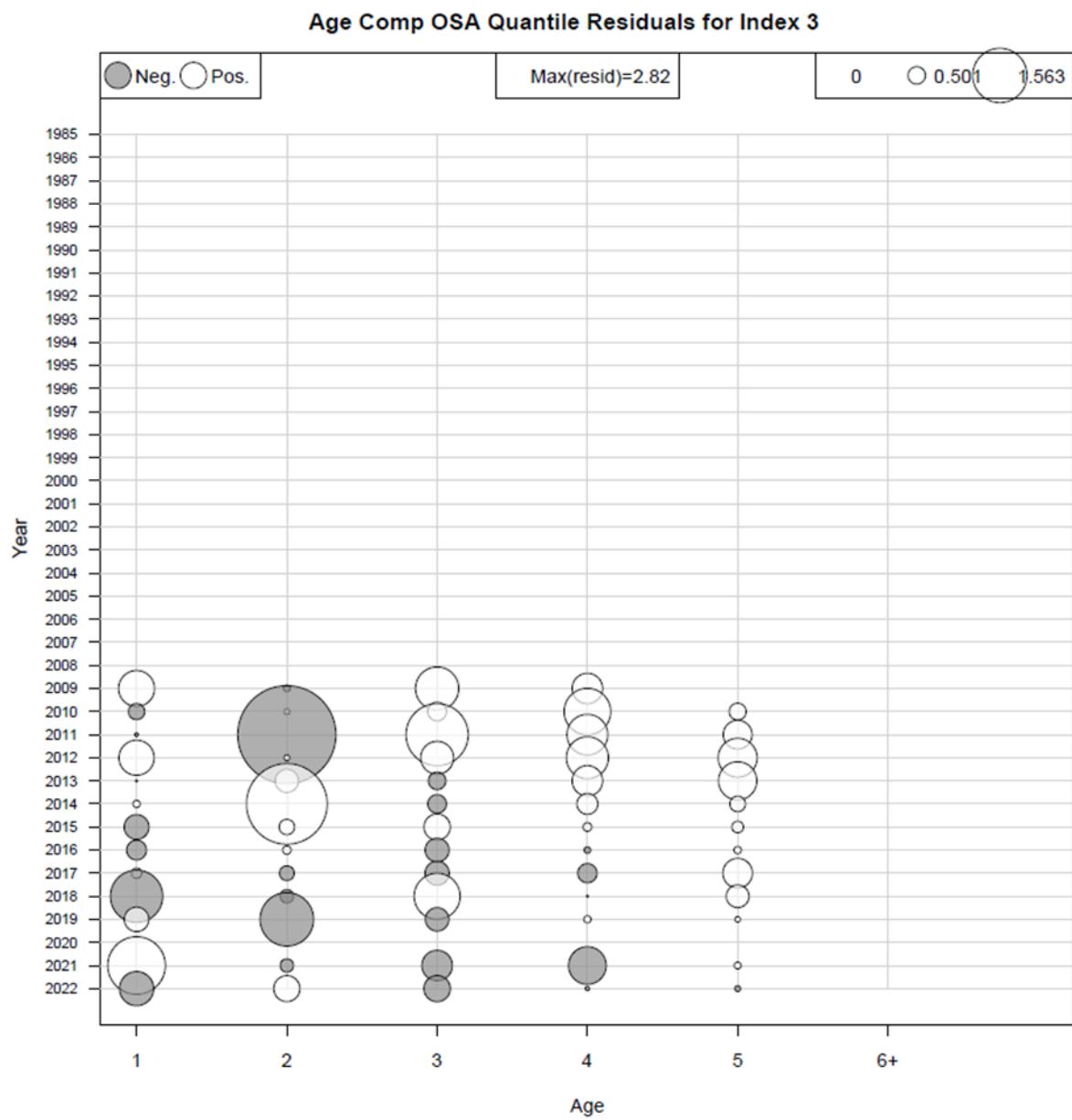


Figure 5.1.26: Bubble plot of OSA residual diagnostics for the NEFSC Spring bottom trawl survey index age composition from the Bigelow vessel (2009-2022).

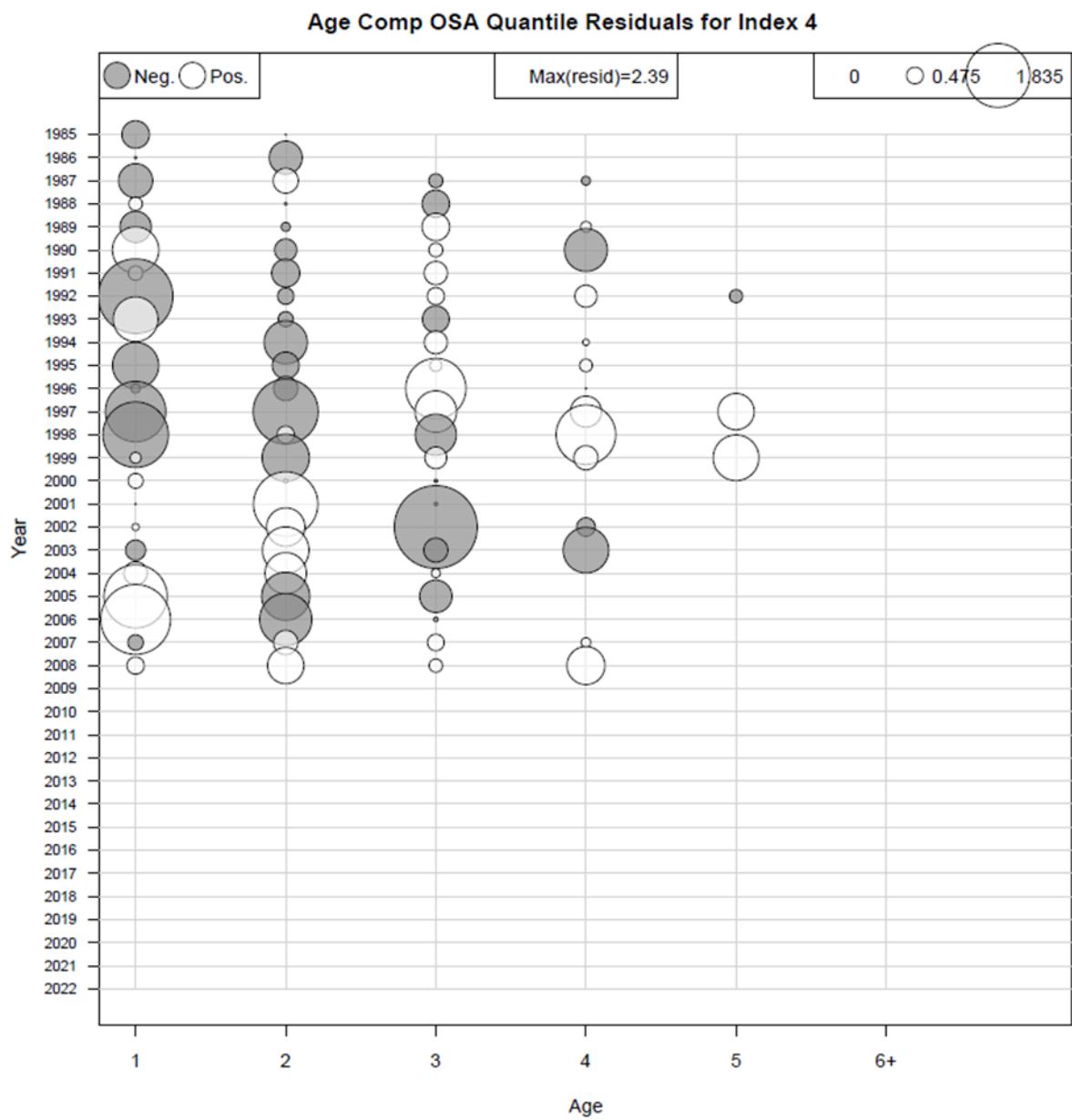


Figure 5.1.27: Bubble plot of OSA residual diagnostics for the NEFSC Fall bottom trawl survey index age composition from the Albatross vessel (1985-2008).

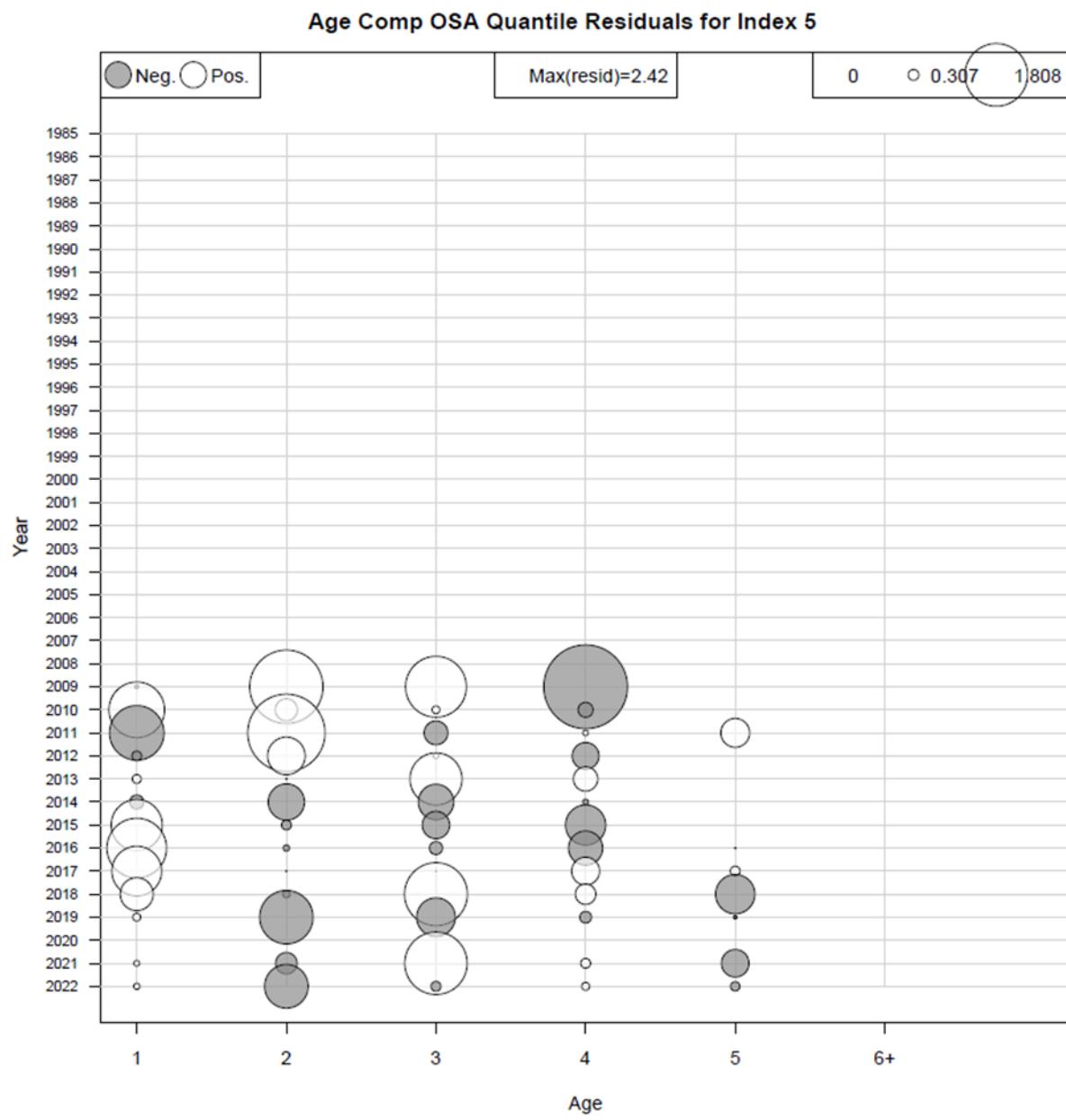


Figure 5.1.28: Bubble plot of OSA residual diagnostics for the NEFSC Fall bottom trawl survey index age composition from the Bigelow vessel (2009-2022).

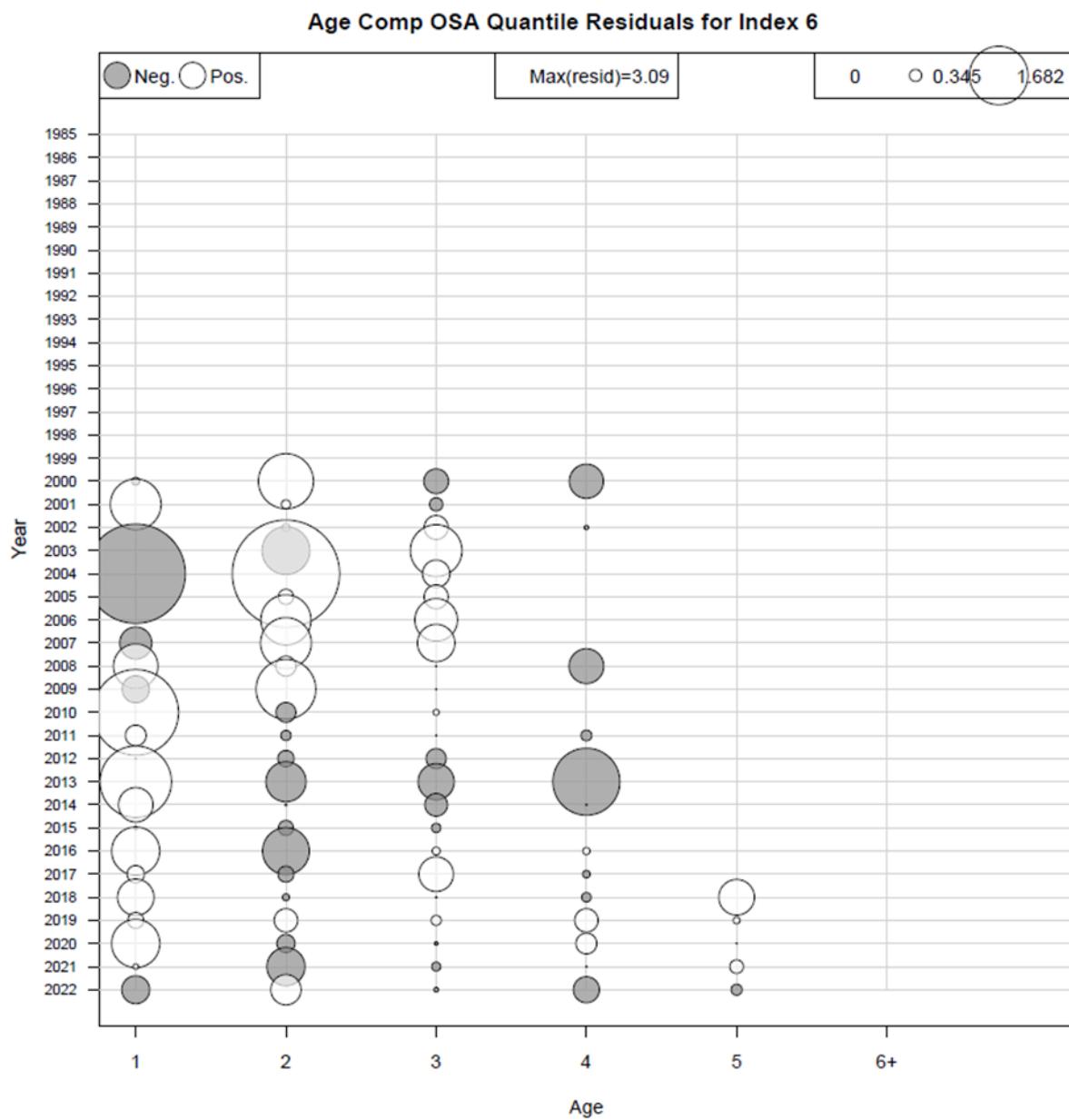


Figure 5.1.29: Bubble plot of OSA residual diagnostics for the Inshore MENH Fall bottom trawl survey index age composition.

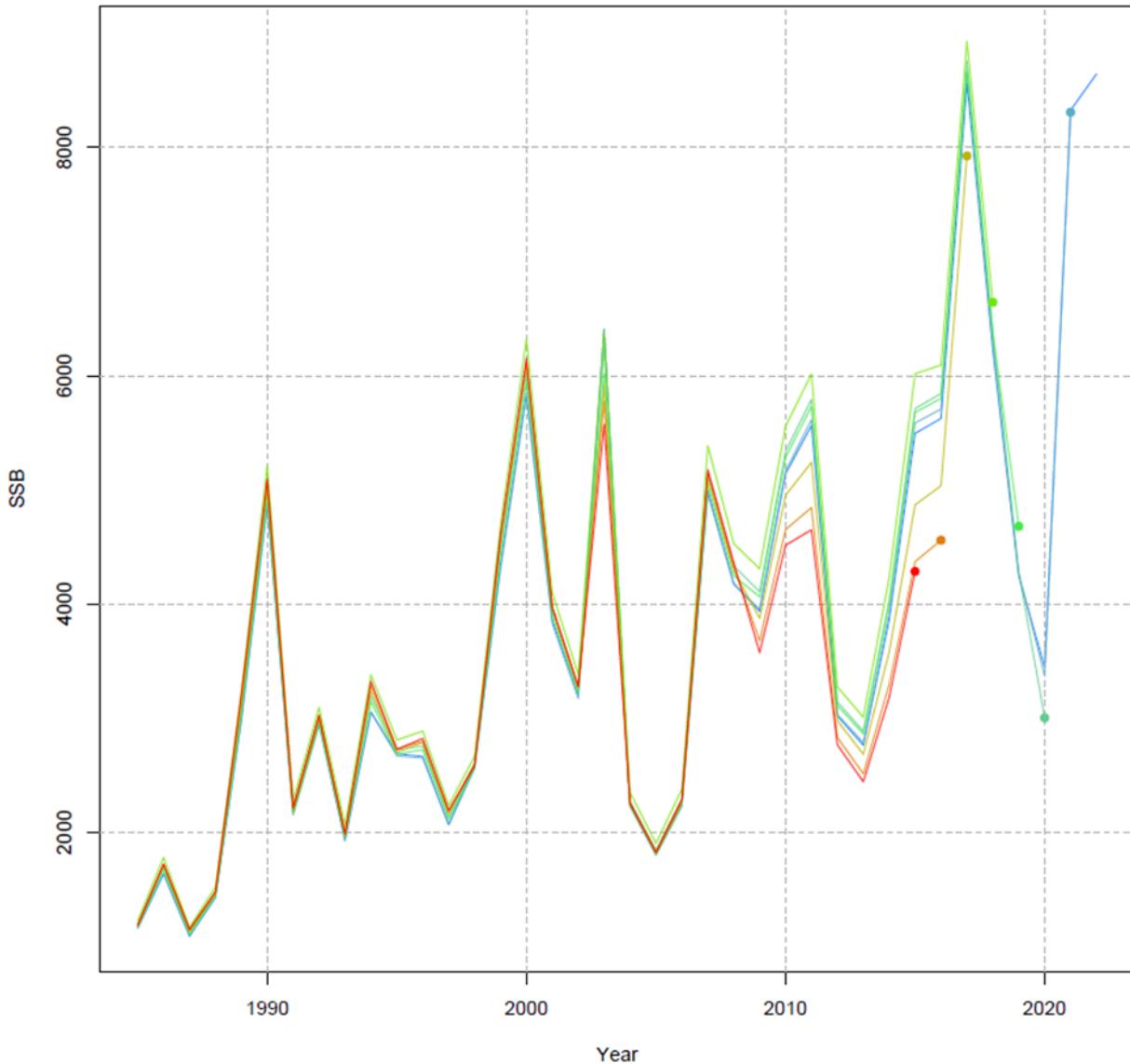


Figure 5.1.30: Retrospective pattern of SSB for the candidate model m452.

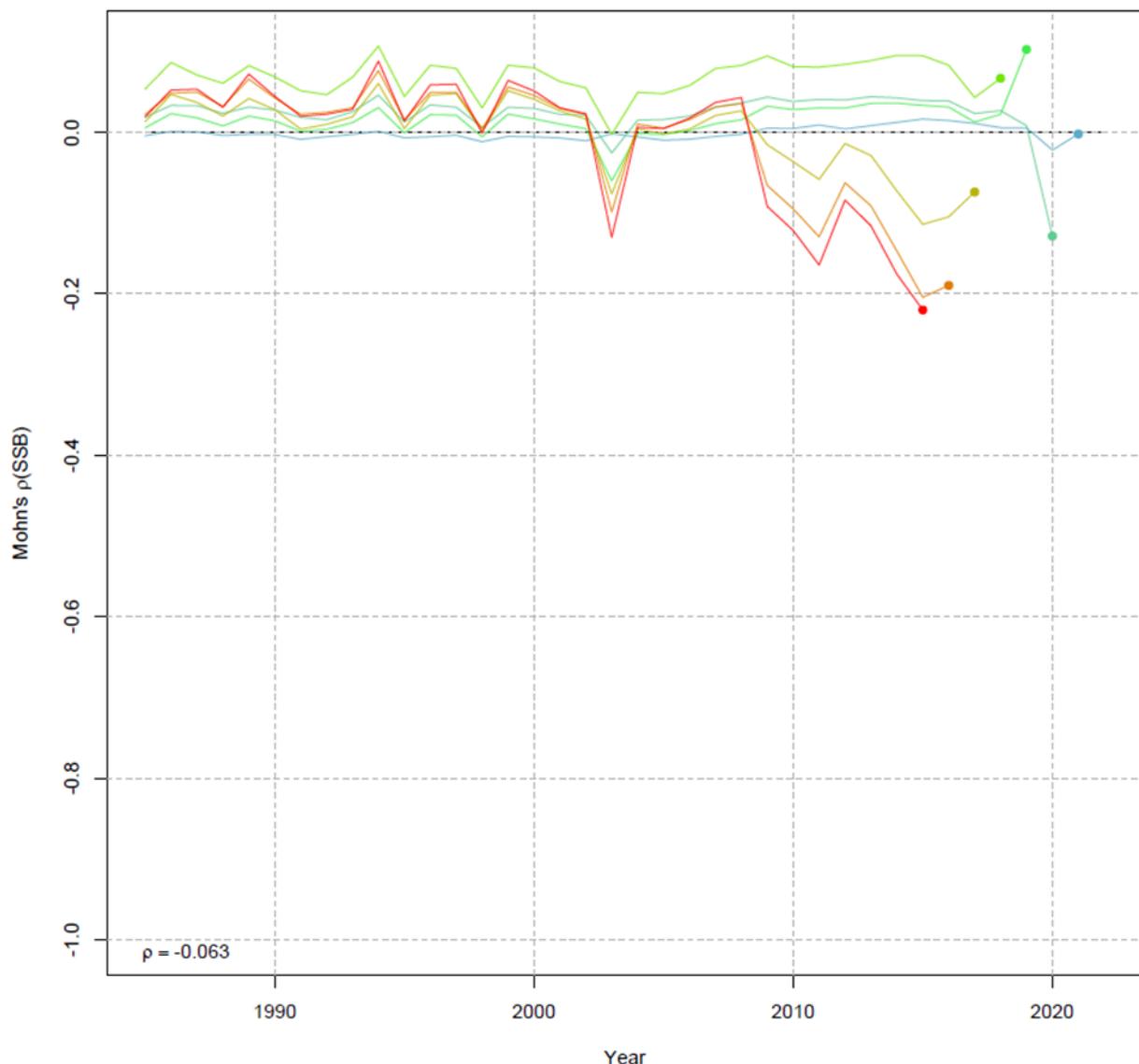


Figure 5.1.31: Relative retrospective pattern of SSB for the candidate model m452.

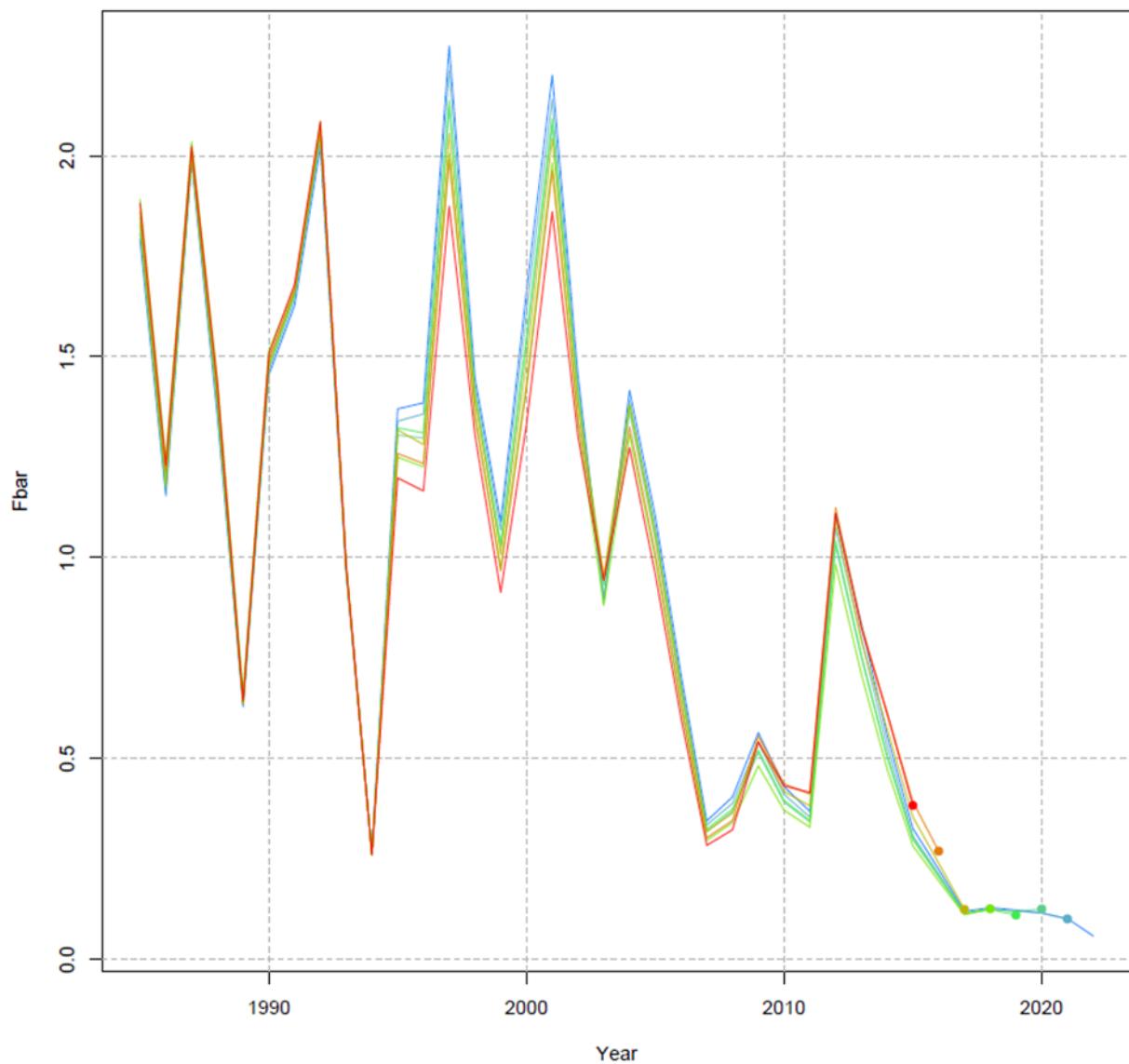


Figure 5.1.32: Retrospective pattern of F for the candidate model m452.

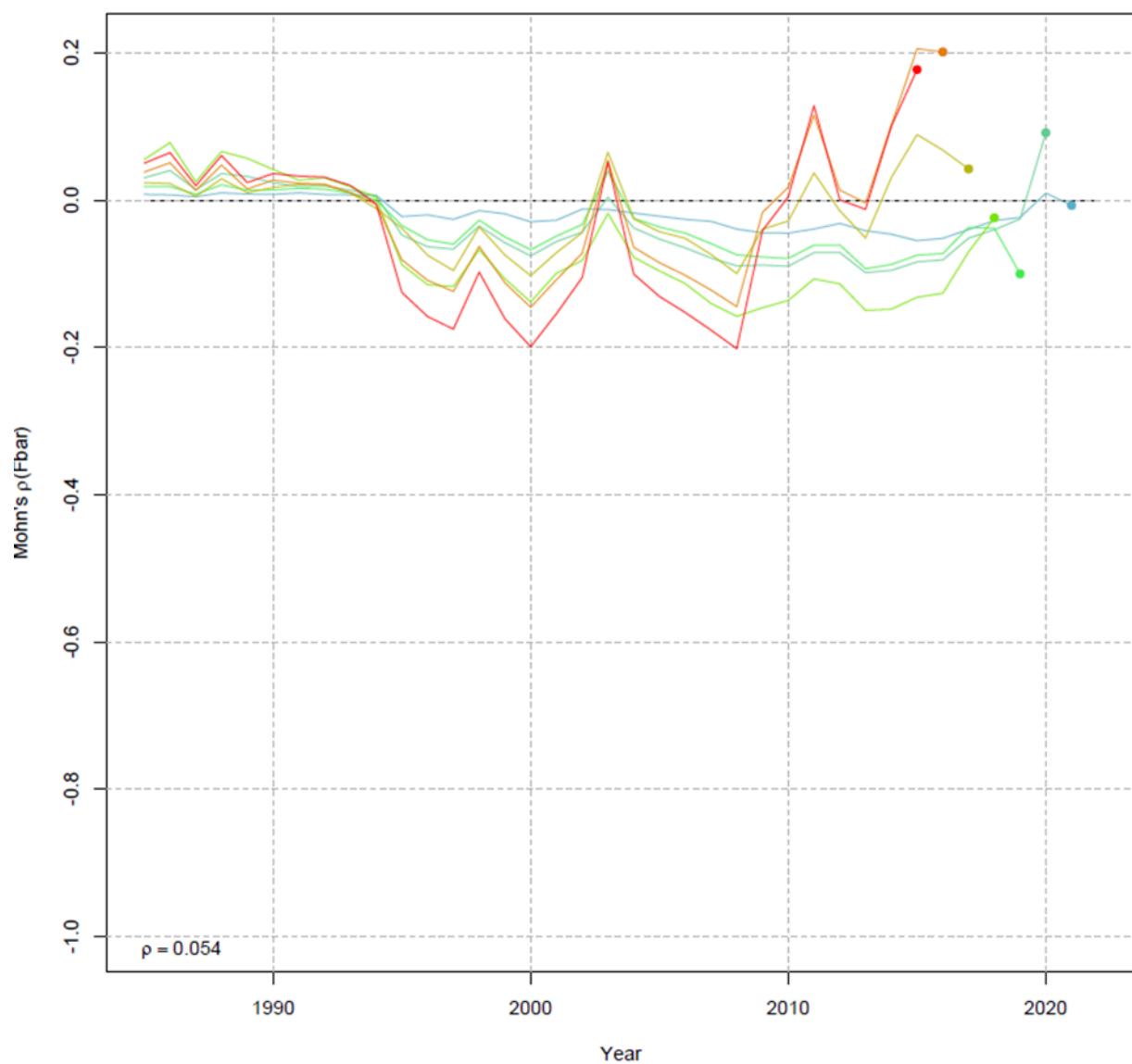


Figure 5.1.33: Relative retrospective pattern of F for the candidate model m452.

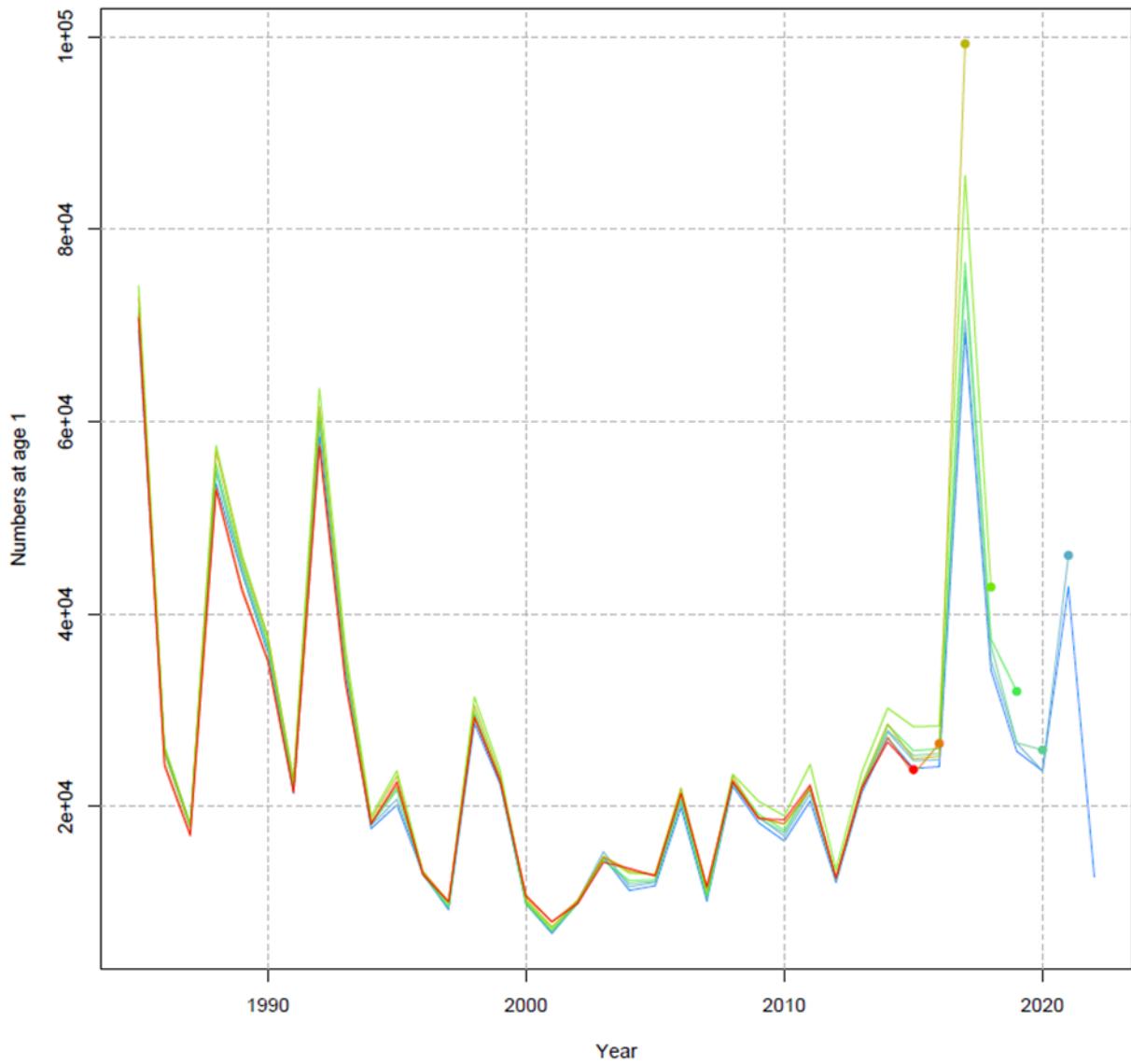


Figure 5.1.34: Retrospective pattern of age-1 recruitment for the candidate model m452.

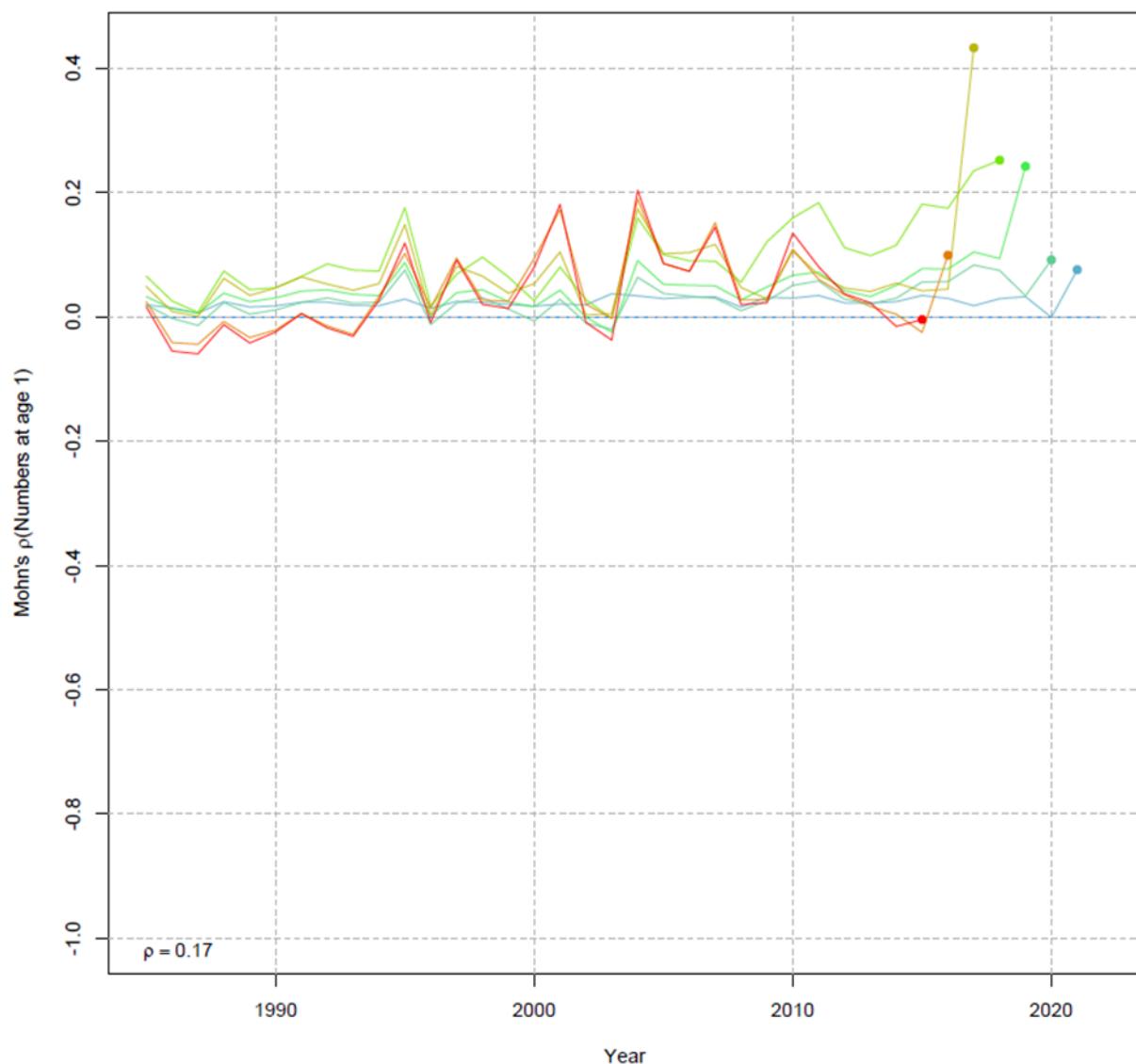


Figure 5.1.35: Relative retrospective pattern of age-1 recruitment for candidate model m452.

Self-test mean (blue) and median bias with 5%, 25%, 75%, 95% quantiles

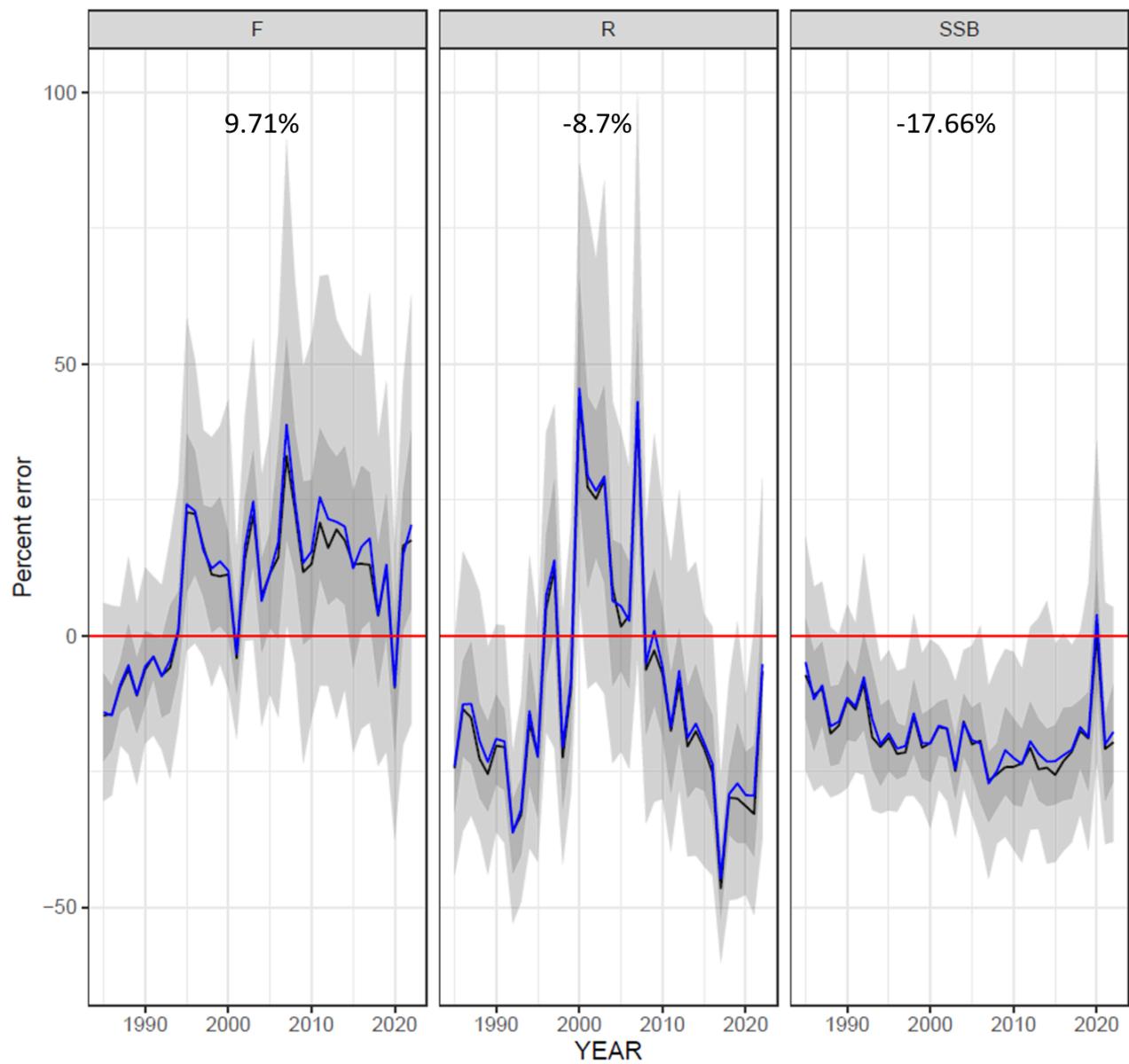


Figure 5.1.36: Self-test results of the candidate model m452.

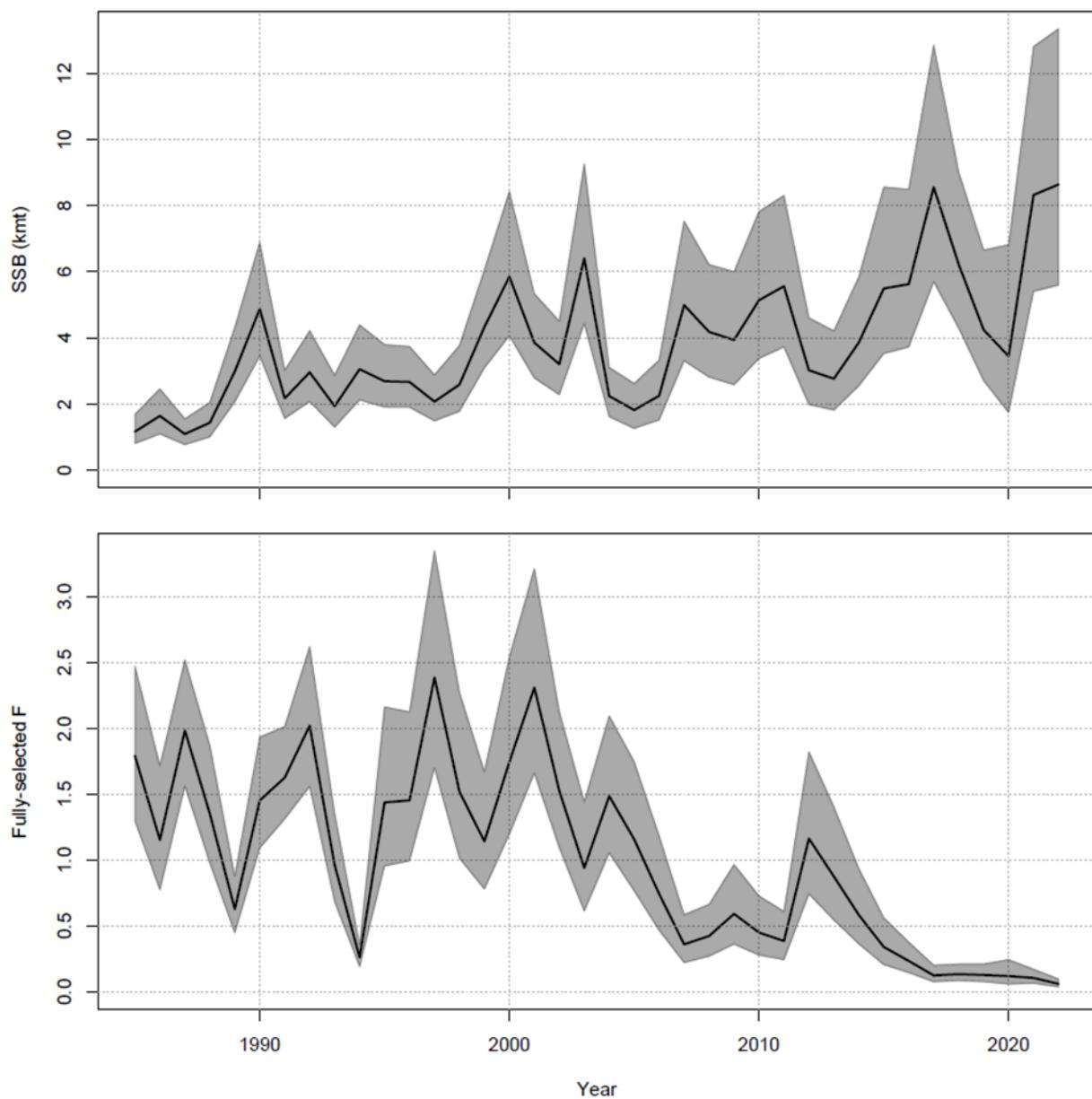


Figure 5.2.1: SSB (top) and F (bottom) from the candidate model m452.

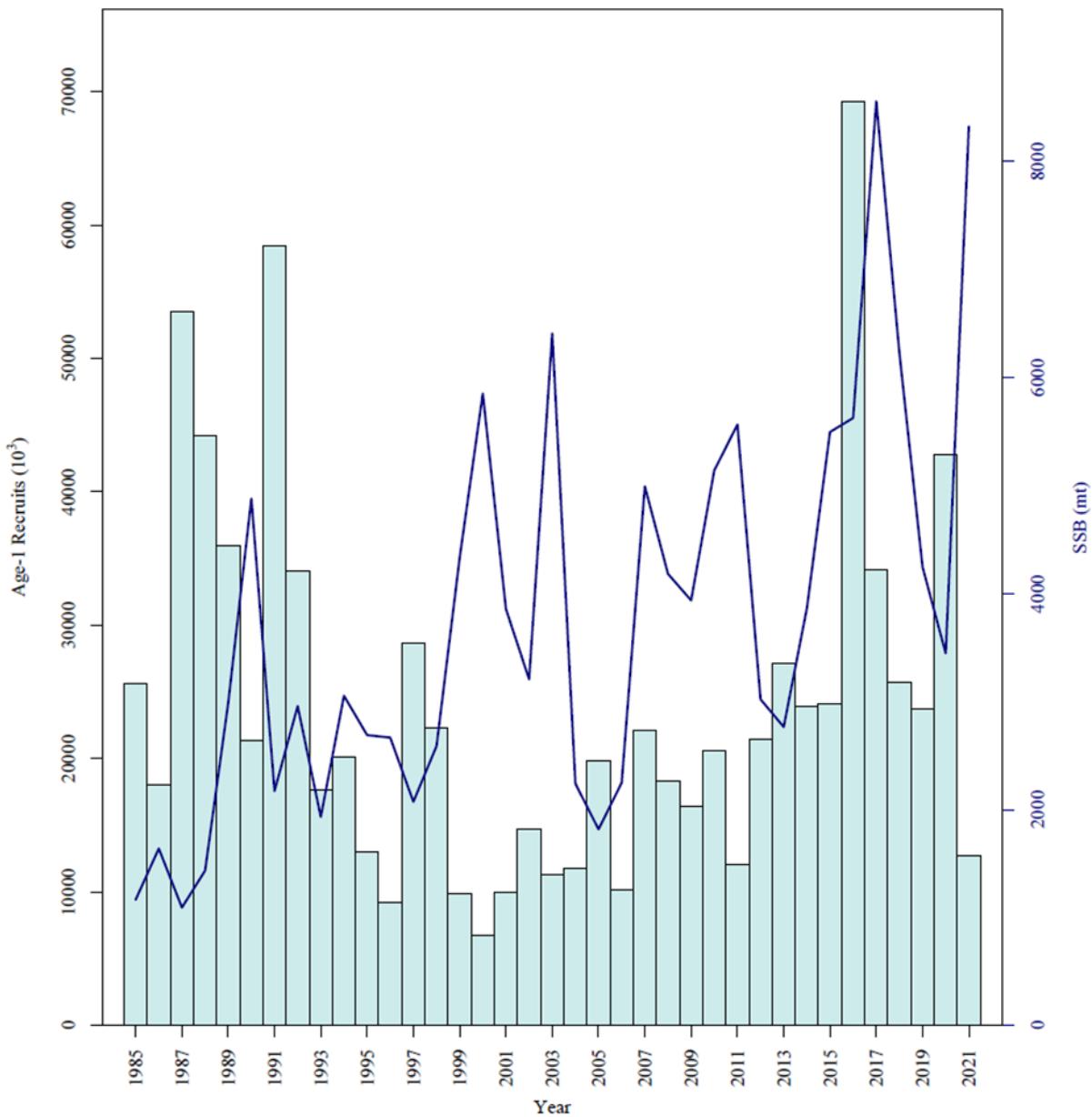


Figure 5.2.2: SSB (line) and age-1 recruitment (bars) from the candidate model m452.

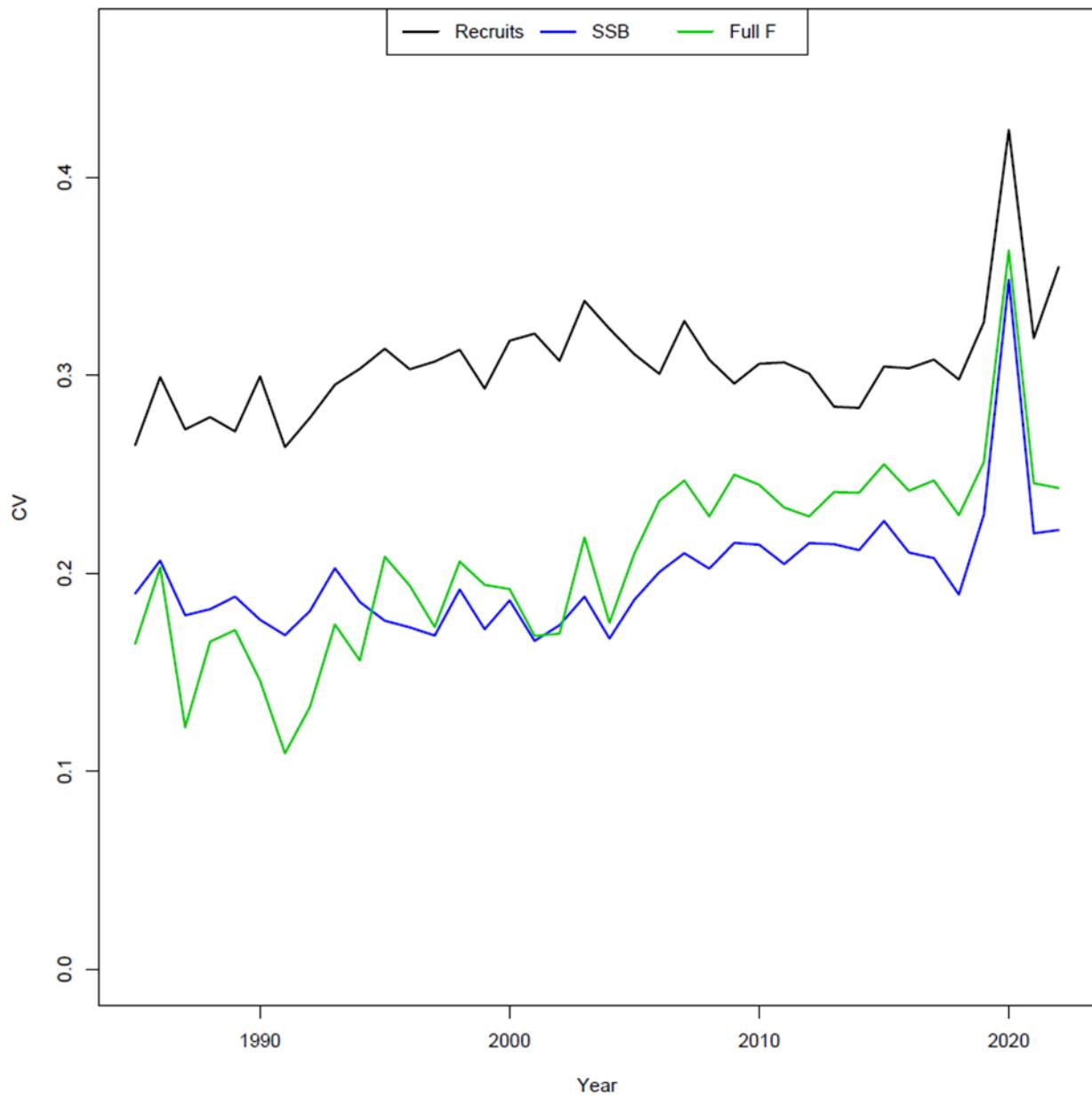


Figure 5.2.3: CV's of SSB, F and R from the candidate model m452.

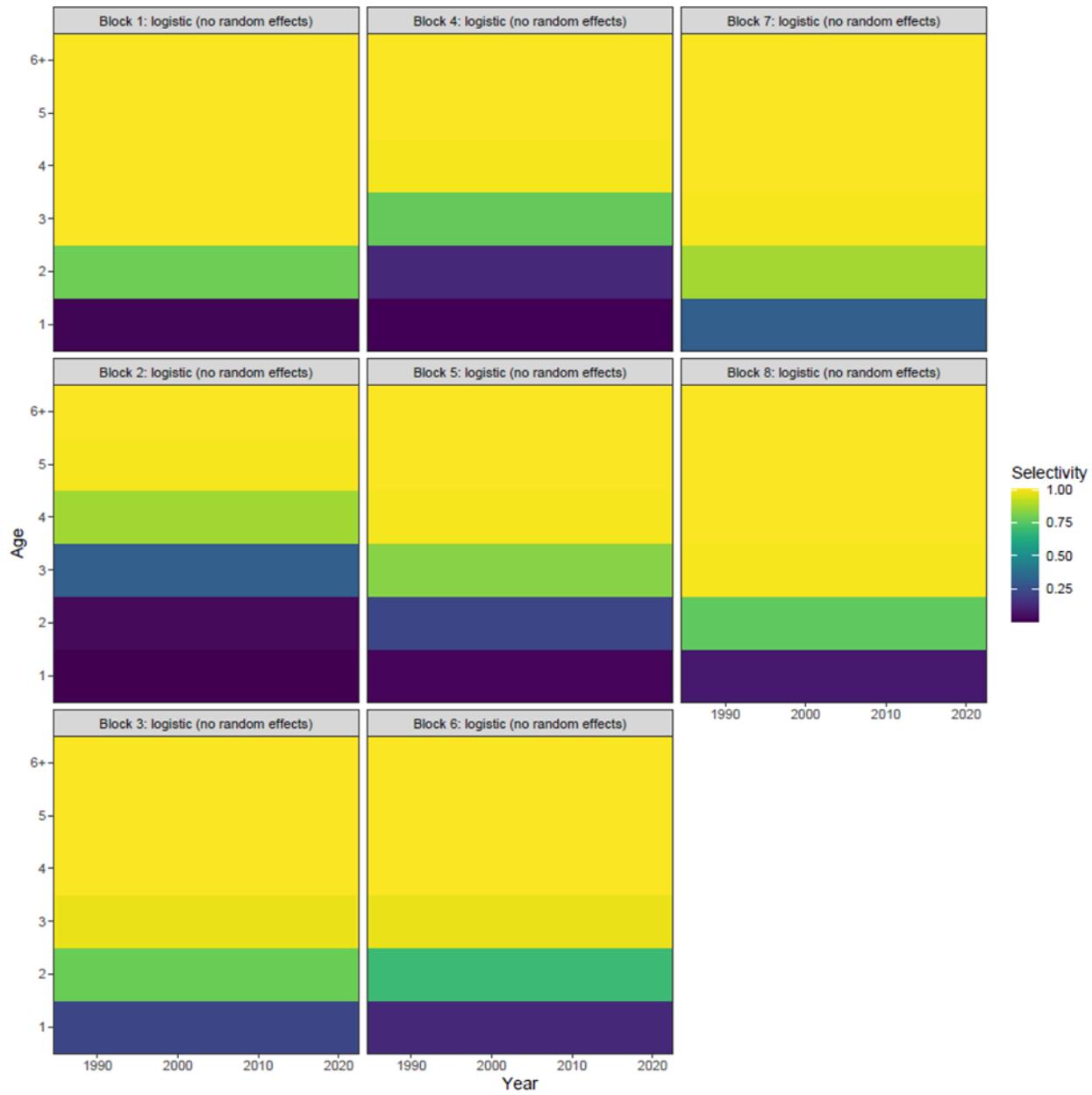


Figure 5.2.4: Aggregate fleet 1985-1993 Sel. block (Block1), Aggregate fleet 1994-2022 block (Block 2), MADMF Fall (Block 3), NEFSC spring Albatross (Block 4), NEFSC spring Bigelow (Block 5), NEFSC fall Albatross (Block 6), NEFSC fall Bigelow (Block 7) and MENH fall (Block 8) selectivities-at-age from the candidate model m452.

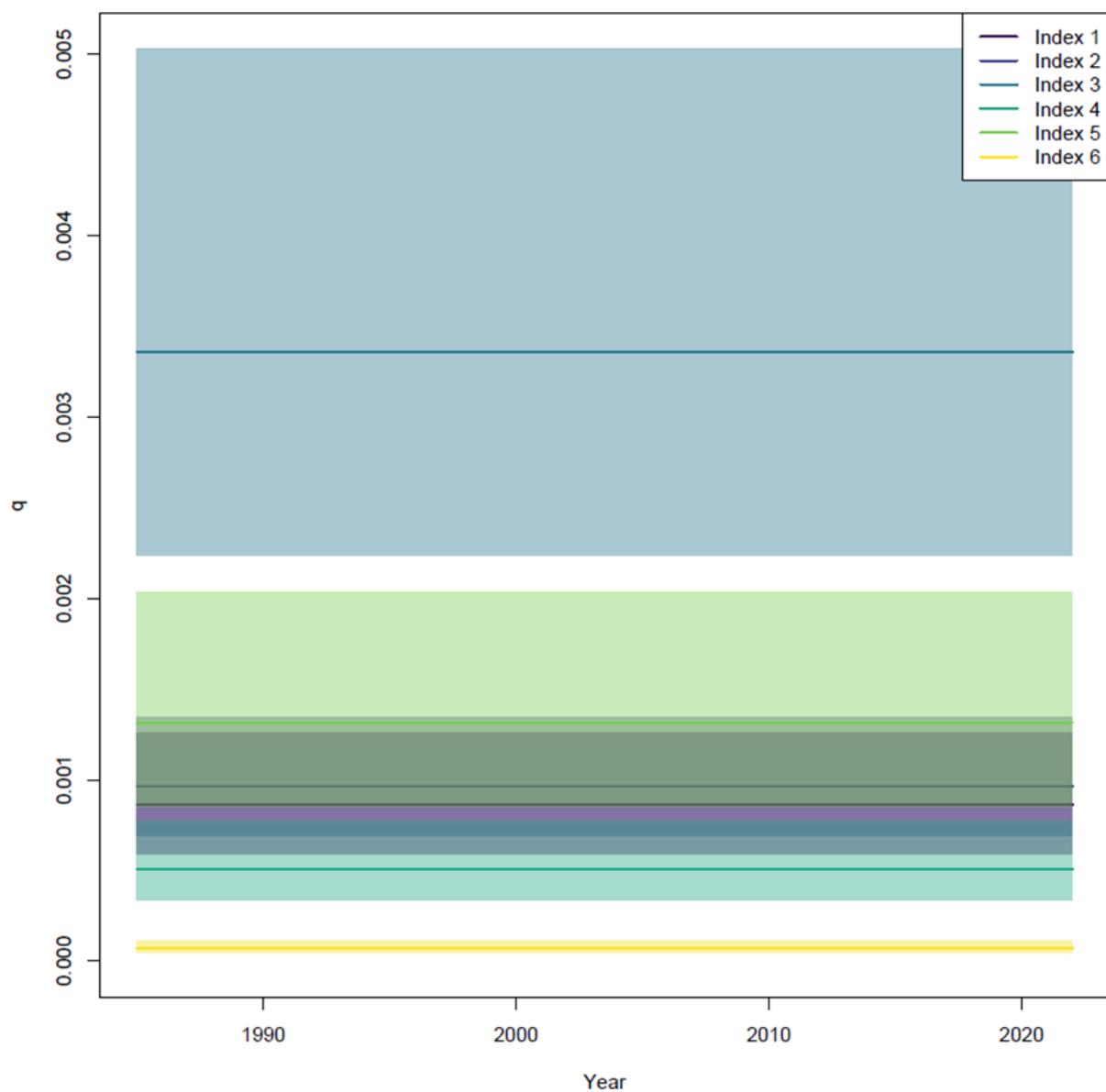


Figure 5.2.5: MADMF Fall (Index1), NEFSC spring Albatross (Index 2), NEFSC spring Bigelow (Index 3), NEFSC fall Albatross (Index 4), NEFSC fall Bigelow (Index 5) and MENH fall (Index 6) catchabilities from the candidate model m452.

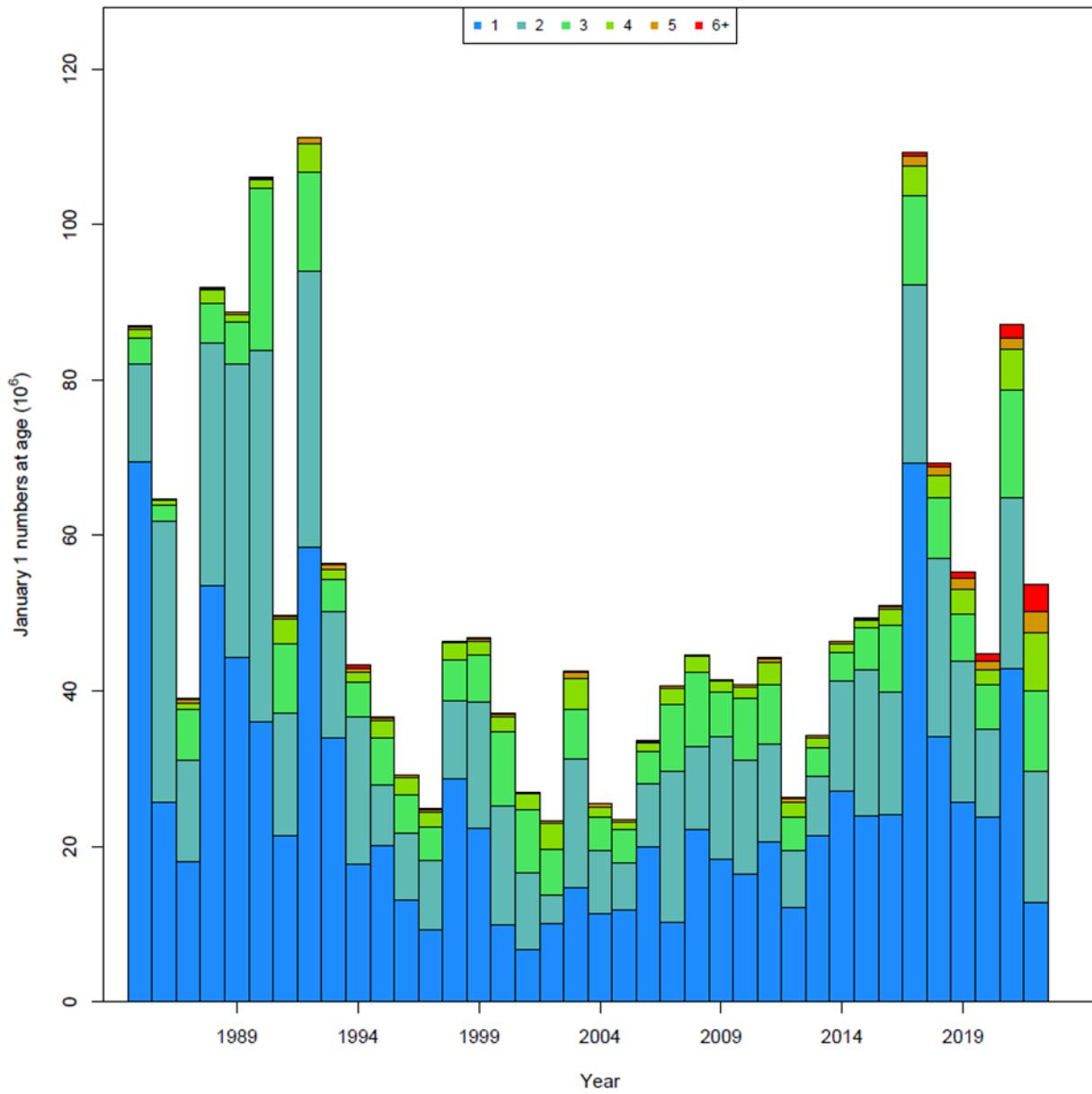


Figure 5.2.6: January 1st NAA from candidate model m452.

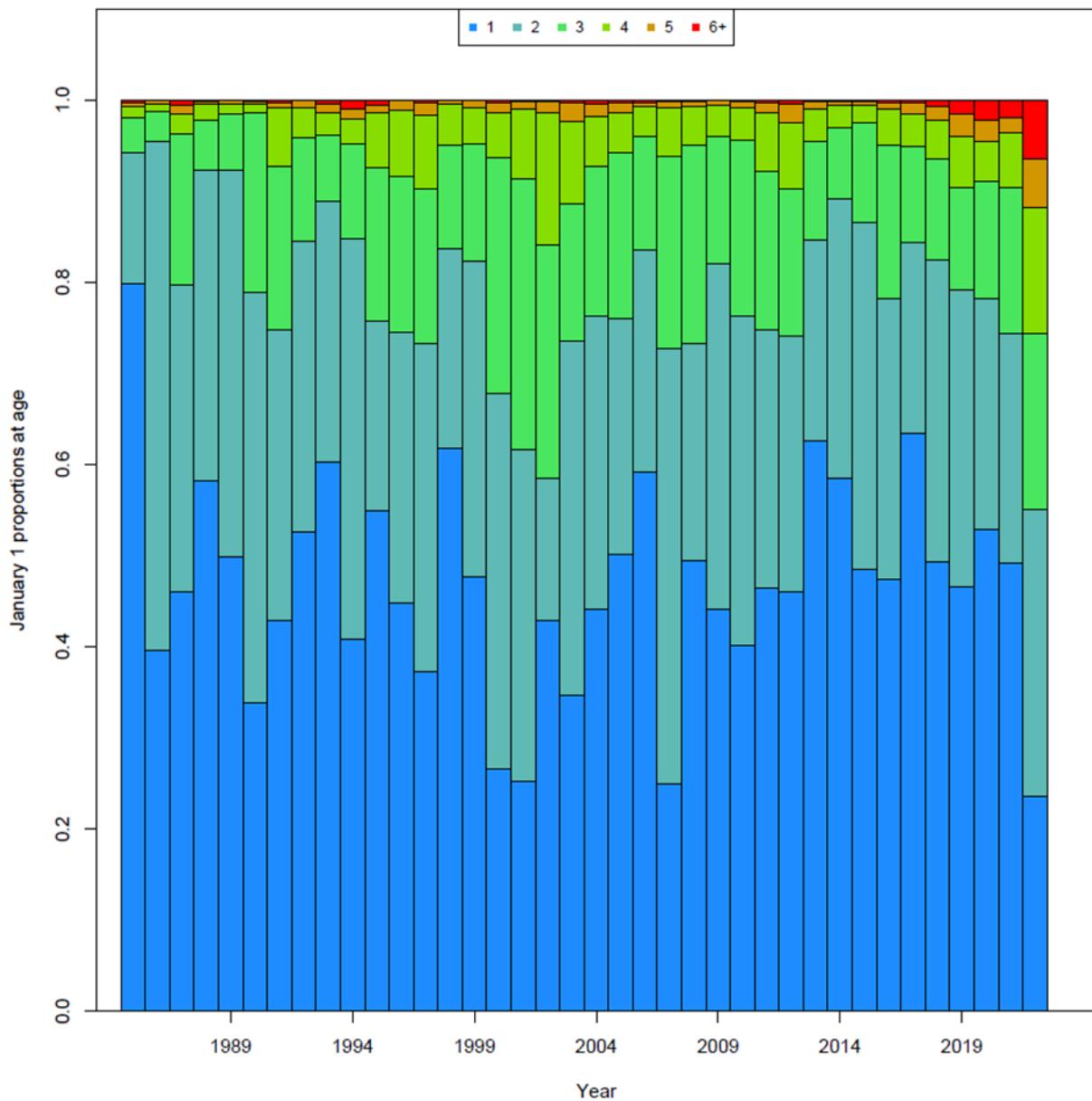


Figure 5.2.7: Proportion January 1st NAA from the candidate model m452.

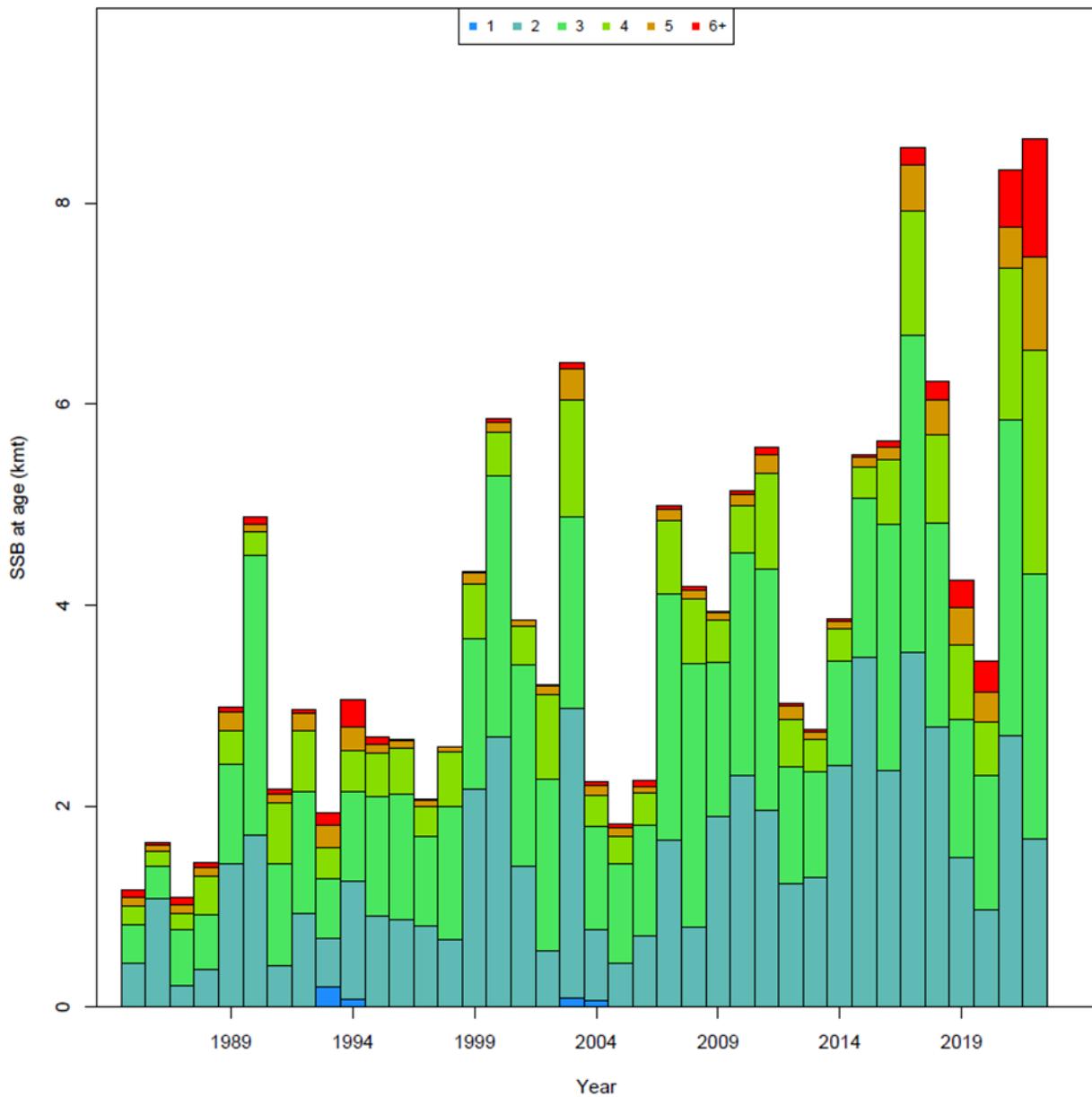


Figure 5.2.8: SSB-at-age from candidate model m452.

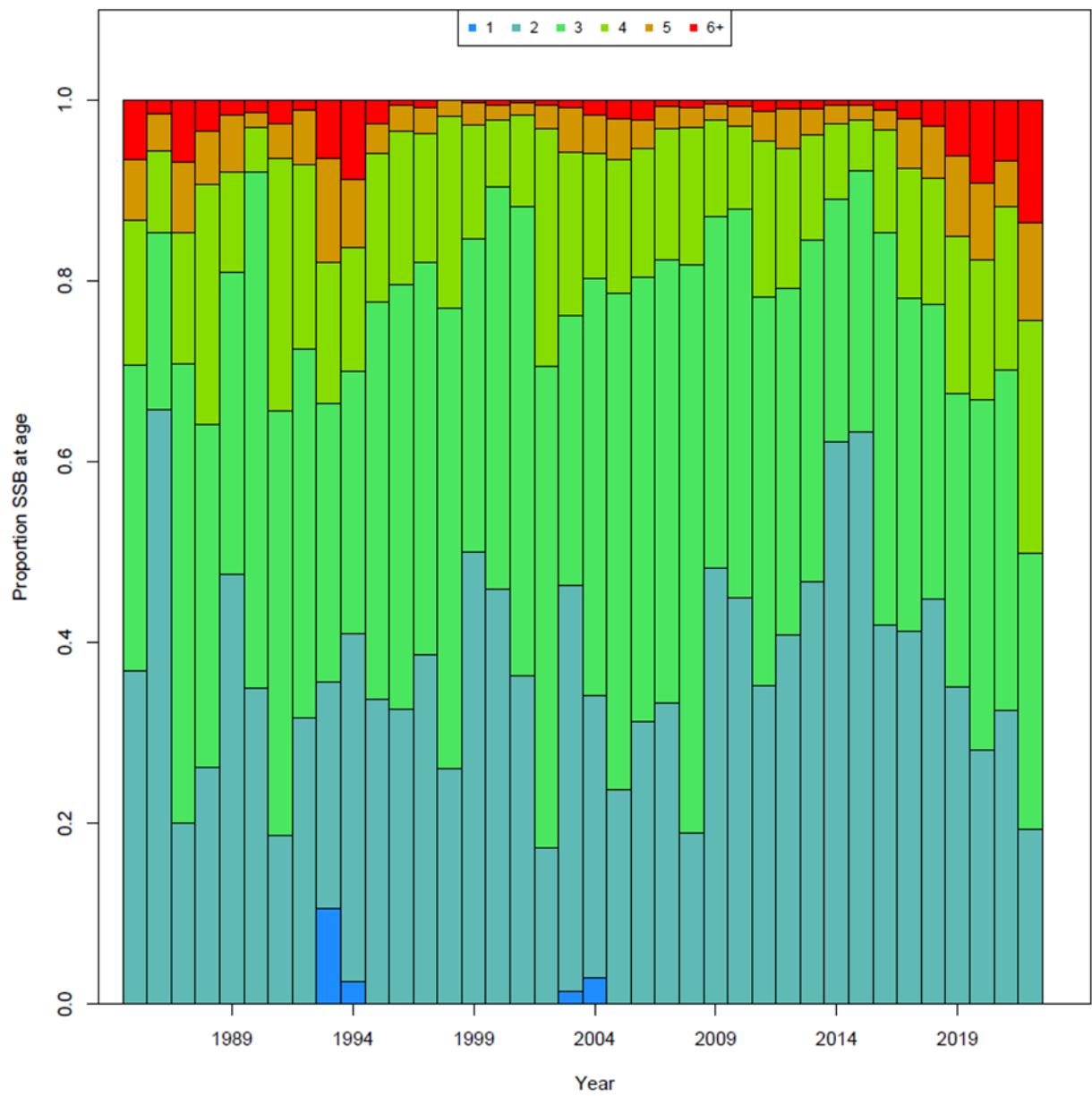


Figure 5.2.9: Proportion SSB-at-age from candidate model m452.

Cape Cod-Gulf of Maine WAA

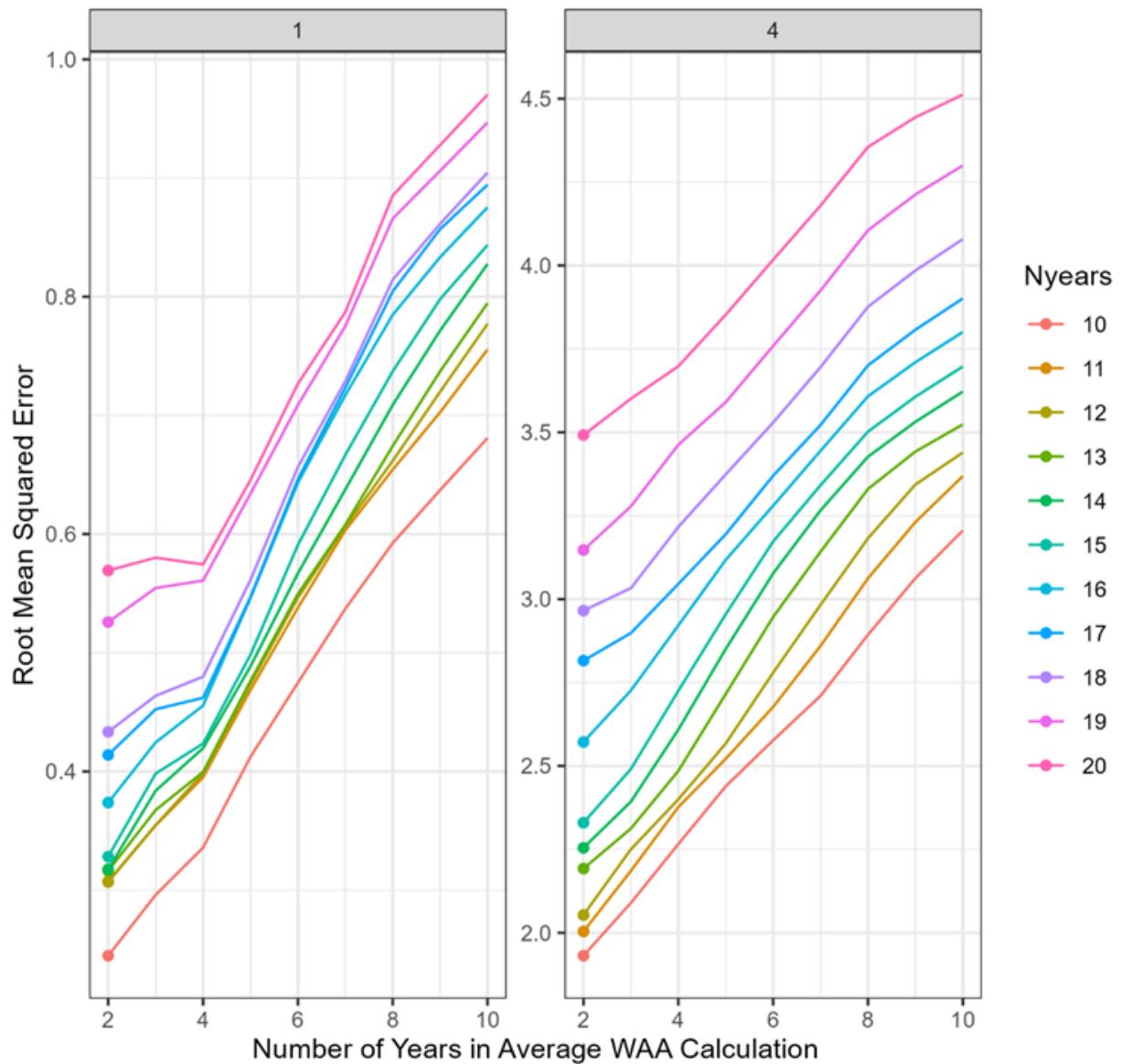


Figure 6.1.2: Moving window analysis for determining appropriate averages for WAA, Left is predicting one year ahead; right is predicting four years ahead. Colors represent the number of peels to calculate the total RMSE.

Cape Cod-Gulf of Maine Maturity

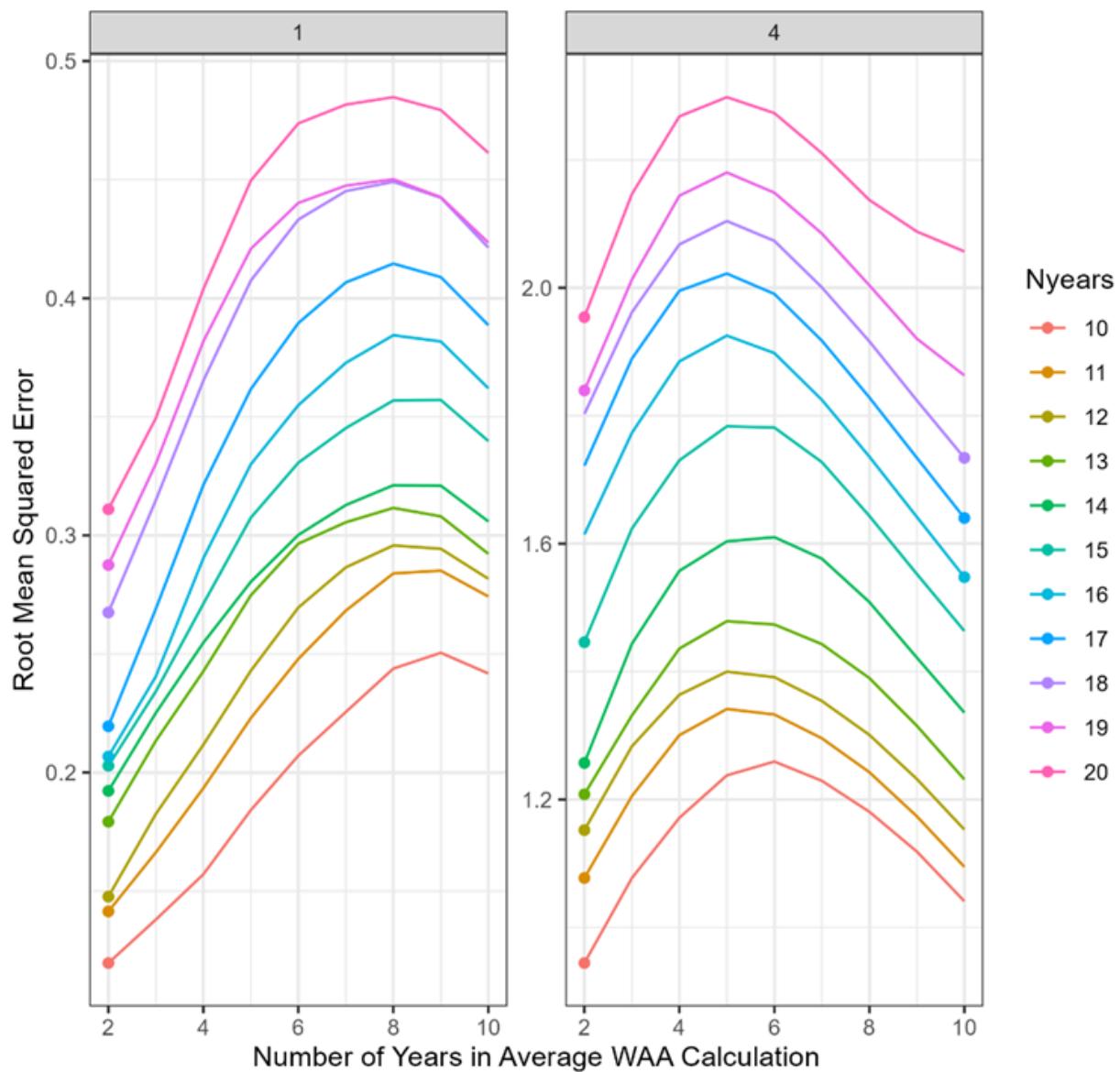


Figure 6.1.3: Moving window analysis for determining appropriate averages for maturity-at-age; Left is predicting one year ahead; right is predicting four years ahead. Colors represent the number of peels to calculate the total RMSE.

Annual Weight-at-Age for Total Catch

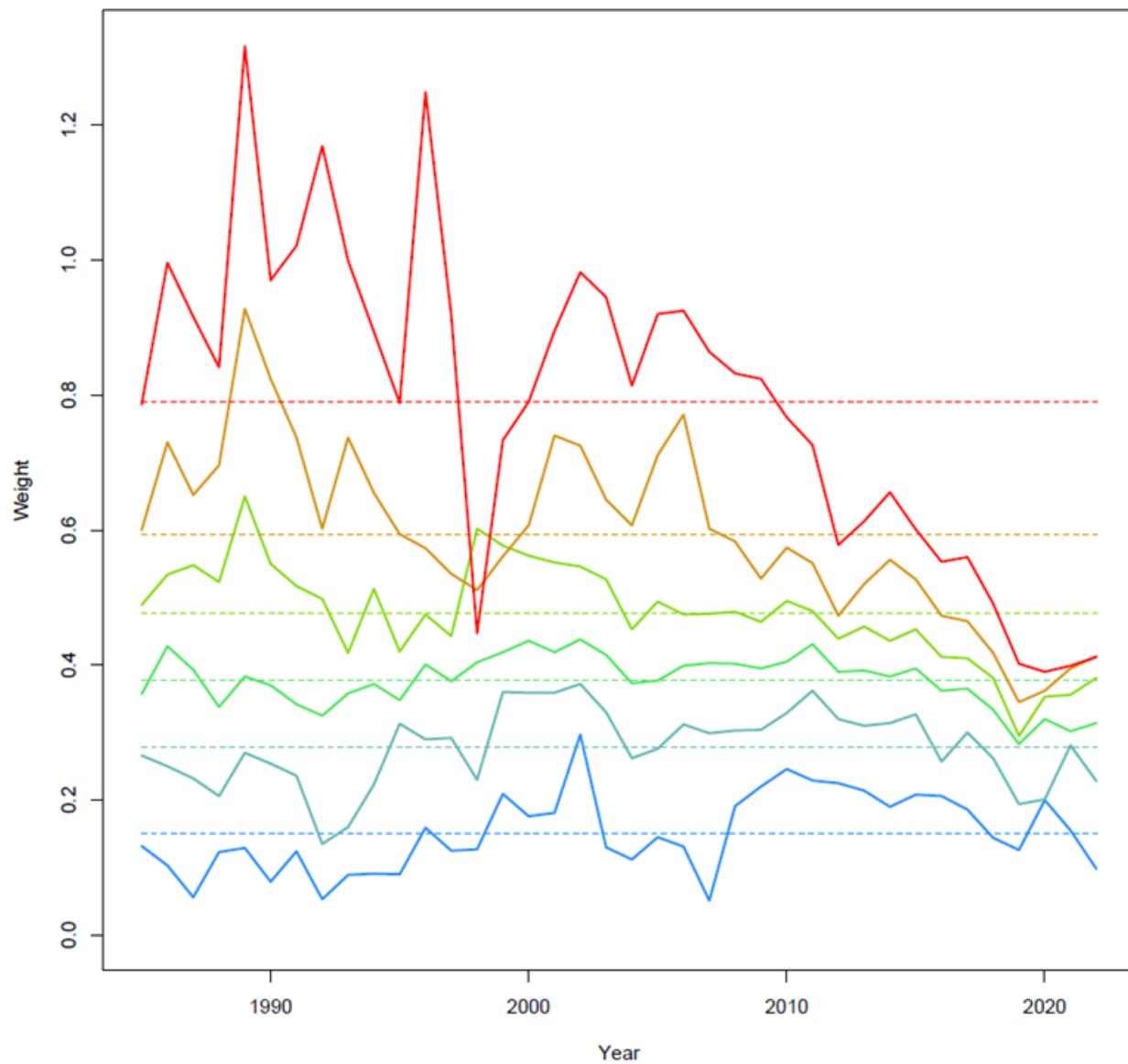


Figure 6.1.4: Aggregate fleet Weight-at-age over the entire time series (1985-2022).

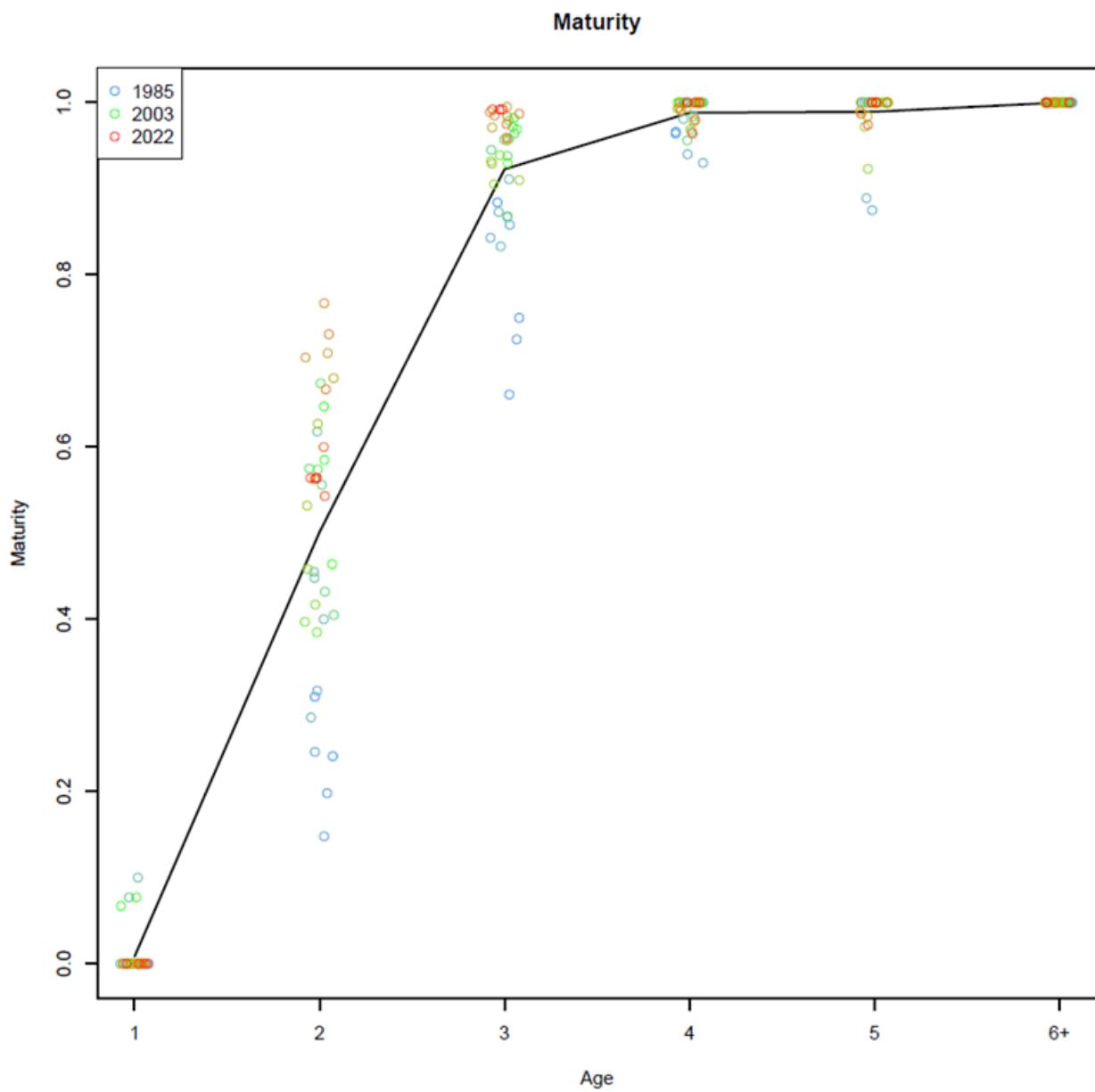


Figure 6.1.5: Proportion mature over the entire time series (1985-2022).

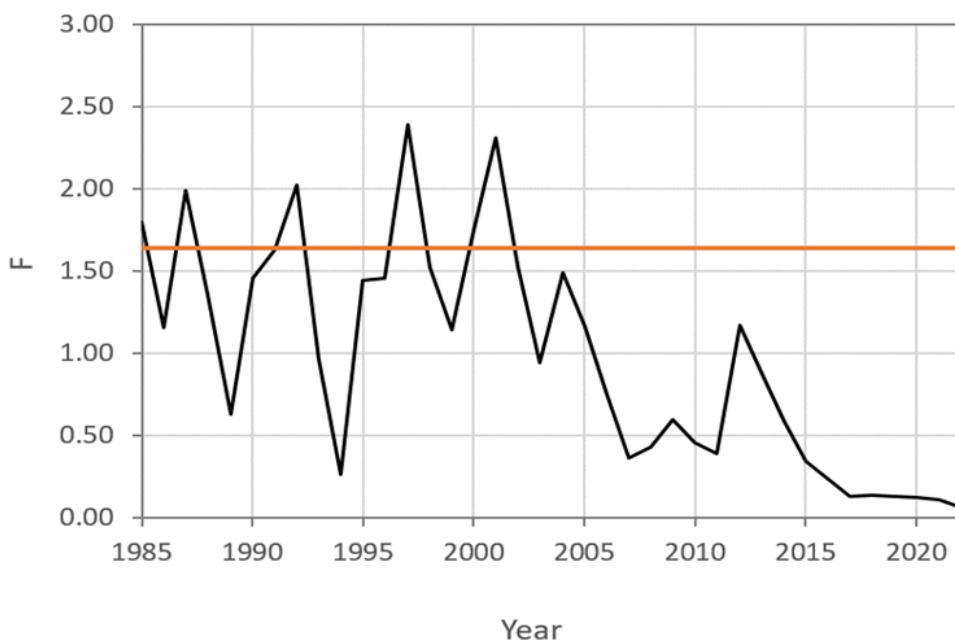
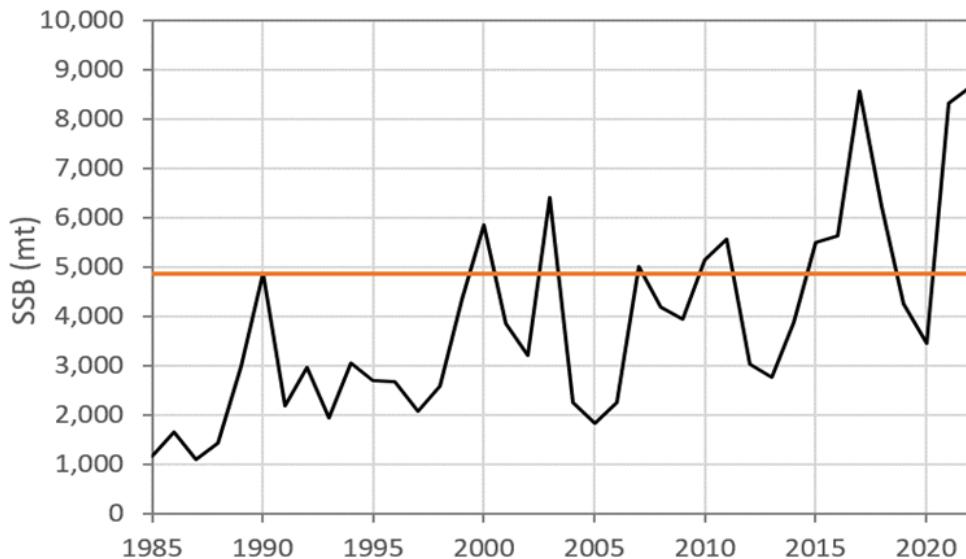


Figure 6.2.1: SSB 1985-2022 with a red line at $\text{SSB40\%} = 4870 \text{ mt}$ (top) and F 1985-2022 with a red line at $\text{F40\%} = 1.64$ (bottom). **Note:** Due to the large change in fleet selectivity, as well as changes in weights at age and maturity, the two red lines are not truly comparable with SSB and F at the start of the time series. Interpretation of the plots should be based on the recent 5-years.

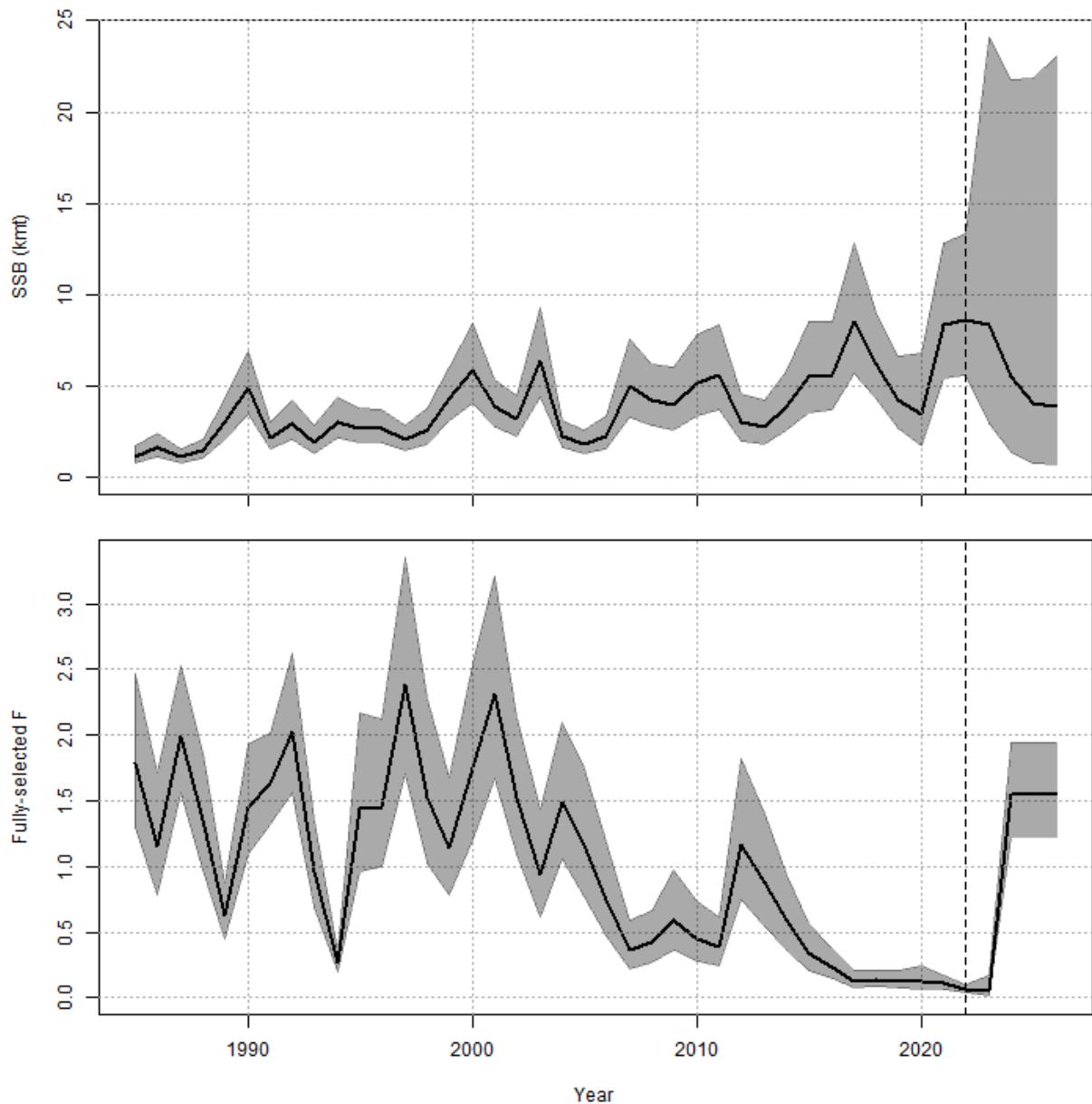


Figure 7.2.1: Short-term projections of SSB (top) and F (bottom).