Proposed Model Changes to the Alaska Sablefish Stock Assessment for 2025 Including Analysis of NOAA Longline Survey Design Changes

Daniel Goethel and Matt Cheng

September 2025 NPMFC Groundfish Joint Plan Team

Executive Summary

For 2025, a full assessment is scheduled for the Alaska-wide sablefish (*Anoplopoma fimbria*) stock. Major modeling changes to the stock assessment were undertaken, which aimed to address model stability, better align parametrizations with stock assessment good practices, and refine data inputs. A number of sensitivity runs were also explored to address outstanding concerns previously raised by the North Pacific Fishery Management Council's (NPFMC) groundfish Joint Plan Team (JPT) and Science and Statistical Committee (SSC). Finally, further analyses were developed using the sablefish stock assessment to better understand potential impacts (e.g., on catch advice) of survey design changes that will be implemented in 2025 for the NOAA sablefish longline survey.

The Alaska sablefish assessment is a single region, sex-specific model, which was last revised in 2023, primarily to address data inputs (e.g., switching to a standardized Catch-per-Unit Effort, CPUE, index and including non-commercial catch). The resulting model 23.5 was utilized for the 2024 update assessment. Model updates for 2025 began with a bridging exercise to move from the current AD Model Builder (ADMB) framework, which is no longer being actively maintained, to the widely recognized successor, Template Model Builder (TMB), using the more user-friendly R Bindings for TMB (R-TMB) package. The R-TMB version of the sablefish model was built using the Stochastic Population over Regional Components (SPoRC) model. SPoRC was originally developed for sablefish, but is completely generalized and capable of accommodating any number of regions, ages, sexes, and fishery or survey fleets (https://github.com/chengmatt/SPoRC).

Following successful model bridging from ADMB model 23.5 to R-TMB model *25.1_SPoRC_Cont* (Section 1), a stepwise model development process was undertaken wherein model changes were implemented incrementally within groupings of alike model explorations. Three primary model development groupings are detailed in the build from model *25.1_SPoRC_Cont* to the author recommended model *25.12_Drop_TS_Upd_M* and include: code fixes and good practice implementation (Section 2), disaggregation of age composition data (Section 3), and updates to data and model assumptions (Section 4). See Table 1 for a full list of models and Table 2 for a summary of results for the final model in each grouping. Further model explorations that were deemed mostly exploratory for 2025 are then provided as sensitivity runs in Section 5. For all model runs in this document, no new data was utilized compared to the 2024 update assessment. Choice of final models within each grouping as well as the author recommended model were made based on a combination of whether a model led to improved data fits or diagnostics (particularly retrospective bias), reductions in model tension among data sets, increased model stability, and/or improved biological realism. In many instances, results and diagnostics were similar across model runs, which required the use of expert judgment to identify which models were likely to be more robust to the primary uncertainties associated with sablefish population dynamics while providing consistent management advice.

Finally, to address potential impacts from ongoing changes to the NOAA sablefish longline survey footprint and experimental design (i.e., surveying only the Gulf of Alaska in odd years and only the Bering Sea and Aleutian Islands in even years), a retroactive survey analysis is provided in Section 6. This analysis reruns the assessment with refined survey indices and age compositions meant to emulate future data scenarios. Two appendices then provide a full mathematical description of the underlying SPoRC assessment framework (Appendix 1) and additional figures comparing the continuity model (*25.1_SPoRC_Cont*) with the 2025 author recommended model (*25.12_Drop_TS_Upd_M;* Appendix 2).

The R-TMB version of the sablefish model (*25.1_SPoRC_Cont*) was able to match the estimates and performance of the ADMB continuity model with negligible differences (< 0.001% for likelihood values, time-series estimates, selectivity, and key estimates including reference points). Thus, the R-TMB model (*25.1_SPoRC_Cont*) was used as the basis for all subsequent model and code updates. When the R-TMB code was updated to align with assessment good practices (*25.6_Upd_Likes*) it resulted in improved model stability, better fit to the catch data, and a slight increase in natural mortality and spawning stock biomass (SSB; Table 3; Figure 1). Disaggregating age compositions by sex and removing length compositions when age data was available (model *25.9_Age_Drop_Len*) slightly improved fits to age composition data, but at the cost of moderately degraded fits to the recent longline survey index. Increased natural mortality estimates led to an increase in SSB scale, while removal of the length composition data led to refined recent recruitment estimates, suggesting fewer extreme recent cohorts (Figure 1).

The final author recommended model (*25.12_Drop_TS_Upd_M*) revised the prior used on natural mortality to ensure a more biologically realistic and lower mean value (0.085 compared to 0.1), while also removing the trawl survey index as it was deemed to be non-representative of the Alaska-wide sablefish population. The resulting estimate of natural mortality was lower (~0.1) compared to the continuity model (~0.115) due to the reduced prior mean value, which resulted in a moderate decline in the scale of SSB and reductions in catch advice (Table 3; Figure 1). Removal of the trawl survey data led to less variable recent recruitment estimates and reduced recruitment retrospective bias. However, model *25.12_Drop_TS_Upd_M* appears to initially underestimate very large year classes (e.g., the 2016 year class) likely causing the model to provide precautionary management advice when extreme year classes first recruit to the fishery. Overall, given that the model and data updates as represented in model *25.12_Drop_TS_Upd_M* lead to a combination of better fits to primary data sources (i.e., age compositions), reduced retrospective bias, improved model stability, and increased biological realism of the sablefish assessment, this is the model being recommended by the authors as the basis of management advice for 2025.

A number of model sensitivities were explored to better understand model performance, diagnose potential model misspecification, and address requests or concerns posed by the groundfish Plan Team or SSC. Given the additional complexity associated with many of these model configurations, none are being recommended for use in 2025. However, each model provided useful functionality that might be considered in future sablefish assessments, but further analysis of stability and performance were deemed necessary before being recommended as the basis of operational management advice. The first sensitivity model run (*25.13_TV_ALK*) incorporated time-varying growth and weight-at-age (Cheng et al., 2024c). Slight improvement in fits to the

CPUE index and trawl fishery lengths were observed along with lower SSB estimates in recent years due reductions in weight-at-age. Next, a model estimating continuous time-varying selectivity for the fixed-gear fishery was developed (*25.14_TV_Selex*). Again, slight improvements in fits to the CPUE index and length compositions resulted, while no substantive improvements in fits to age compositions occurred. Estimates of selectivity were fairly consistent between the continuous and time block approaches. A model using a self-weighting, two-dimensional (i.e., incorporating age- and sex-specific correlations) logistic normal composition likelihood (*25.16_2dLN*) for age-composition data was then implemented. The model demonstrated improved fits to the longline survey index in recent years along with to the age-composition data. Lower estimates of natural and fishing mortality (the latter due to changes in fixed-gear selectivity estimates) led to modest increases in population SSB scaling. The final sensitivity run described the results of a spatially-explicit model (*25.15_Spatial*; Cheng et al., 2025c), which included five regions and directly integrated tagging data to estimate movement among regions. Overall, estimates of Alaska-wide SSB and recruitment were similar between model *25.12_Drop_TS_Upd_M* and *25.15_Spatial*. However, regional differences in population dynamics emerged, where SSB tended to be highest in the central and eastern Gulf of Alaska, while age-2 recruitment tended to originate in the Bering Sea and the Aleutian Islands. Collectively, these sensitivity runs provide a foundation for exploring alternative model parameterizations in future assessments, with the spatial model offering a promising approach to address the non-uniform sampling design of the longline survey.

Based on the longline survey retroactive analysis, it is expected that the new survey design will lead to reduced survey relative population numbers (RPNs) and less informative age compositions. Although catch advice will not likely be impacted in the short-term, increased recruitment variability and underestimation of large year classes is increasingly possible. Moreover, the new survey design will necessitate extrapolating (i.e., using constant hold-over values for non-surveyed regions) ~50% of sablefish abundance in a given year. Additionally, age compositions will only reflect year class strength in more localized regions. This analysis was only conducted during a period of rapid population increase, which inherently makes the results more conservative. Using holdover RPNs for unsurveyed regions during a period of population decline would lead to inflated RPNs, quotas being set too high, and potentially deleterious effects on the sablefish stock.

The following document summarizes model performance and results of select model runs, primarily in comparison to the continuity model (*25.1_SPoRC_Cont*). However, full model results and diagnostics are available at the Alaska sablefish assessment GitHub site, where individual folders are provided for every model run produced for the September NPFMC JPT containing SPoRC input and output R files along with PDFs of results (https://github.com/dgoethel-noaa/2025_Sablefish_SAFE/tree/main/Sept%20PT%20Model%20Runs).

Tables

Table 1. Model runs undertaken as part of the 2025 sablefish assessment model update process. Model development proceeded in a stepwise manner within each grouping, where the final model in each group (bold) included all updates considered for that group. Moreover, the final model from the previous group was typically used as the starting model for the next group of model runs. All model runs with prefix '23' utilize ADMB, while all model runs with prefix '25' utilize R-TMB.

Model Group	Model Name	Major Changes	Rationale
Bridge to RTMB (Section 1)	*23.5_Cont*	None	2024 Continuity Model
	23.5a_No_Dev_Vectors	Remove dev_vectors from ADMB	Remove functionality not natively available in RTMB.
	23.5b_No_Dev_No_Max	Remove dev_vectors and max calls from ADMB	Remove functionality not natively available in RTMB.
	25.1_SPoRC_Cont	Match 23.5b in RTMB (SPoRC framework)	R-TMB bridge to the *23.5_Cont* model.
Code Fixes and Good Practices (Section 2)	*25.2_Fix_Legacy_Code*	(25.1+) Fix legacy code issues	Improve model by removing legacy code issues.
	25.3_Fix_SigR	(25.1+) Fix sigma_R = 1.1	Implement good practices by fixing sigma_R.
	25.4_Upd_Selex	(25.1+) Add selectivity priors and reduce parameter linkages	Add loose selectivity priors to increase stability and reduce linked parameters.
	25.5_All_Code_Upd	All updates in model 25.1 – 25.4	Simultaneously implement all previous model updates.
	25.6_Upd_Likes	(25.5+) Set all likelihood lambdas = 1.0 and utilize appropriate implementation of probability functions (e.g., catch likelihood)	Implement good practices for MLE likelihoods and remove legacy code for fitting catch data.
Disaggregate Age Compositions (Section 3)	*25.7_Disagg_Age*	(25.1+) Disaggregate age composition data	Disaggregate age comps to help inform sex-specific dynamics.
	25.8_Age_Upd_Code	All updates 25.6 – 25.7	Implement all code updates and age comp disaggregation.
	25.9_Age_Drop_Len	(25.8+) Remove length composition data for years/data sources with age composition data (except for Japanese Longline Survey; LLS)	Remove length comps when age comps exist in a given year for a given data source (except for JPN LLS) to improve model performance.

Model Group	Model Name	Major Changes	Rationale
Update Data and Model Assumptions (Section 4)	*25.10_Drop_Trwl_Sur*	(25.9+) Remove the trawl survey	Trawl survey is a contradictory data source adding to model tension, so remove it to reduce variability in recruitment estimation.
	25.11_Upd_M_Pr	(25.9+) Change the M prior to ~N(ln(0.085), 0.1^2)	The M prior is relatively high for a long-lived species, so the prior mean value is reduced to be more consistent with tagging estimates of M and observed longevity.
	25.12_Drop_TS_Upd_M	All updates 25.10 – 25.11	Remove trawl survey and decrease M prior mean.
Sensitivity Runs (Section 5)	*25.13_TV_Growth_WAA*	(25.12+) Implement time-varying growth and weight	Address SSC/JPT requests by exploring time-varying growth and weight-at-age.
	25.14_TV_Selex	(25.12+) Implement continuous time-varying fixed gear fishery selectivity	Address SSC/JPT requests by exploring continuous time-varying selectivity.
	25.15_Spatial	A spatially explicit model for sablefish	Provides further inference on regional dynamics and depletion.
	25.16_2dLN	(25.12+) Implement logistic normal likelihood for age compositions	Evaluate how alternative data-weighting likelihoods impacts model inference.

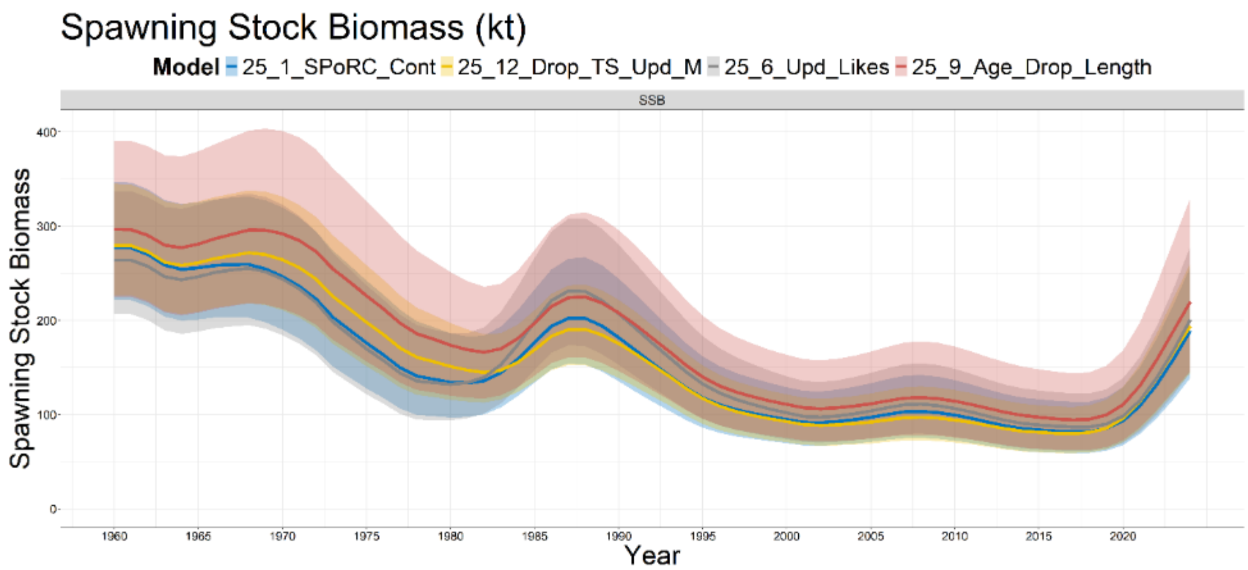
Table 2. Summary of results for the final model within each grouping, including the author recommended model *25.12_Drop_TS_Upd_M* (bold).

Model Group	Model Name	Major Changes	Conclusion	Author Recommendation
Bridge to RTMB (Section 1)	*25.1_SPoRC_Cont*	Move the ADMB assessment into RTMB with minor adjustments (no *dev_vectors* or *max* calls).	RTMB is a more flexible, user friendly framework, and the SPoRC package was able to adequately match the ADMB version of the sablefish assessment.	Use the SPoRC framework as the basis of the sablefish operational SAFE assessment.
Code Fixes and Good Practices (Section 2)	*25.6_Upd_Likes*	Address legacy code, fix Sigma_R, implement selectivity priors, reduce parameter linkages, and use appropriate likelihoods.	Code and assessment changes led to a more stable model parametrization, which better adheres to assessment good practices. Minor changes in model scaling occur, primarily driven by selectivity and M estimates.	All model updates represent defensible good practices, while providing a more stable parametrization of sex-specific selectivity.
Disaggregate Age Compositions (Section 3)	*25.9_Age_Drop_Len*	Disaggregate age compositions by sex and implement all previous model updates.	Disaggregating age compositions improves model stability and refines understanding of recent recruitment events, but increases tension with length composition data.	Disaggregating age data by sex is good practice and the sample sizes support this approach, so all composition data should be disaggregated.
		Remove length composition data for years/data sources with age composition data (except for Japanese LLS).	Removing length compositions further improves stability, and alleviates model tension among age and length comps (to a degree), but implicitly upweights the influence of the trawl survey data.	Length composition is a less reliable indicator of year class strength (given high uncertainty in the estimation of growth needed to convert lengths to ages), so removing lengths in years that ages are available improves model performance.
Update Data and Model Assumptions (Section 4)	***25.12_Drop_TS_Upd_M***	Drop the trawl survey and associated length compositions.	The trawl survey provides contradictory data to the LLS and only partially samples the sablefish population (GOA only, < 500m, primarily juveniles) adding to model tension, while removing it reduces variability in recruitment estimation.	The trawl survey was initially integrated to aid estimation of recent year classes, but appears to be providing contradictory information to the dedicated sablefish LLS. Removing the trawl survey improves recruitment estimation and consistency.
		Change the M prior to $\sim N(\ln(0.085), 0.1^2)$.	The current M prior mean value $\sim N(\ln(0.1), 0.1^2)$ is high for a long-lived species, where the new reduced prior mean better aligns with estimates from tagging data, observed longevity, and M estimates in other regions. However, M is an important model scalar, so the choice is influential. A lower M estimate increases F and reduces the scale of SSB.	Basic biology, observed maximum ages, and tagging data all support the use of the new, lower prior mean on M. Although it scales down SSB and ABC, it better aligns with known sablefish biology and should be maintained given that the model cannot adequately estimate M freely.

Table 3. Estimates of key assessment outputs from the final models in each model development grouping. See Table 2 for a summary of model changes and Table 3 for a summary of results.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_1_SPoRC_Cont	1	189.0804	0.04037044	49.68588	121.0229	0.08593	1.56235	0.46979
25_6_Upd_Likes	1	200.9191	0.03842778	52.95886	128.0622	0.08665	1.56892	0.44350
25_9_Age_Drop_Length	1	219.7952	0.04162539	50.00421	125.1969	0.09062	1.75560	0.45932
25_12_Drop_TS_Upd_M	1	194.0907	0.04842486	39.55329	126.6058	0.08308	1.53303	0.58289

Figures



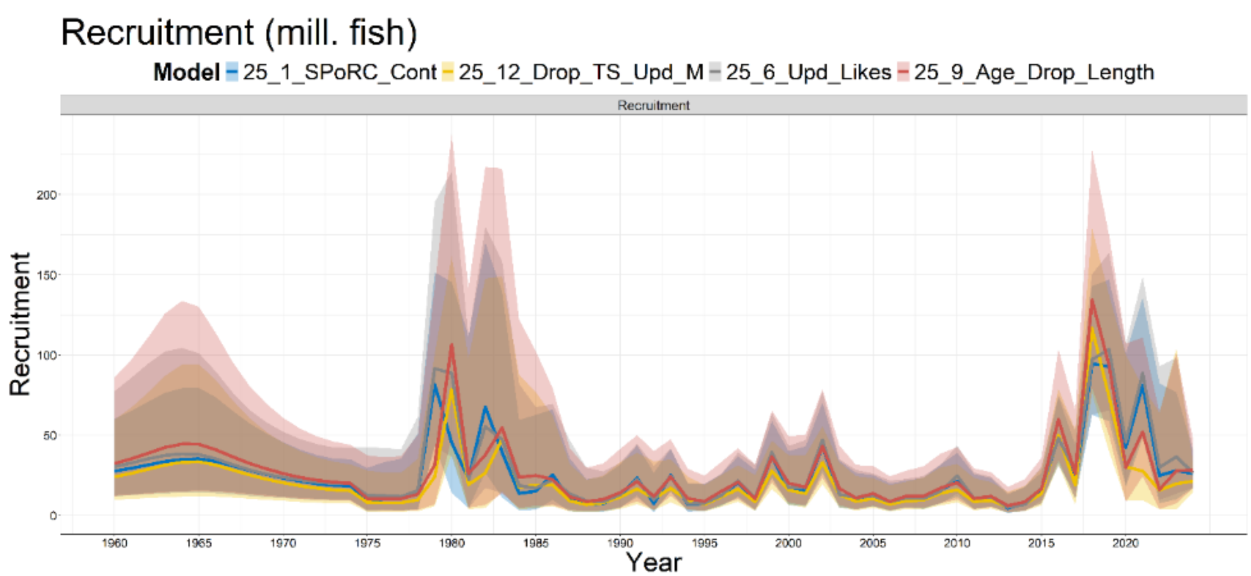


Figure 1. Comparison of spawning stock biomass (SSB, in kilotons; top panel) and recruitment (in millions of fish; bottom panel) across the four main model runs, where shaded regions represent the 95% confidence intervals.

1. Bridging Exercise from ADMB to RTMB

Introduction

The Alaskan sablefish assessment is a separable age- and sex-structured assessment model implemented in the Automatic Differentiation Model Builder [ADMB; Fournier *et al.* (2012)] platform to estimate parameters and provide management advice. However, ADMB is no longer being actively developed. Template Model Builder [TMB; Kristensen *et al.* (2016)] is widely considered the successor to ADMB, offering improved capabilities for estimating random effects through the Laplace approximation. Additionally, TMB models can be written in the R language (R-TMB), eliminating the need to develop models using C++ and providing an accessible framework for developing contemporary assessment models. To date, several assessments at the Alaska Fisheries Science Center have successfully transitioned to the TMB or R-TMB platforms (Monnahan *et al.*, 2023; Williams *et al.*, 2024).

Several differences between the ADMB and TMB platforms exist, which could lead to variations in parameter estimation. Mainly, unlike ADMB, TMB lacks:

1. Native parameter phasing capabilities, generally requiring all parameters to be estimated simultaneously;
2. Parameter bounds cannot be *readily* defined;
3. The ability to define a *dev_vector* parameter class, which penalizes a vector to sum to zero;
4. The inclusion of *max* functions, which are commonly used to scale selectivity values in ADMB (though these are likely inappropriate given the non-differentiable nature of the max function).

Note that many of these ADMB capabilities, if used in an assessment, may not be considered 'good' practice.

Here, we present a bridging exercise of the 2024 accepted ADMB model [model 23.5; Goethel and Cheng (2024)] to R-TMB. We utilize a generalized stock assessment package, initially developed to account for the inherent complexities associated with sablefish (e.g., sex, fleet, and spatial structure). The Stochastic Population over Regional Components model is completely generalized and capable of accommodating any number of spatial dimensions, ages, sexes, fishery fleets, and survey fleets (https://github.com/chengmatt/SPoRC). Full documentation of the model dynamics and equations are provided in Appendix 1 and available from https://chengmatt.github.io/SPoRC/articles/c_model_equations.html.

Methods

The sablefish model bridging exercise took a step-wise approach and evaluated each individual impact of explicitly identified differences between the ADMB and R-TMB platforms. We first developed three variants of the ADMB model to sequentially remove ADMB-specific functionality not included in R-TMB:

1. Model 23.5: 2024 SAFE model;
2. Model 23.5a: remove the use of the *dev_vector* type;
3. Model 23.5b: remove the use of the *dev_vector* type and max function calls.

Model 23.5b, which addresses all of the major difference between estimation platforms, is then compared with a sablefish implementation of SPoRC (R-TMB), model *25.1_SPoRC_Cont*. As noted, SPoRC was initially developed with sablefish in mind, so it was readily able to address existing sex and fleet structure along with parameter sharing, priors, and data fitting assumptions (e.g., sex-aggregated age compositions and sex-disaggregated length compositions). While porting the sablefish ADMB model to SPoRC, a handful of coding bugs and inconsistent modeling decisions were encountered. To ensure an accurate representation of the operational assessment (as adapted for TMB functionality, model 23.5b), the SPoRC source code was updated to be able to retain all aspects of the sablefish model. However, subsequent model updates (see Section 2) aim to address these to provide a more robust sablefish assessment.

Initial comparisons involved first initializing the R-TMB model *25.1_SPoRC_Cont* using the direct parameter estimates from the ADMB model without performing model optimization. As a kind of self-test, using the same parameter values likely demonstrates whether the two models had identical structural population dynamics. Initial comparisons indicated that differences in model outputs were < 1e-5% (results not shown for brevity), indicating that the underlying population dynamics and likelihood components between the models were equivalent.

In the following, we present comparisons between models 23.5b and *25.1_SPoRC_Cont* post-optimization using the same data inputs. The post-optimization comparisons help assess convergence towards similar parameter solutions across platforms. Comparisons focus on differences in likelihood components, time-series quantities of age-2 recruitment (millions) and spawning stock biomass (SSB; kt) along with their estimated standard errors (SE), total fishing mortality (fully selected instantaneous mortality), total female and male abundance (millions of fish), fleet-specific selectivity curves, and key parameter estimates [e.g., reference points ($B_{40\%}$ and $F_{40\%}$), fishery and survey catchability, natural mortality, and mean recruitment]. Model outputs for each ADMB model are available here and those for model *25.1_SPoRC_Cont* here.

Results

In general, post-optimization results from the three ADMB models demonstrate negligible differences in time-series estimates (Figure 2). Similarly, comparisons of post-optimization results for model 23.5b with the R-TMB model *25.1_SPoRC_Cont* show negligible differences. In particular, relative differences are all within < 0.001% for likelihood values, time-series estimates, selectivity, and key estimates (including reference points) (Table 4, Figures 3 – 6). Thus, we conclude that model 23.5b and model *25.1_SPoRC_Cont* are essentially identical, providing effectively the same stock status and management advice (Figures 3 and 6).

Although there are some minor differences between estimation platforms, R-TMB's lack of a *dev_vector* class and inability to use *max* functions should not preclude a transition to the platform, given its improved capacity to estimate random effects and ease of implementation. Moreover, recruitment deviates are already constrained to a mean of zero through penalties (in both the ADMB and R-TMB implementations of the sablefish assessment), and enforcing a sum-to-zero constraint (as is done in the ADMB implementation) has the potential to introduce irregular behavior in optimization processes. Likewise, *max* functions are theoretically non-differentiable, making their use problematic in automatic differentiation frameworks. Given the close alignment between model 23.5b with the R-TMB model *25.1_SPoRC_Cont*, we propose that the R-TMB model *25.1_SPoRC_Cont* should be the starting basis for the 2025 sablefish SAFE model development process. Moreover, the SPoRC framework provides greater flexibility for exploring alternative model structures (e.g., spatial structure, fishery fleets, and survey fleets) and improved capacity for estimating random effects.

Tables

Table 4. Likelihood values for model 23.5b and the R-TMB model *25.1_SPoRC_Cont* (post-optimization). Difference is calculated as ADMB – R-TMB, while relative difference is calculated as (ADMB – R-TMB) / ADMB.

Likelihood Component	ADMB (model 23.5)	R-TMB (model 25.1)	Diff	Relative Diff (%)
jnLL	768.83	768.83	0.00	0.00
Fixed Gear Fishery Age	46.63	46.63	0.00	0.00
Fixed Gear Fishery Length (F)	85.55	85.55	0.00	0.00
Fixed Gear Fishery Length (M)	91.07	91.07	0.00	0.00
Trawl Gear Fishery Length (F)	25.46	25.46	0.00	0.00
Trawl Gear Fishery Length (M)	17.90	17.90	0.00	0.00
Domestic Survey LL Age	173.99	173.99	0.00	0.00
Domestic Survey LL Length (F)	48.07	48.07	0.00	0.00
Domestic Survey LL Length (M)	34.74	34.74	0.00	0.00
Domestic Trawl Survey Length (F)	23.12	23.12	0.00	0.00
Domestic Trawl Survey Length (M)	17.64	17.64	0.00	0.00
Japanese LL Survey Length (F)	27.92	27.92	0.00	0.00
Japanese LL Survey Length (M)	17.54	17.54	0.00	0.00
Catch	3.67	3.67	0.00	0.00
Domestic LL Survey Index	33.76	33.76	0.00	0.00
Japanese LL Survey Index	10.16	10.16	0.00	0.00
Trawl Survey Index	20.56	20.56	0.00	0.00
Domestic LL Fishery Index	13.49	13.49	0.00	0.00
Japanese LL Fishery Index	24.23	24.23	0.00	0.00
FMort Penalty	6.09	6.09	0.00	0.00
M Prior	0.79	0.79	0.00	0.00
Rec Penalty	23.27	23.27	0.00	0.00

Figures

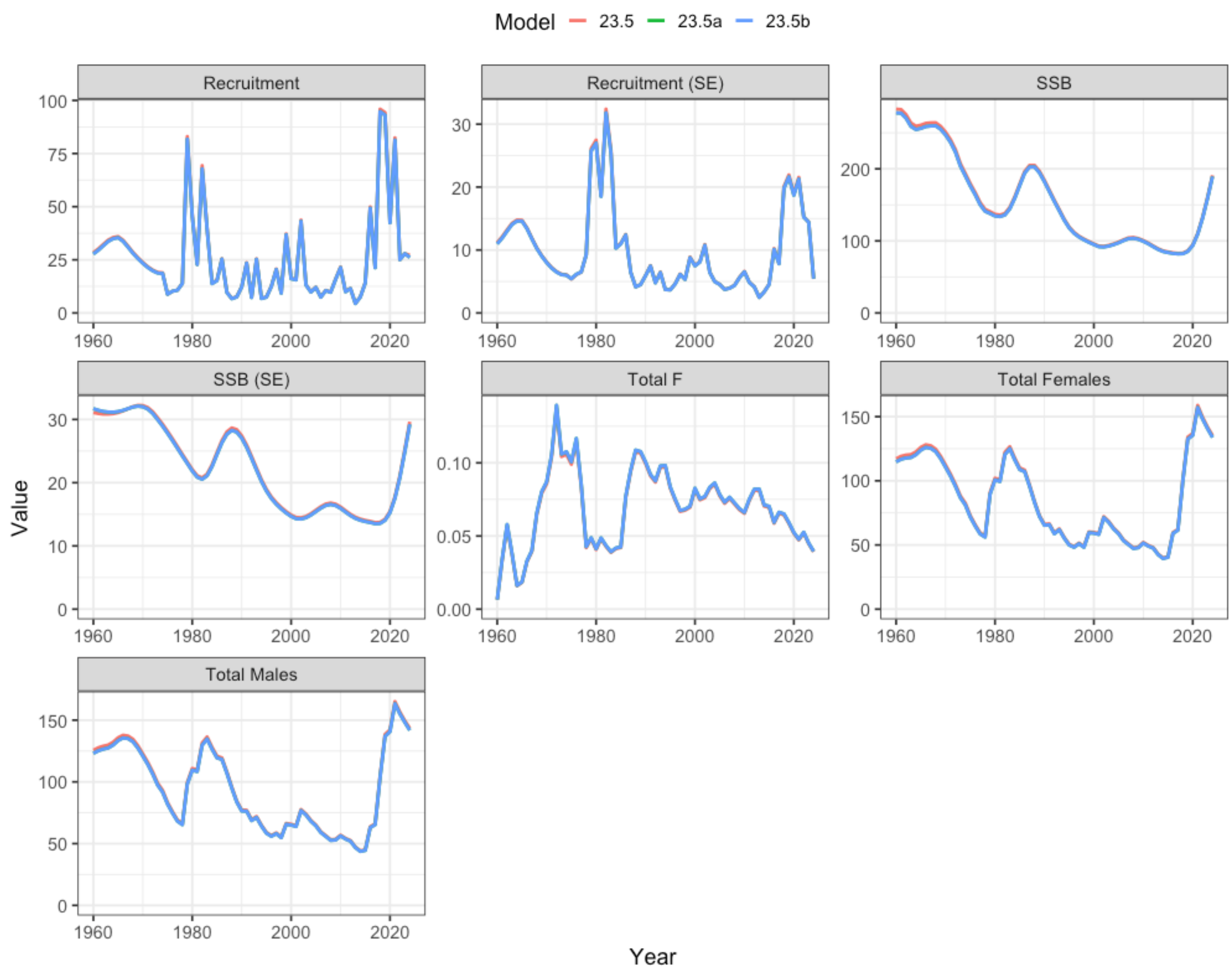


Figure 2. Comparisons of post-optimization time-series estimates among the ADMB model variants. Model 23.5 (red) is the 2024 operational assessment, model 23.5a is the operational assessment without the *dev_vector* class, and model 23.5b is the operational assessment without the *dev_vector* class and max function calls.

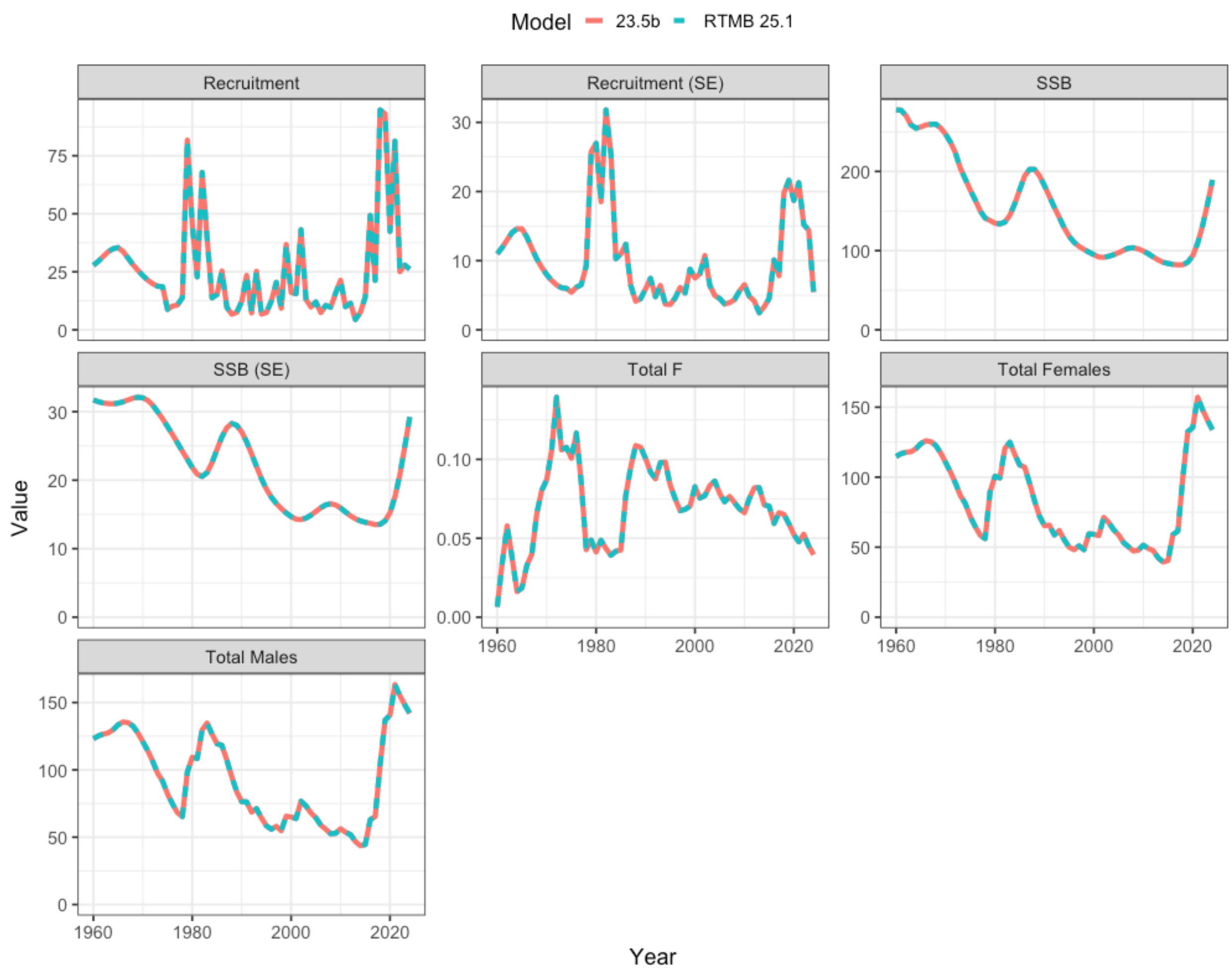


Figure 3. Comparison of time-series quantities (post-optimization) between model 23.5b and the RTMB model *25.1_SPoRC_Cont*. Dashed blue lines indicate the R-TMB model *25.1_SPoRC_Cont*, while the red line denotes model 23.5b. Note that time-series estimates from the two models are overlaid essentially on top of each other.

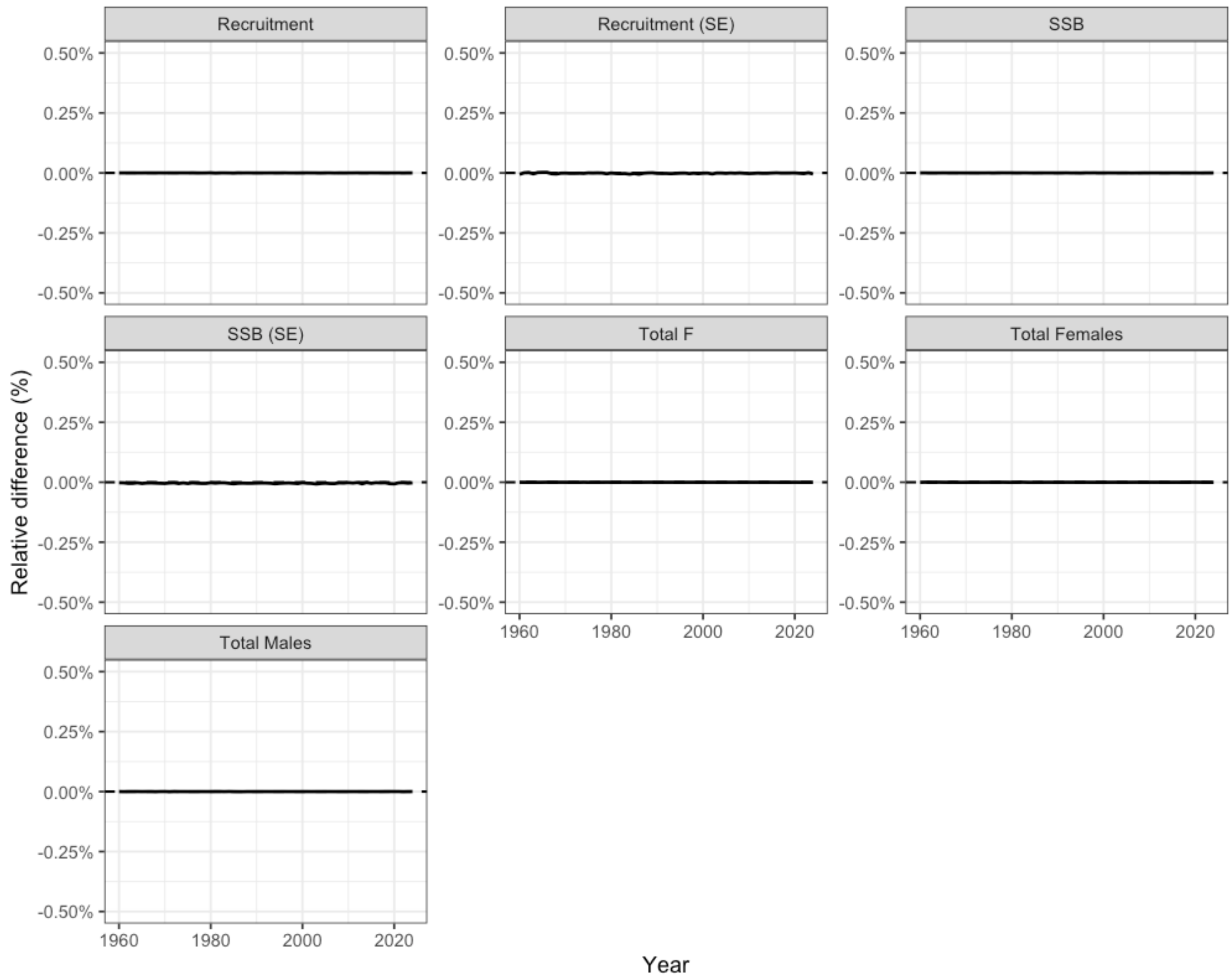


Figure 4. Post-optimization relative percent difference ((ADMB - R-TMB) / ADMB) in time-series quantities between model 23.5b and the R-TMB model *25.1_SPoRC_Cont*. The black dashed horizontal line indicates 0% difference between models.

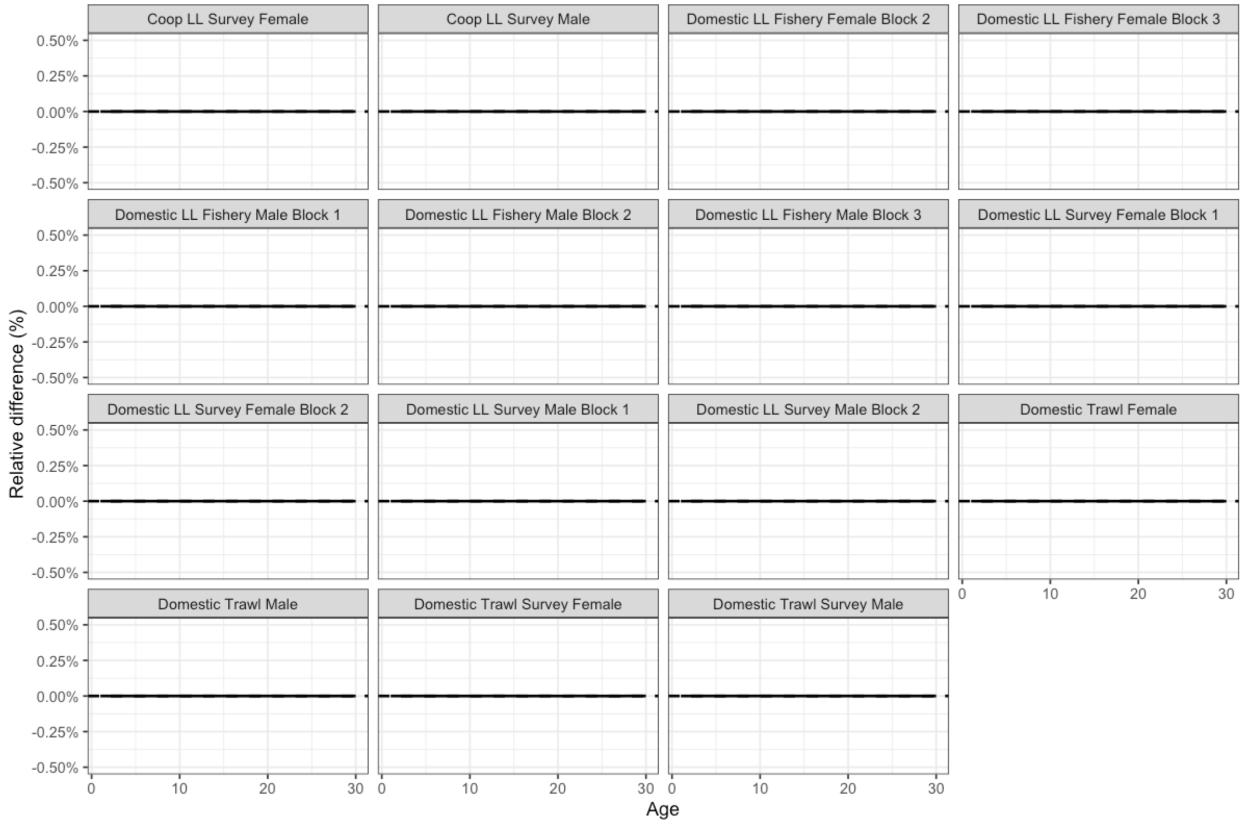


Figure 5. Post-optimization relative percent difference ((ADMB - R-TMB) / ADMB) in selectivity values between model 23.5b and the R-TMB model *25.1_SPoRC_Cont*. The black dashed horizontal line indicates 0% difference between models.

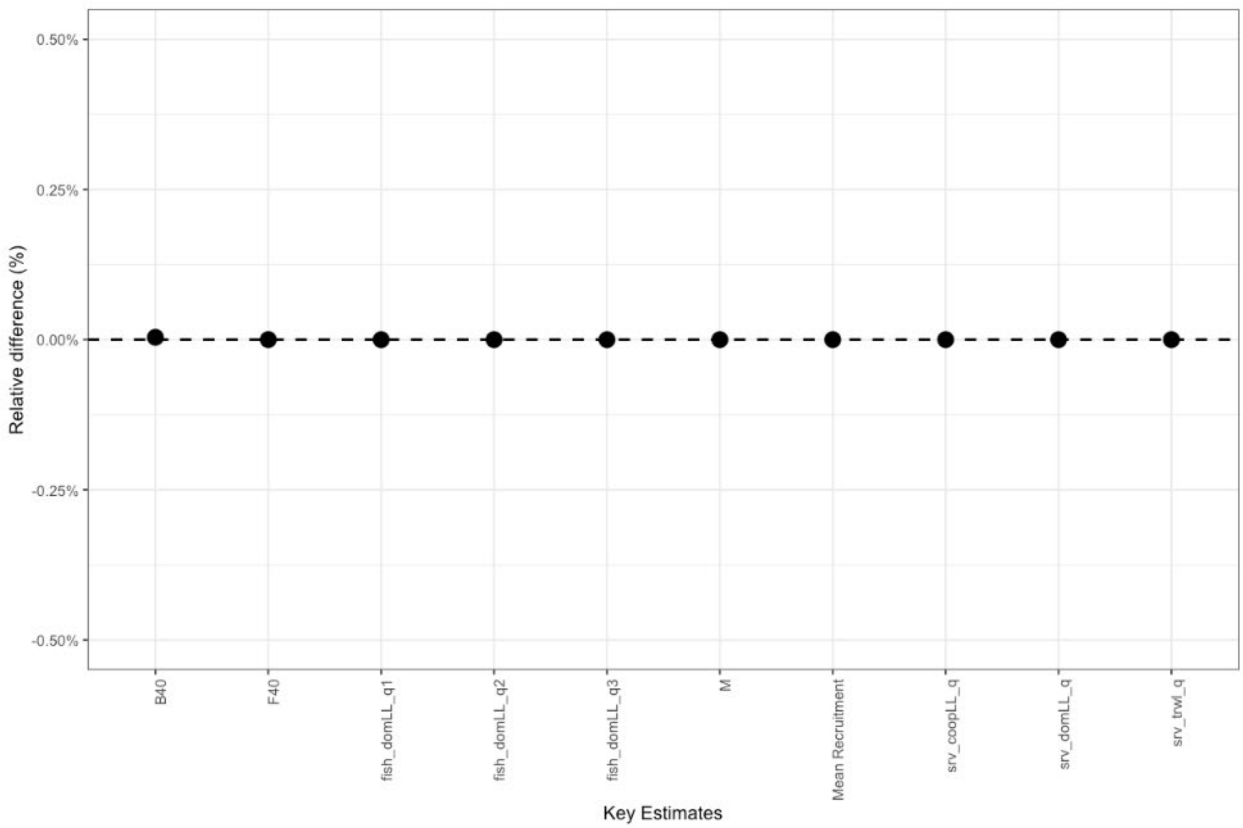


Figure 6. Post-optimization relative percent difference ((ADMB – R-TMB) / ADMB) in key parameter estimates between model 23.5b and the R-TMB model *25.1_SPoRC_Cont*. The black horizontal line indicates 0% difference between models.

2. Code Fixes and Good Practices

Introduction

Over the last five years, the sablefish assessment has undergone three periods of extensive code review by the lead assessment author upon taking over the assessment (D. Goethel, 2019), as part of the development of a TMB-based spatial model (post-doc C. Marsh, 2023), and while building SPoRC (post-doc M. Cheng; 2024). During these reviews, a number of code bugs, legacy code issues, and sub-optimal code parametrizations were identified. Moreover, recent SSC and JPT recommendations have noted specific issues that warranted modification in the assessment (i.e., fixing the recruitment variance term instead of estimating it). Given the change in modeling platform to R-TMB (Section 1) and the relatively straightforward nature of changing to more appropriate modeling choices in the new generalized platform, it was deemed an opportune time to rectify outstanding sablefish assessment issues.

Methods

Model updates in this group aimed to fix code bugs and inconsistencies (*25.2_Fix_Legacy_Code*), fix the recruitment variance term (as requested by both the groundfish PT and SSC) instead of estimating it in an MLE context as is currently done (*25.3_Fix_SigR*), improve model stability through an improved selectivity module (*25.4_Upd_Selex*), and adopt good practices for handling probability functions and data weighting (*25.6_Upd_Likes*). The first three models (25.2 – 25.4) were implemented incrementally (only the noted changes were implemented and not combined with other changes) from the continuity model (*25.1_SPoRC_Cont*), while model *25.5_All_Code_Upd* implemented all of the model changes for these three models simultaneously. The final model in this grouping (*25.6_Upd_Likes*) then built upon model *25.5_All_Code_Upd* and additionally implemented appropriate data weighting and likelihoods. Only model *25.6_Upd_Likes* underwent Francis reweighting, given the change in the likelihood structure, while all previous models maintained the same relative data weights as the continuity model (*25.1_SPoRC_Cont*). As noted, the description of results focuses on the final model in the group (*25.6_Upd_Likes*) primarily in comparison to the continuity model (*25.1_SPoRC_Cont*).

Primary model updates included:

- *25.2_Fix_Legacy_Code*:
    - Use a single, sex-invariant natural mortality estimate (previously a fixed male natural mortality offset was accidentally set to a nonzero value, fixed at -0.008).
    - Rectify coding bugs in the calculation of mortality for CPUE and composition data.
    - Implement a consistent approach to applying size-age transition and aging error across data sets.
    - Remove the additive constant from the catch likelihood (it is no longer needed because years with no catch are no longer included in the likelihood).
- *25.3_Fix_SigR*:
    - Fix the recruitment standard deviation at the previously estimated value of 1.1.
- *25.4_Upd_Selex*:

- Implement loose/uninformative priors on selectivity parameters to aid model stability, including (note that mean values are somewhat arbitrary given the high variance terms used, but more informative priors may be explored in the future):
    - Logistic selectivity functions: both parameters $\sim N(\ln(1),5^2)$.
    - Gamma selectivity functions (trawl fishery): δ $\sim N(\ln(1),2^2)$, b_{max} $\sim N(\ln(2),1^2)$.
- Improve sex-specific parameter linkages (previously these were haphazardly implemented):
    - δ is linked across sexes (within a given fleet) for the Japanese LLS and pre-IFQ (pre-1995) fixed gear fishery.
    - b_{max} is linked across sexes for the trawl fishery.

- *25.5_All_Code_Upd*:
    - Integrate all previously noted code updates in models *25.2 – 25.4*.
- *25.6_Upd_Likes*:
    - Integrate all previously noted code updates in models *25.2 – 25.4*.
    - Use full (and correct) likelihood probability functions (see Appendix 1).
    - Remove arbitrary weights (lambdas) for indices (= 0.448) and catch (= 50) and reset all lambdas to 1.0 (prior to Francis reweighting).
    - Assume a catch standard deviation = 0.05.
    - Double index CVs to address unrealistically low uncertainty:
        - Previous average CVs were 0.1 for the fishery CPUE, 0.05 for the Japanese LLS, 0.05 for the NOAA LLS, and 0.135 for the NOAA trawl survey.
    - Fix σ_R = 0.9 (this lower value was used to account for the previous SRR penalty weight of 1.5 that was removed).

Results

None of the model updates had a significant impact on fits to the data (Table 5; Figure 7), except for a slight improvement in fits to the length compositions with the update to the selectivity parametrizations (model *25.4_Upd_Selex*). Moreover, utilizing more appropriate likelihood functions and data weighting led to improved fit to the fixed gear catch data in model *25.6_Upd_Likes* (Figure 8). Because the old ADMB model used a least squares approach for fitting the catch data with a high data weight (λ = 50), it is likely that the approach was not adequately weighting the catch data (i.e., the weighting was roughly equivalent to using a standard deviation of 0.1 which is a low 'weight' for catch data).

Notable changes in model dynamics included:

- *25.2_Fix_Legacy_Code* (full results here):
    - A relative increase in total natural mortality (due to the removal of the male offset that previously reduced male natural mortality) led to an upward scaling of SSB to allow the same level of catch despite higher total mortality.
- *25.3_Fix_SigR* (full results here):
    - Essentially no change from the continuity model.
- *25.4_Upd_Selex* (full results here):

    - Increased model stability resulted from loose priors on selectivity parameters, while minor changes to the shape of the trawl fishery selectivity occurred resulting in reduced estimates of M (given the inherent correlation between the domed selectivity, F, and M).
    - Slight reductions in M resulted in minor downward scaling in SSB to allow for the same catch despite reduced total mortality.
- *25.5_All_Code_Upd* (full results here):
    - Implementing all changes in tandem led to minor differences from the continuity model, mostly related to overall increases in natural mortality.
- *25.6_Upd_Likes* (full results here)
    - Improved fit to the catch data.
    - Faster rebuilding of SSB and increased scale of SSB in recent years.
    - Slight increases in catch advice due to higher estimate of M.

The final model in the grouping (*25.6_Upd_Likes*) led to moderate differences in stock status and catch advice compared to the continuity model (Table 6). Mainly, to better fit the catch data given changes noted in previous models (slight changes in trawl fishery selectivity and moderate increases in natural mortality), the rate and magnitude of SSB rebuilding increases during the 1980s compared to other models and remains relatively higher at the end of the time series (Figure 9). Concomitantly, the average recruitment was slightly higher (Figure 10), which led to a moderate increase in the biomass reference point ($B_{40\%}$). In total, despite essentially equivalent stock status as the continuity model, catch advice increases slightly due to higher natural mortality (more fish can be removed because longevity is assumed to be lower). Generally, population trajectories from model *25.6_Upd_Likes* fall within the 95% confidence intervals of model *25.1_SPoRC_Cont*, while fits to the data are not significantly altered. However, given improved model stability and a more defensible model parametrization that better aligns with good practices, model *25.6_Upd_Likes* is recommended. In the future, the authors will look into treating recruitment deviations as random effects and directly estimate the recruitment variance term to better utilize the capabilities of TMB (this was not done currently for various reasons, but once the RTMB model is approved for operational assessment it will be a primary model update in the future).

Tables

Table 5. Negative log-likelihood components for each model run in the 'Code Fixes and Good Practices' grouping, including data fit components and parameter prior or penalty components. Note that the fundamental likelihood structure changed for model *25.6_Upd_Likes* making it not directly comparable to previous models.

Model	jnLL	M Prior	Recruitment Penalty	Catch nLL	Fishing Mortality Penalty	Survey Index nLL	Fishery Index nLL	Initial Age Penalty	Survey Age nLL	Fishery Age nLL	Survey Length nLL	Fishery Length nLL	Selectivity Penalty
25_1_SPoRC_Cont	768.8210	0.7838709	23.11463	3.688151	6.088752	64.49362	37.72432	0.1631310	197.0703	46.62799	169.0396	220.0267	NA
25_2_Fix_Legacy_Code	771.1194	0.7662659	22.94943	3.384739	6.091542	64.77658	37.44798	0.1714670	196.7168	47.31267	170.3107	221.1912	NA
25_3_Fix_SigR	768.8556	0.7502505	22.94854	3.718270	6.090218	64.50946	37.74435	0.1572363	197.0988	46.62127	169.1287	220.0885	NA
25_4_Upd_Selex	766.4771	0.5465163	23.13318	4.193411	6.134515	66.13810	36.99018	0.2104431	196.8007	46.87685	161.7206	221.2670	2.465497
25_5_All_Code_Upd	768.9416	0.5279751	22.84171	3.908407	6.136092	66.35713	36.67317	0.2109372	196.4084	47.58893	162.9960	222.8158	2.476966
25_6_Upd_Likes	835.7924	-0.2354034	76.09974	-83.779841	146.315255	-29.36051	-10.47291	0.3208338	189.2229	46.83455	192.9026	235.4203	72.524762

Table 6. Estimates of key assessment outputs for each model run in the 'Code Fixes and Good Practices' grouping.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_1_SPoRC_Cont	1	189.0804	0.04037044	49.68588	121.0229	0.08593	1.56235	0.46979
25_2_Fix_Legacy_Code	1	203.2045	0.03882372	51.65097	128.6548	0.08583	1.57945	0.45235
25_3_Fix_SigR	1	188.3272	0.04053991	49.38646	120.9461	0.08577	1.55712	0.47264
25_4_Upd_Selex	1	173.3655	0.04385504	45.00124	118.3230	0.08437	1.46519	0.51982
25_5_All_Code_Upd	1	184.9366	0.04237330	46.46873	125.1706	0.08415	1.47748	0.50356
25_6_Upd_Likes	1	200.9191	0.03842778	52.95886	128.0622	0.08665	1.56892	0.44350

Figures

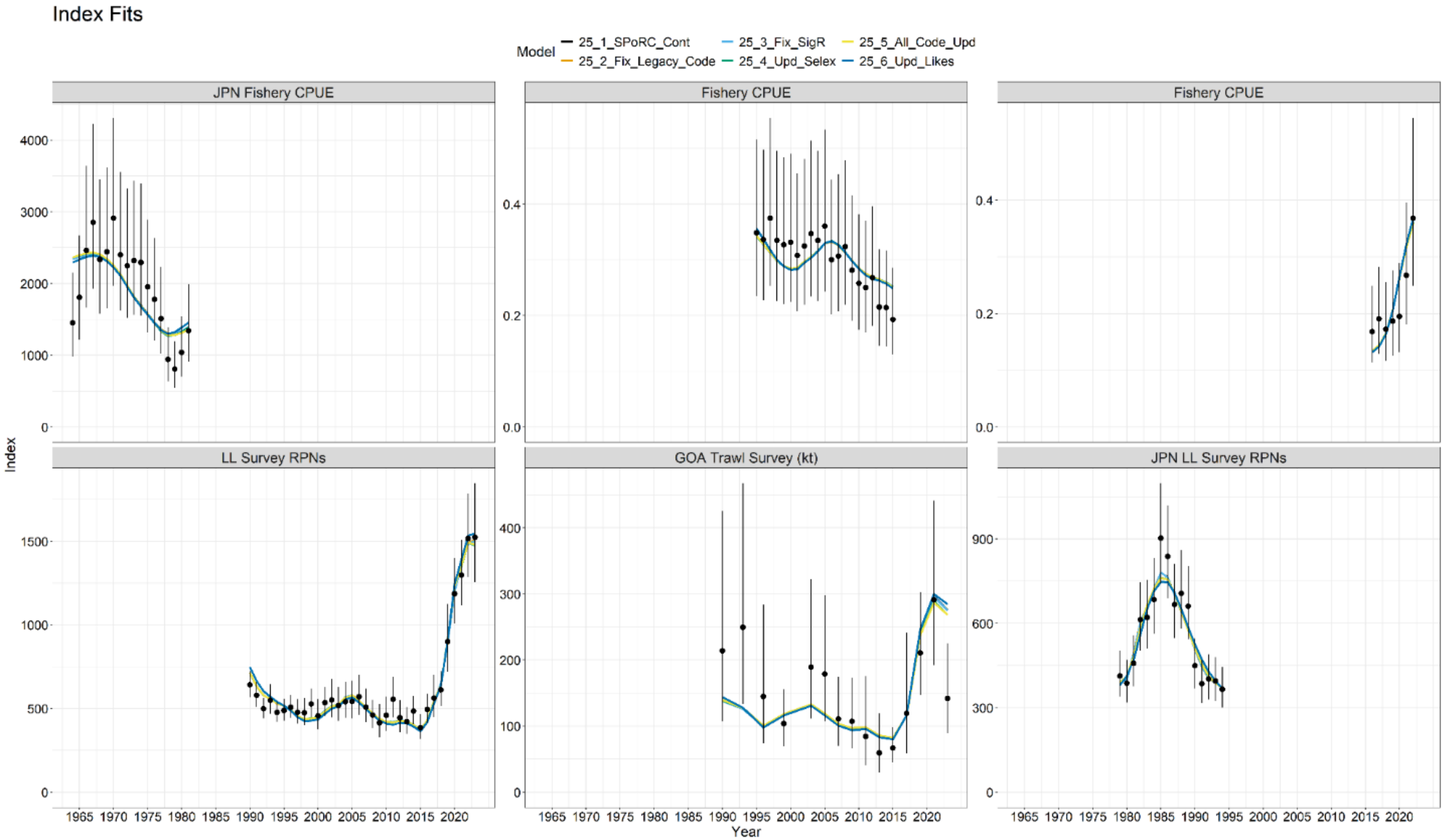


Figure 7. Fits (lines) to the various observed indices of abundance (dots; vertical lines represent 95% confidence intervals) used in the sablefish assessment for each model run in the 'Code Fixes and Good Practices' grouping.

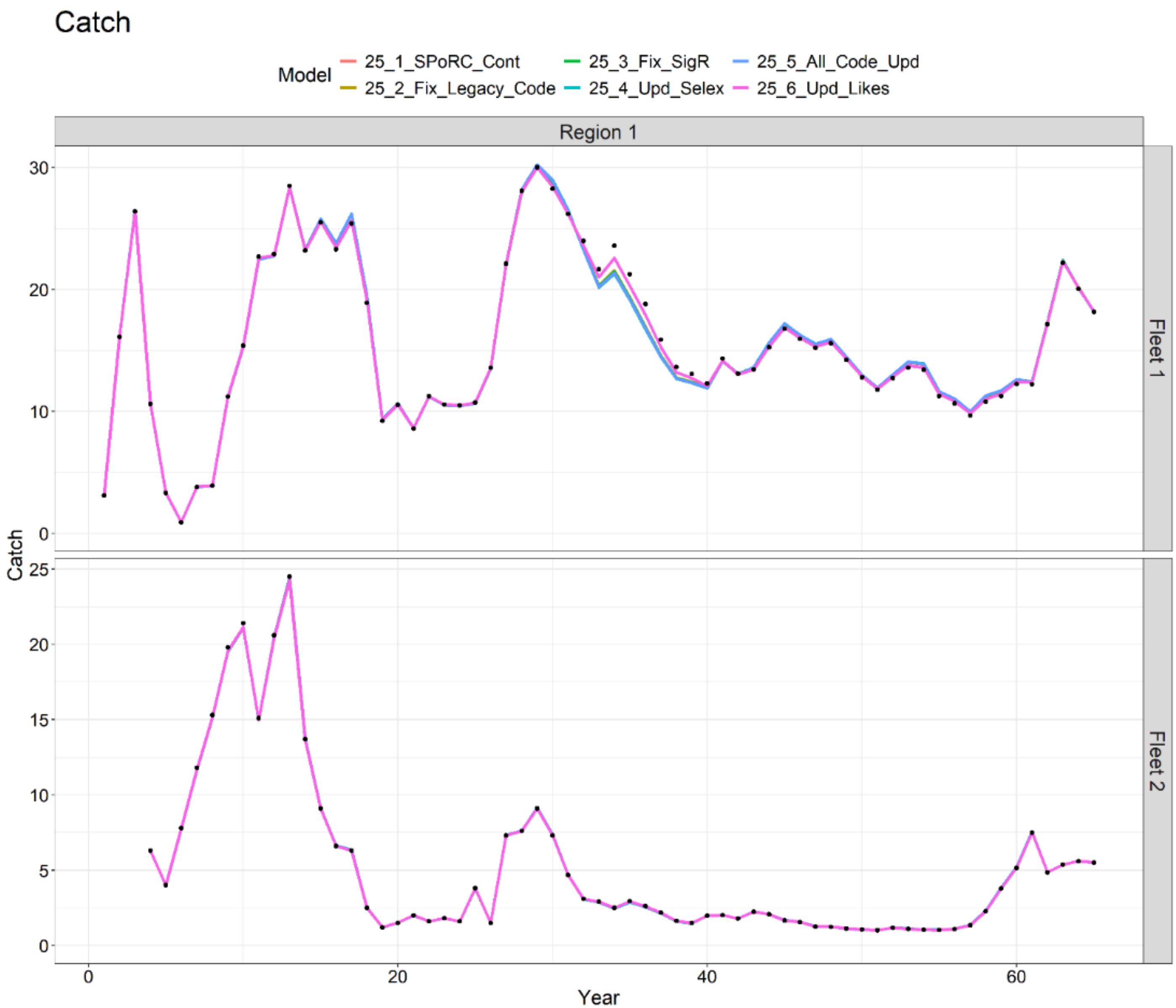


Figure 8. Fits to the fleet specific (fixed gear in first row, trawl gear in second row) catch used in the sablefish assessment for each model run in the ‘Code Fixes and Good Practices’ grouping.

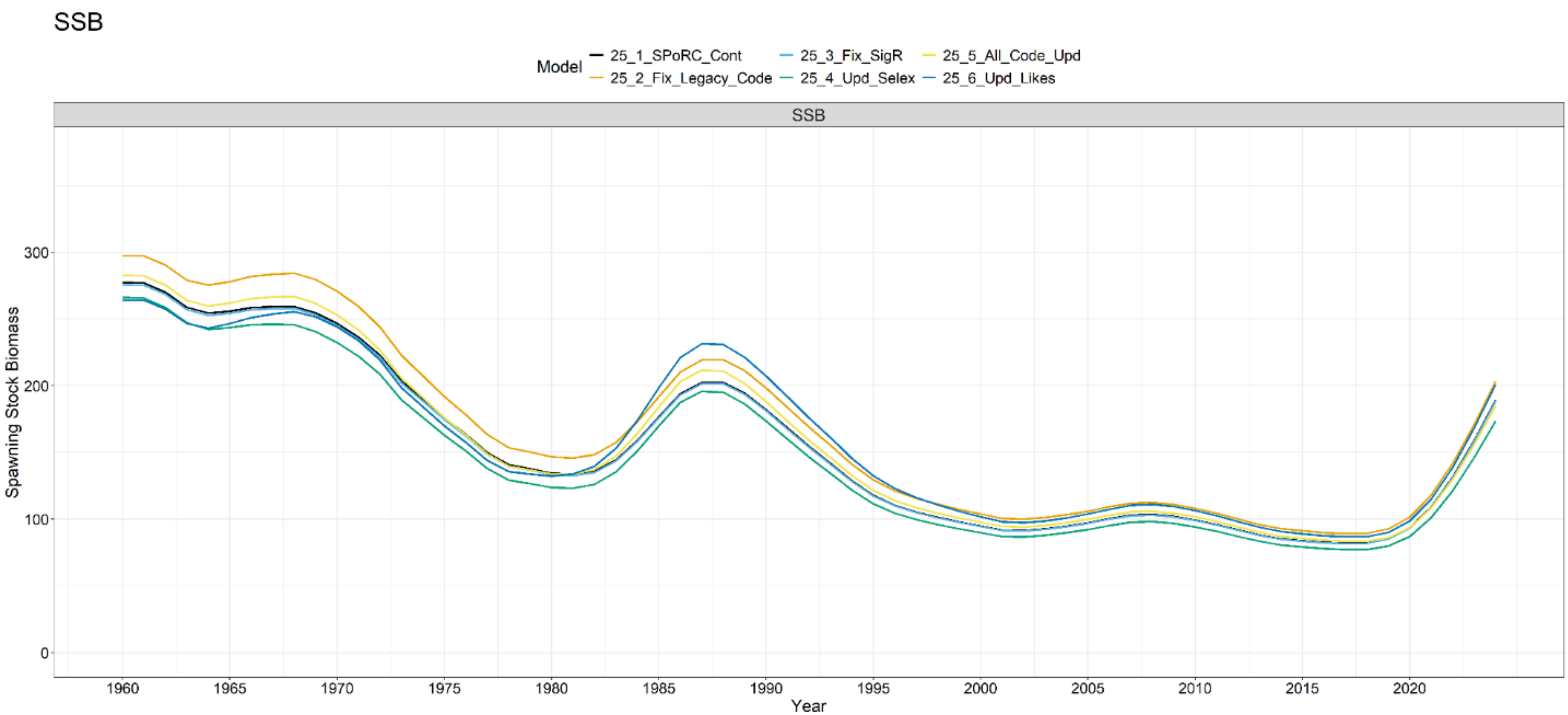


Figure 9. Comparison of spawning stock biomass (SSB, in kilotons) for each model run in the 'Code Fixes and Good Practices' grouping.

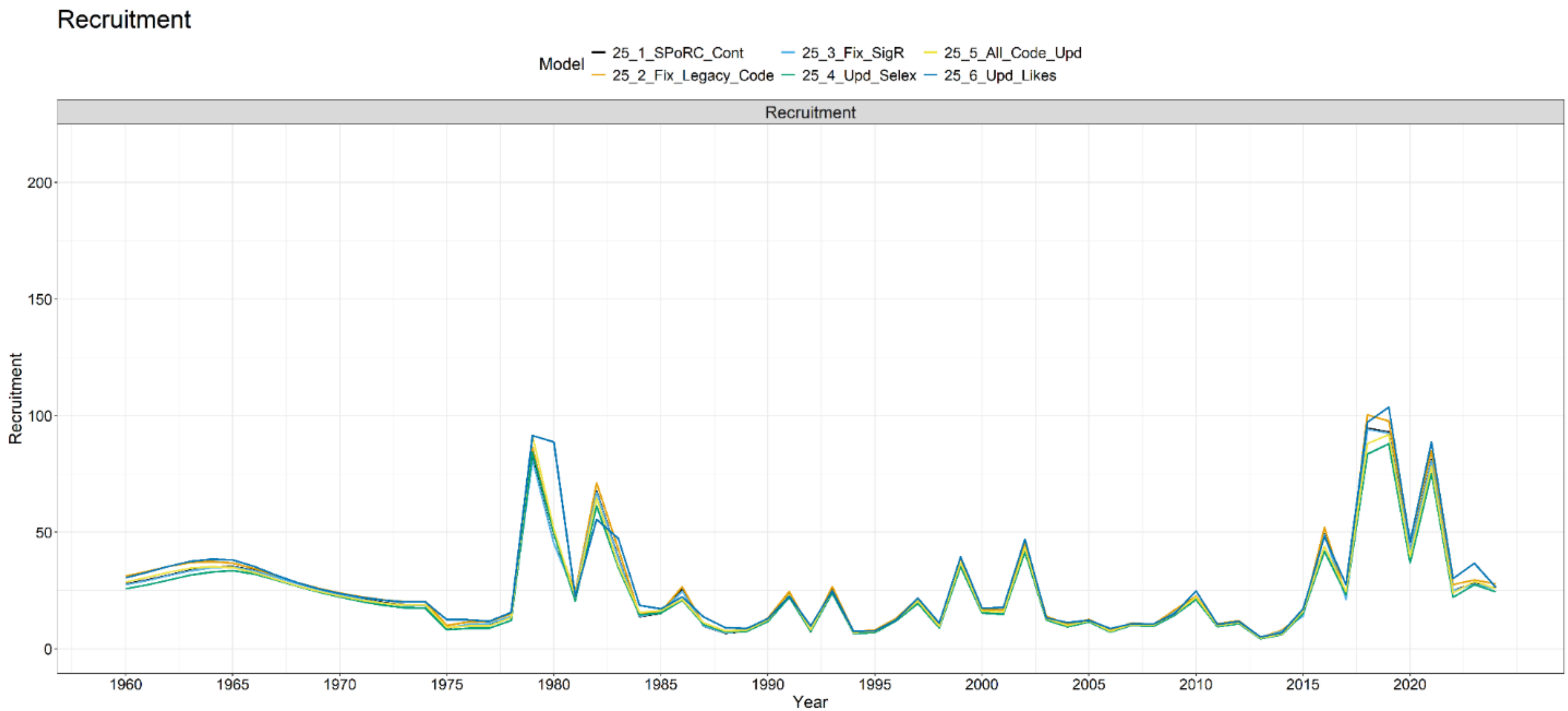


Figure 10. Comparison of recruitment (in millions of fish) for each model run in the ‘Code Fixes and Good Practices’ grouping.

3. Disaggregate Age Compositions

Introduction

Sablefish exhibit sexually dimorphic growth, which manifests as differences in selectivity-at-age by sex in the age-structured assessment model. A key source of model instability is the estimation of selectivity, given the increase in model parameters due to integrating a sex partition. Historically, the sablefish assessment has always fit sex-disaggregated length data, but sex-aggregated age data. Age data has more limited sample sizes and a shorter time series compared to length data. However, age data now exists from the fishery since 1999 and from the NOAA longline survey since 1996, with annual sample sizes of ~1200 otoliths aged for each. The impact of fitting age-aggregated data in a sex-specific assessment is not clear, but it likely smears signals of sex-specific dynamics (e.g., mortality, selectivity, and recruitment), while hindering estimation of sex-specific selectivity. Thus, this section explores the impact of disaggregating age compositions by sex.

Next, model runs were explored that removed length compositions when age compositions were available for a given data source in a given year. For many NPFMC assessed groundfish species where age compositions are routinely collected (e.g., most rockfish), lengths are typically not fit in the assessment if age data are available in that year for a given data source. The approach ensures that age-length pairs used to develop the age-length key are not double counted and then later fit as independent data in the assessment model.

Generally, age compositions provide more direct indications of year class strength and cohort mortality compared to length data. These signals can be lost when fitting length data if there is high variation in length-at-age or small, young fish are not highly selected (Ono et al., 2015). The benefit of simultaneously fitting age and length data in an integrated assessment is an open area of research (Ono et al., 2015), but fitting length data is not recommended when growth is misspecified (i.e., if time-varying growth cannot be adequately replicated in the model; Minte-Vera et al., 2017). Given that sablefish demonstrate density-dependent growth that is not captured in the assessment (Cheng et al., 2024c), there is likely important model tension between age and length data, particularly for recent cohorts. Generally, directly addressing potential model misspecification is recommended to handle data conflicts (Maunder and Wang, 2017). However, when the process causing conflict is either unknown or cannot be adequately modeled, the removal of a specific data set may be warranted, especially if other data sets provide similar information on specific processes (e.g., selectivity and recruitment; Maunder and Piner, 2017). For sablefish, exploratory time-varying growth models were investigated (see Section 5, model *25.13_TV_ALK*), but removal of the length data (when ages were available) was deemed a more parsimonious approach for 2025.

Methods

Whereas model *25.1_SPoRC_Cont* utilizes the split by sex approach for fitting sex disaggregated length compositions (albeit sex-aggregated age compositions), all models in this group implement the joint by sex approach for fitting sex-disaggregated age and length compositions (Cheng et al., 2025a,b). In the split by sex approach, proportions of catch-at-age sum to 1.0 within a given sex, and no sex-ratio information is provided from the composition data. By contrast, the joint by sex approach assumes that compositions sum to 1.0 across both ages and sexes, where sex-ratio information is implicitly provided within these data. Thus, the latter approach inherently provides more information regarding the relative sex structure (e.g., sex ratio at birth), sex-specific mortality, and sex-specific selectivity, given that proportions are fit across sexes simultaneously (Cheng et al., 2025a,b).

The first model in this grouping (model *25.7_Disagg_Age*) explores the dual impact of disaggregating age compositions by sex, while also utilizing the joint approach by sex for fitting compositional data, but otherwise matches the assumptions of the continuity model (*25.1_SPoRC_Cont*). Next, model updates from Section 2, as implemented in model *25.6_Upd_Likes*, are all included, along with disaggregating age compositions and fitting composition data using the joint by sex approach (model *25.8_Age_Upd_Code*). Finally, when age compositions exist for a given data source in a given year, length compositions are removed and not fit in the assessment (model *25.9_Age_Drop_Len*). However, the one exception is for the Japanese longline survey for which limited age (available every other year) and length compositions exist, the data has more uncertainty, and selectivity estimation was somewhat unstable. Thus, both age and length data were kept for all years that they were available for the Japanese longline survey, and since these data were not used to develop age-length keys the threat of double counting was not applicable.

All models in this section underwent Francis reweighting, given the inherent changes in data structure and availability. As noted, the description of results focuses on the final model in the group (*25.9_Age_Drop_Len*) primarily in comparison to continuity model (*25.1_SPoRC_Cont*).

Primary model updates included:

- *25.7_Disagg_Age*:
    - Uses the model settings of the continuity model (*25.1_SPoRC_Cont*).
    - Disaggregates age compositions by sex.
    - Fits all data assuming the joint by sex approach (Cheng et al., 2025a,b).
- *25.8_Age_Upd_Code*:
    - Uses the model settings of model *25.6_Upd_Likes*.
    - Disaggregates age compositions by sex.
    - Fits all data assuming the joint by sex approach (Cheng et al., 2025a,b).
- *25.9_Age_Drop_Len*:
    - Uses the model settings of model *25.8_Age_Upd_Code*.
    - Drops length data for data sources (e.g., fixed gear fishery and NOAA longline survey) and years for which age compositions are available.
    - Maintains both age and length compositions for the Japanese longline survey due to time series and sample size limitations.

Results

By disaggregating age compositions, a number of data conflicts that have caused internal model tension became more readily apparent. In previous models, age composition data was being inherently downweighted, because each sex-disaggregated length composition data set was initially given essentially equivalent weight as each sex-aggregated age composition data set.

Moreover, because the Francis reweighting procedure primarily emphasizes adequate fit to the indices of abundance (as these are assumed to provide the most reliable information on population scale and trend), the reweighting procedure tended to favor composition data that provided consistent signals as the indices (while accounting for correlations among ages or lengths by ensuring mean age or length reflects the mean age and lengths observed in the compositional data; Francis, 2017). Thus, length data has often been given consistently higher weights than age data after Francis reweighting (Table 8), given that the length composition data better supports the index estimates of scaling parameters (e.g., R_0; Figure 11).

Not surprisingly, tension between fits to the age compositions and the longline survey index were readily apparent in the data fits, whereby disaggregating the age compositions by sex (and providing inherently more weight to the age data) led to declining fits to the recent longline survey index (Figure 12). Moreover, the tension among the longline survey index and age composition data was further exacerbated by the increasingly disparate signals between the trawl survey index and the longline survey index in recent years (Figure 12; see Section 4). Overall, disaggregating the age compositions by sex led to moderately improved fits to the age data with minor reductions in fit to the length data (Table 8, Figures 13-14). Removing the length data when age compositions were available (model *25.9_Age_Drop_Len*) led to further improvements in age composition data fits, particularly to the younger (e.g., ages-2-4) and intermediate ages (around age-10), yet some residual patterns remain (Figures 13-14).

Disaggregating the age compositions led to a noticeable increase in natural mortality (~0.12, model *25.8_Age_Upd_Code*), but estimates decreased slightly when length compositions were removed. Thus, it is likely that misspecification in growth is aliasing signals in natural mortality and that the natural mortality prior may be inappropriate (Figure 15). As is expected, changes in natural mortality are concomitant with alterations in selectivity estimates, especially increases in the degree of doming in the trawl fishery selectivity (Figure 16). However, fixed gear fishery selectivity also changed, generally resulting in more moderate estimates of selectivity for ages less than 10 and lower selectivity of age-2 fish (particularly males) when the length compositions were removed. The resultant selectivity curves from model *25.9_Age_Drop_Len* appear more reasonable, while also helping to better fit the age composition data.

Once again, all models provide generally the same trends in SSB, albeit with slight changes in scale (Figure 17). Including all the model updates as well as disaggregating the age compositions results in a moderate increase in scale for model *25.8_Age_Upd_Code* compared to the continuity model *25.1_SPoRC_Cont*, which is likely mostly attributed to the increase in natural mortality. Dropping the length compositions (model *25.9_Age_Drop_Len*) has minimal further impact on SSB estimates, reference points, or management advice. The resultant 2025 ABC from model *25.9_Age_Drop_Len* is nearly identical to that from model *25.1_SPoRC*_Cont, despite considerably lower terminal SSB in the latter (Table 9).

The primary changes across models in the 'Disaggregate Age Compositions' scenario grouping is the interpretation of year class strength (Figure 18). Models with sex-aggregated age compositions tend to support multiple sequential large year classes (e.g., 1977-1978 and 2016-2017). In contrast, models where age compositions are disaggregated by sex tend to estimate singular large pronounced year classes (e.g., 1978, 2016; Figure 18). Moreover, when length compositions are removed (model *25.9_Age_Drop_Len*), the recent recruitment dynamics become less variable and a single large 2016 cohort tends to be emphasized (though most of the 2014-2021 year classes remain generally above average). Therefore, it appears that aggregating age compositions and including length compositions when ages exist act to smear signals in cohort strength. However, removing the length compositions appears to increase the recruitment retrospective pattern in recent years, likely due to the trawl survey length compositions being inherently upweighted and acting as the primary (and often contradictory) indicator of recent year class strength when the longline survey length compositions are removed (Figure 19; See Section 4, model *25.10_Drop_Trwl_Sur*).

Notable changes in model dynamics included:

- *25.7_Disagg_Age* (full results here):
    - Moderate improvements in fit to age compositions, but poorer fit to the recent longline survey index.
    - Minor increases in M, but moderate changes in fishery selectivity which impacts relative fishing mortality estimates.
    - Important refinements in estimates of year class strength, tending to emphasize singular large cohorts instead of multiple sequential large cohorts.
    - Early SSB is higher, but terminal SSB is lower than in the continuity model resulting in a moderate decline in catch advice.
- *25.8_Age_Upd_Code* (full results here):
    - Similar results as *25.7_Disagg_Age*, but with moderate further increases in *M*.
    - SSB is scaled up moderately to account for increasing *M*, which leads to a minor increase in catch advice.
    - Recruitment scale increases, but trends are generally unchanged.
- *25.9_Age_Drop_Len* (full results here):
    - Fits to the age composition data are improved, but further degradation in the fits to the recent longline survey index occur.
    - Natural mortality decreases in line with estimates from the continuity model.
    - SSB scale decreases slightly associated with declines in *M*, and catch advice aligns with that from the continuity model.
    - The recent recruitment time series is further refined, supporting a singular large 2016 year class with large reductions in the size of the 2017 and 2019 year classes, but retrospective patterns in recruitment increase slightly.

In summary, disaggregating age composition data leads to important changes in recruitment trends, especially recent recruitment. Moreover, integrating both age and length composition data appears to be adding model tension and smearing of signals in year class strength. Given that sablefish appear to be undergoing density-dependent changes in growth (Cheng et al., 2024c), it is likely that including both age and length composition data for a given data source may lead to model tension unless time-varying growth can be adequately estimated. Given the added complexity associated with modeling time-varying growth as well as the relatively high quality and quantity of aging data (i.e., ~1200 sablefish otoliths are read annually for the fixed gear fishery and longline survey), dropping length data when age compositions are available is deemed the parsimonious approach at this time. However, work to integrate time-varying growth that can adequately capture cohort-specific dynamics is being explored (see Section 5, model *25.13_TV_ALK*). Therefore, model *25.9_Age_Drop_Len* is recommended from this scenario grouping. However, inherent upweighting of the trawl survey length compositions due to removing the longline survey length data likely gives undue influence to the trawl survey for estimates of recent cohort strength. Moreover, the continued changes in the estimates of natural mortality, likely due to inherent tension between the age and length data, suggest that the M prior should be refined.

Tables

Table 7. Compositional data weights (λ) after Francis reweighting for model *25.1_SPoRC_Cont* (middle column, labelled 'Init_Weight') and model *25.8_Age_Upd_Code* (last column, labelled 'Final_Weight'). The NA values indicate that no weight was needed because the data set was either sex-aggregated for model *25.1_SPoRC_Cont* or the joint by sex approach was utilized such that a single weight was required (as noted in the female row for each data set) for model *25.8_Age_Upd_Code.* Abbreviations include: LLF—longline (fixed gear) fishery; LLS—longline survey; JPN—Japanese; TF—trawl fishery; TS—trawl survey; len—length; F—female; M—male.

Comp_Type	Init_Weight	Final_Weight
LLF_Age_F	0.8261073	0.8219418
LLF_Age_M	NA	NA
LLS_Age_F	3.7922454	3.8083961
LLS_Age_M	NA	NA
JPN_LLS_Age_F	1.3168111	0.5720067
JPN_LLS_Age_M	NA	NA
LLF_Len_F	4.1837057	3.8926972
LLF_Len_M	4.2696935	NA
TF_Len_F	0.3164859	0.3493676
TF_Len_M	0.2293966	NA
LLS_Len_F	1.4379202	1.1628594
LLS_Len_M	1.0705376	NA
TS_Len_F	0.6708833	0.5010063
TS_Len_M	0.4652071	NA
JPN_LLS_Len_F	1.2777281	1.0387938
JPN_LLS_Len_M	0.8575195	NA

Table 8. Negative log-likelihood (nLL) components for each model run in the 'Disaggregate Age Compositions' grouping, including data fit components and parameter prior or penalty components. Note that the fundamental likelihood structure changed for model *25.6_Upd_Likes* making it not directly comparable to previous models, while restructuring of the compositional data and dropping length composition data make comparing nLL values across models in this table inadvisable.

Model	jnLL	M Prior	Recruitment Penalty	Catch nLL	Fishing Mortality Penalty	Survey Index nLL	Fishery Index nLL	Initial Age Penalty	Survey Age nLL	Fishery Age nLL	Survey Length nLL	Fishery Length nLL	Selectivity Penalty
25_1_SPoRC_Cont	768.8210	0.7838709	23.11463	3.688151	6.088752	64.49362	37.72432	0.1631310	197.0703	46.62799	169.03957	220.02667	NA
25_6_Upd_Likes	835.7924	-0.2354034	76.09974	-83.779841	146.315255	-29.36051	-10.47291	0.3208338	189.2229	46.83455	192.90260	235.42033	72.524762
25_7_Disagg_Age	623.2369	1.1178911	22.97678	2.643346	6.345658	57.17526	33.34443	0.1361954	218.6027	50.22008	87.54925	138.07892	5.046348
25_8_Age_Upd_Code	653.9155	0.3244425	75.84462	-84.337557	145.365332	-36.25234	-12.70699	0.1736375	215.7966	51.61723	90.02423	135.82932	72.236892
25_9_Age_Drop_Length	563.6117	-0.4228482	75.25805	-84.405651	145.615963	-37.66538	-11.53491	0.1953826	197.1234	49.75360	75.57116	82.03714	72.085731

Table 9. Estimates of key assessment outputs for each model run in the 'Disaggregate Age Compositions' grouping.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_1_SPoRC_Cont	1	189.0804	0.04037044	49.68588	121.0229	0.08593	1.56235	0.46979
25_6_Upd_Likes	1	200.9191	0.03842778	52.95886	128.0622	0.08665	1.56892	0.44350
25_7_Disagg_Age	1	182.5187	0.04918914	45.61297	113.8113	0.09799	1.60370	0.50200
25_8_Age_Upd_Code	1	221.2632	0.03979598	53.31742	126.1573	0.09078	1.75387	0.43839
25_9_Age_Drop_Length	1	219.7952	0.04162539	50.00421	125.1969	0.09062	1.75560	0.45932

Figures

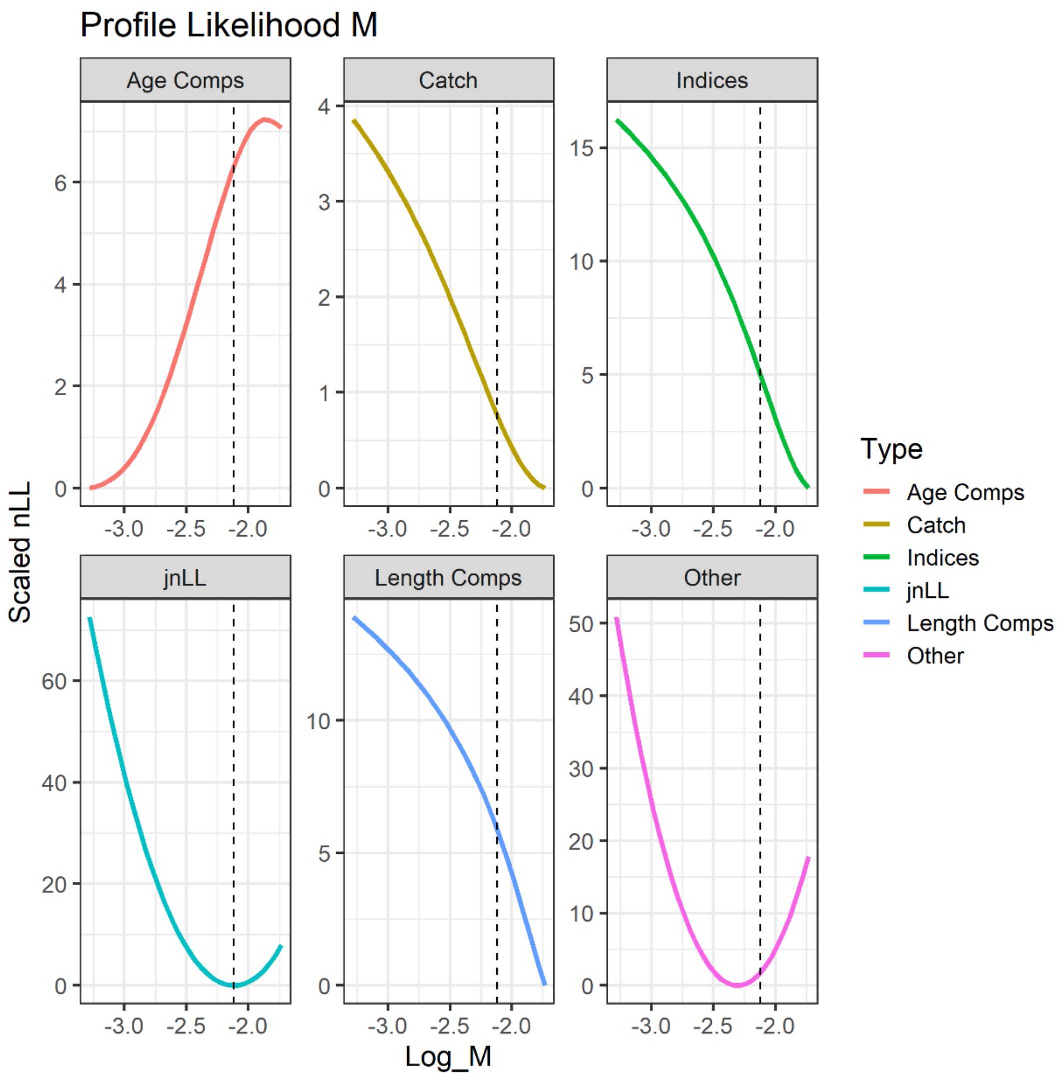


Figure 11. Likelihood profile for the mean recruitment parameter (R_0) from model *25.8_Age_Upd_Code* where panels represent the profile for each data type used in the model.

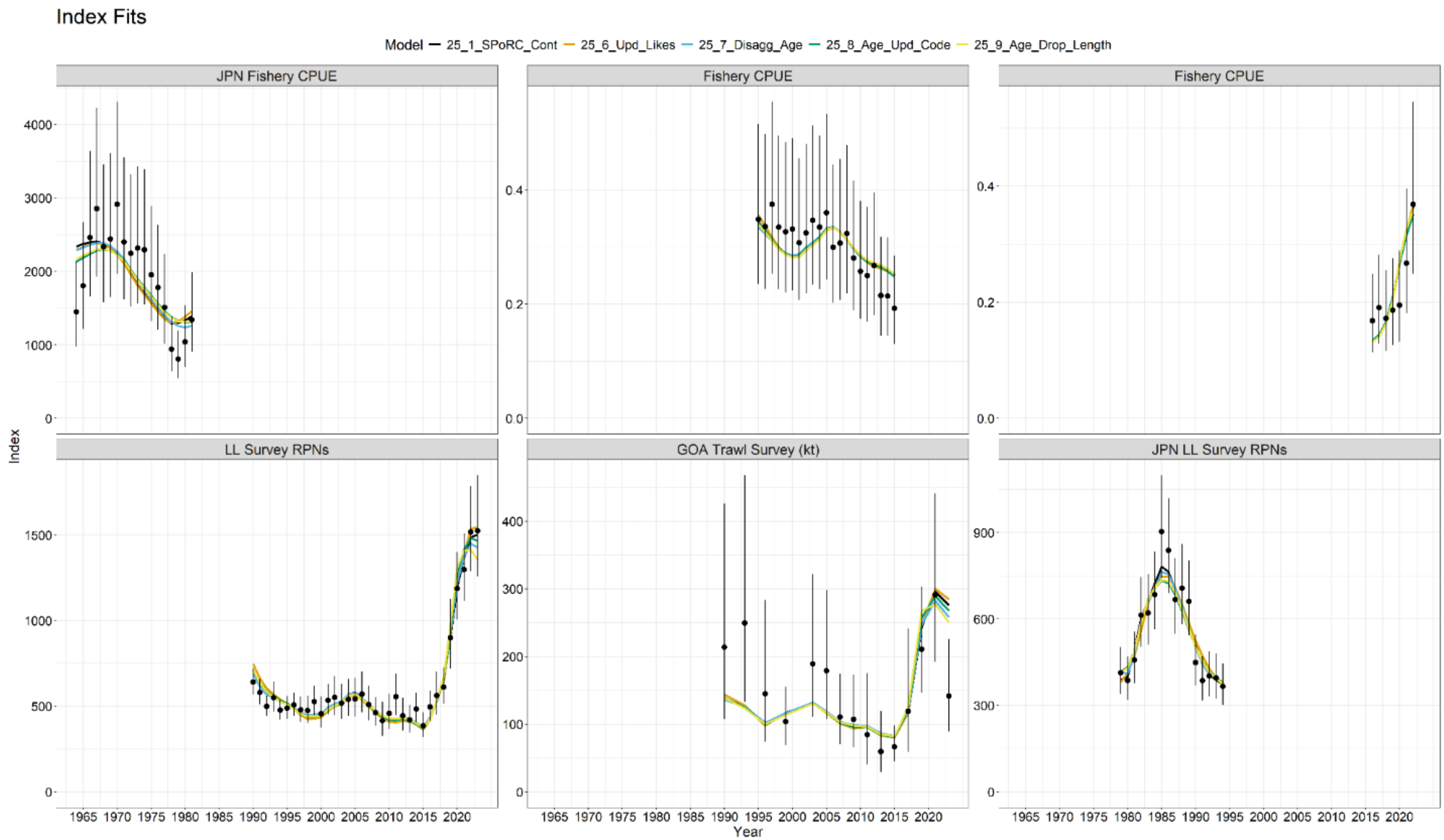


Figure 12. Fits (lines) to the various observed indices of abundance (dots; vertical lines represent 95% confidence intervals) used in the sablefish assessment for each model run in the 'Disaggregate Age Compositions' grouping.

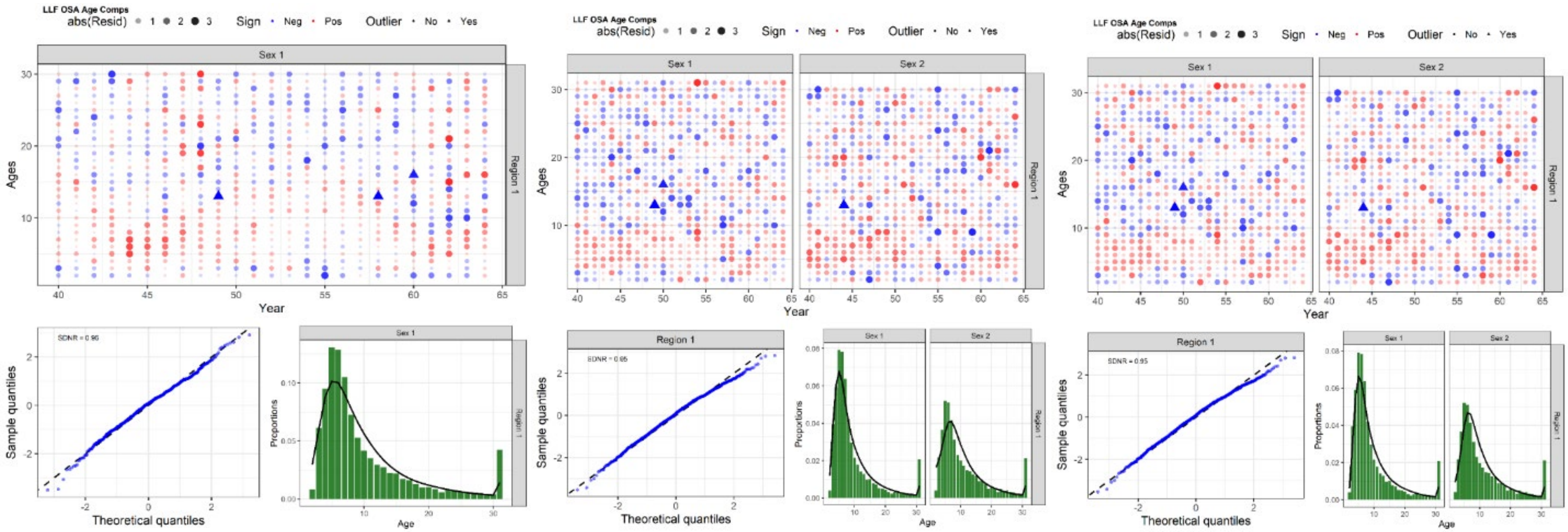


Figure 13. Fits to fixed gear fishery age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model runs in the 'Disaggregate Age Compositions' grouping (left--*25.1_SPoRC_Cont*; middle--*25.8_Age_Upd_Code*; right--*25.9_Age_Drop_Len*).

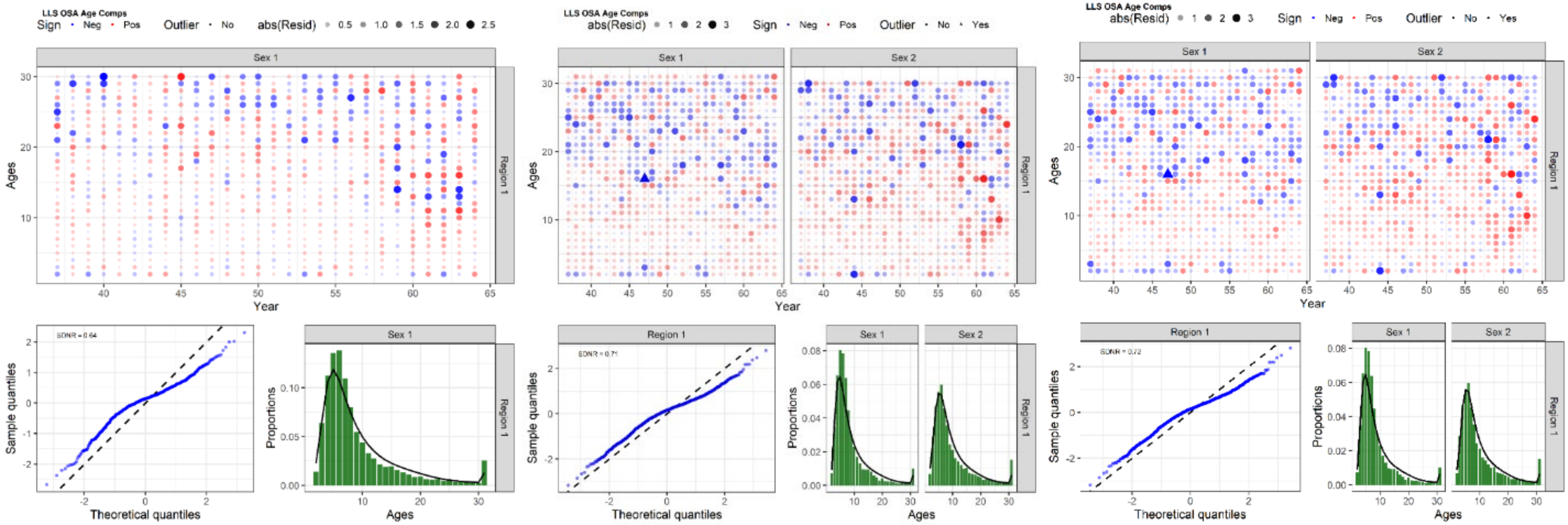


Figure 14. Fits to NOAA longline survey age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model runs in the ‘Disaggregate Age Compositions’ grouping (left--*25.1_SPoRC_Cont*; middle--*25.8_Age_Upd_Code*; right--*25.9_Age_Drop_Len*).

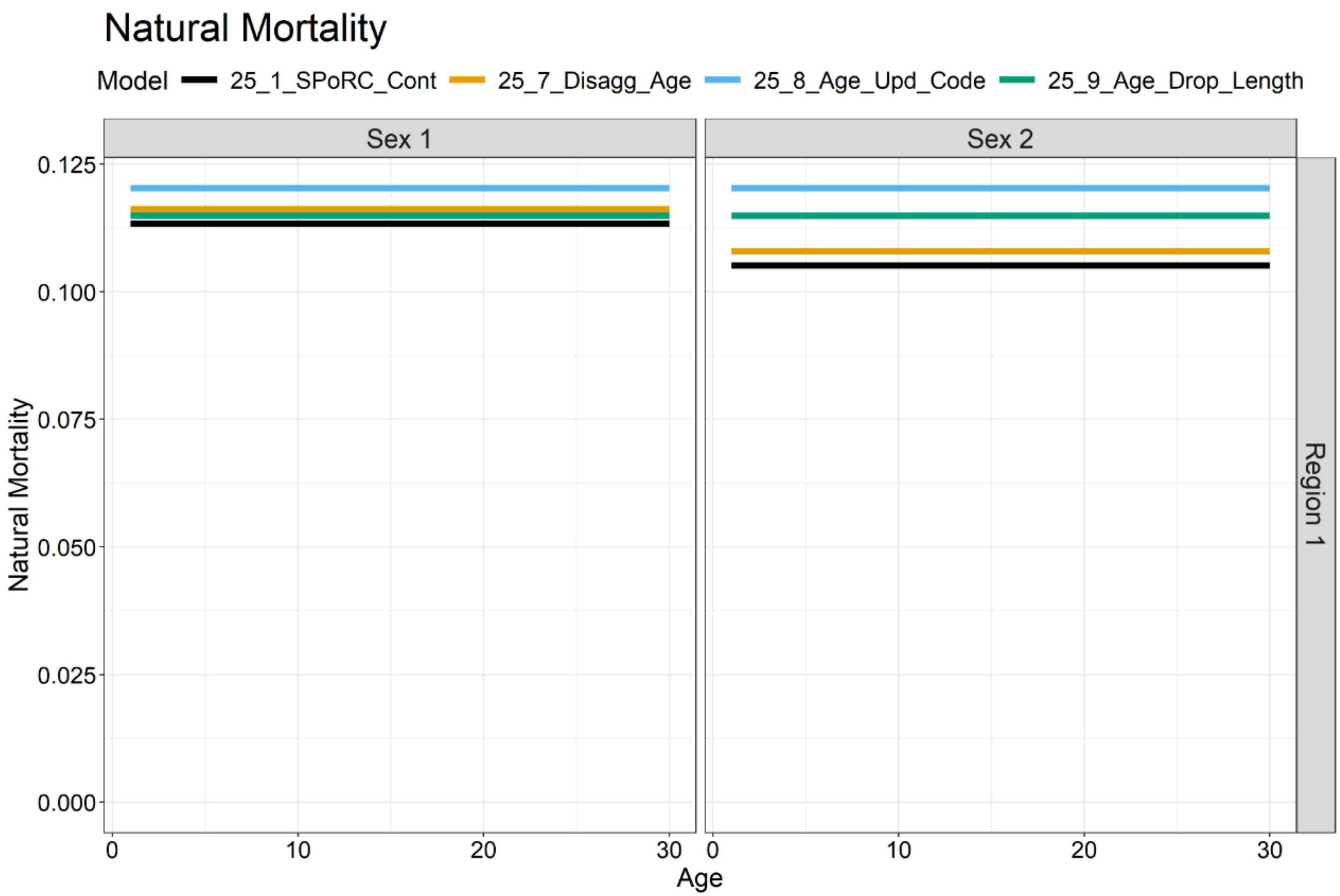


Figure 15. Comparison of natural mortality estimates for each model run in the 'Disaggregate Age Compositions' grouping. Sex 1 is females and sex 2 is males. Note that male natural mortality offset values in models *25_1_SPoRC_Cont* and *25_7_Disagg_Age* were not estimated and were inadvertently fixed less than female natural mortality. Models *25_8_Age_Upd_Code* and *25_9_Age_Drop_Length* correct this issue, and the natural mortality values shown for both sexes are now equivalent (i.e., sex-invariant).

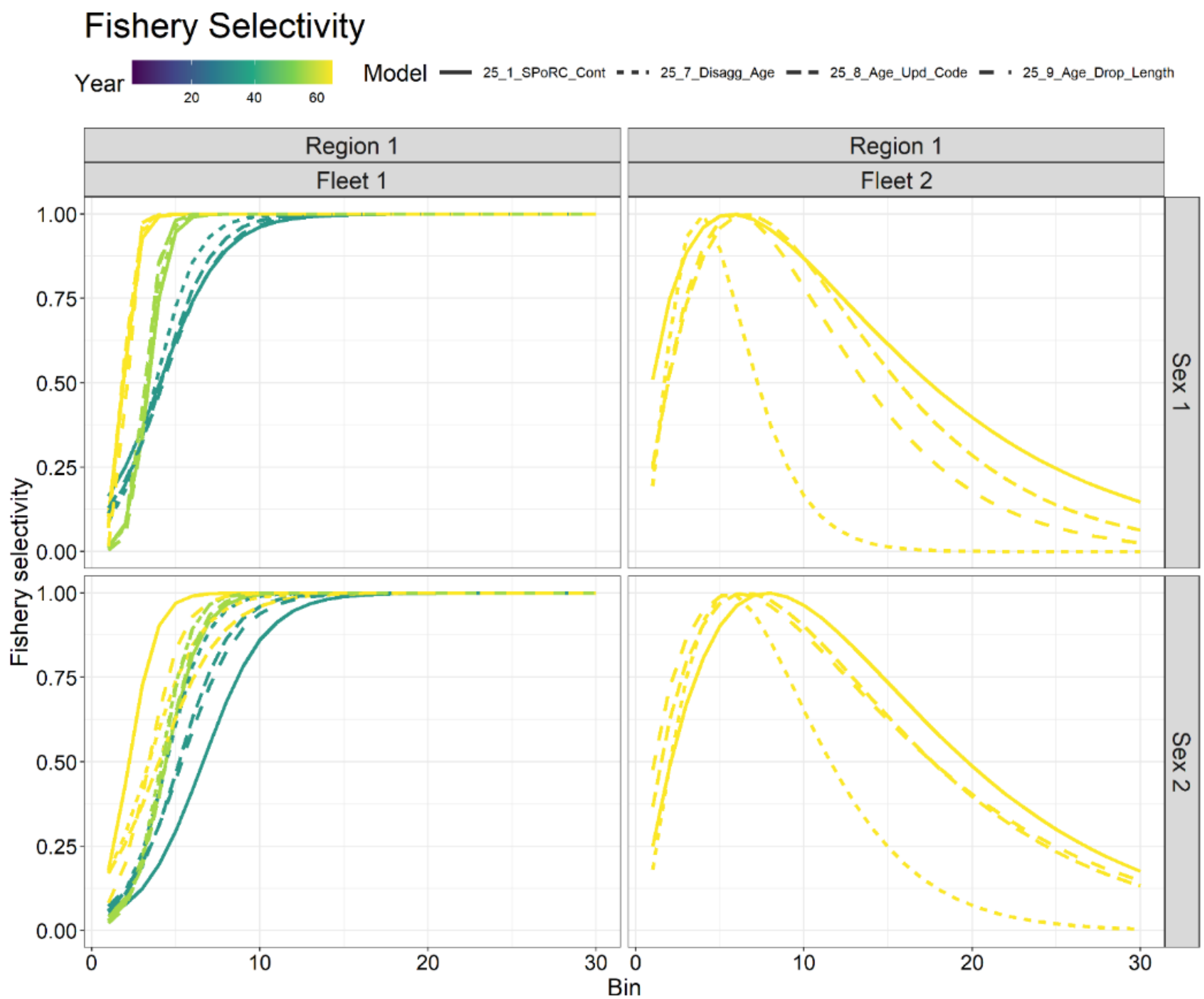


Figure 16. Comparison of fishery selectivity for each model run in the 'Disaggregate Age Compositions' grouping. Sex 1 is females and sex 2 is males. Fleet 1 is the fixed gear fishery and Fleet 2 is the trawl fishery.

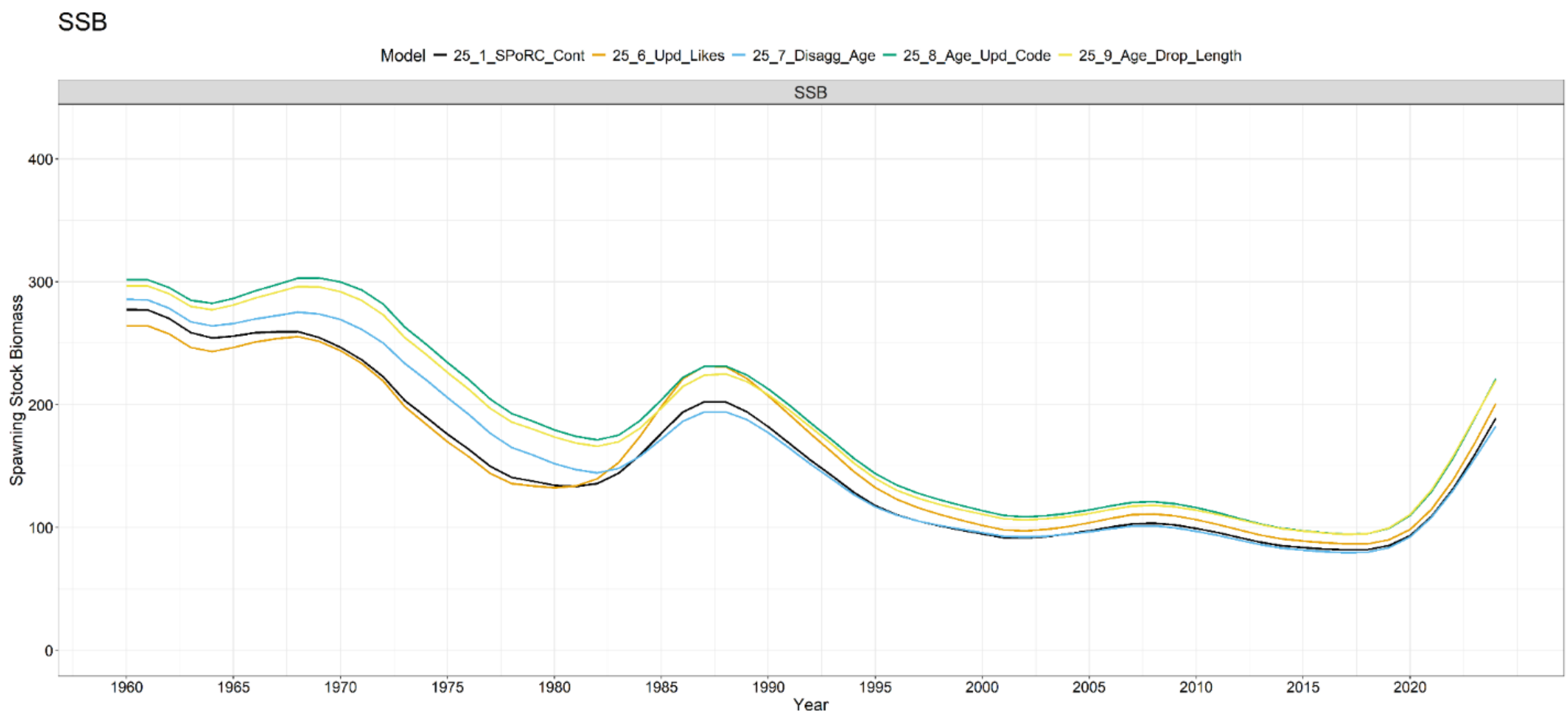


Figure 17. Comparison of spawning stock biomass (SSB, in kilotons) for each model run in the 'Disaggregate Age Compositions' grouping.

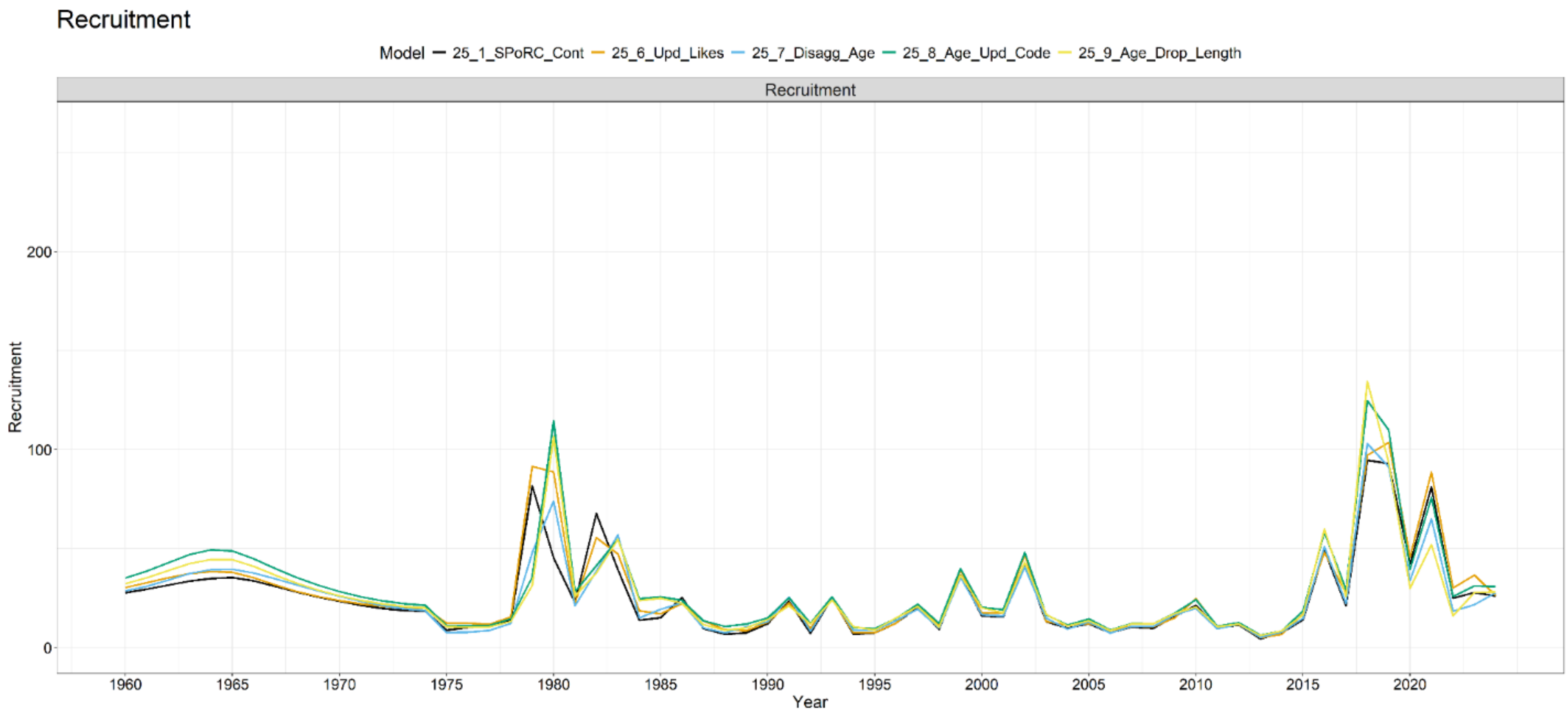


Figure 18. Comparison of recruitment (in millions of fish) for each model run in the 'Disaggregate Age Compositions' grouping.

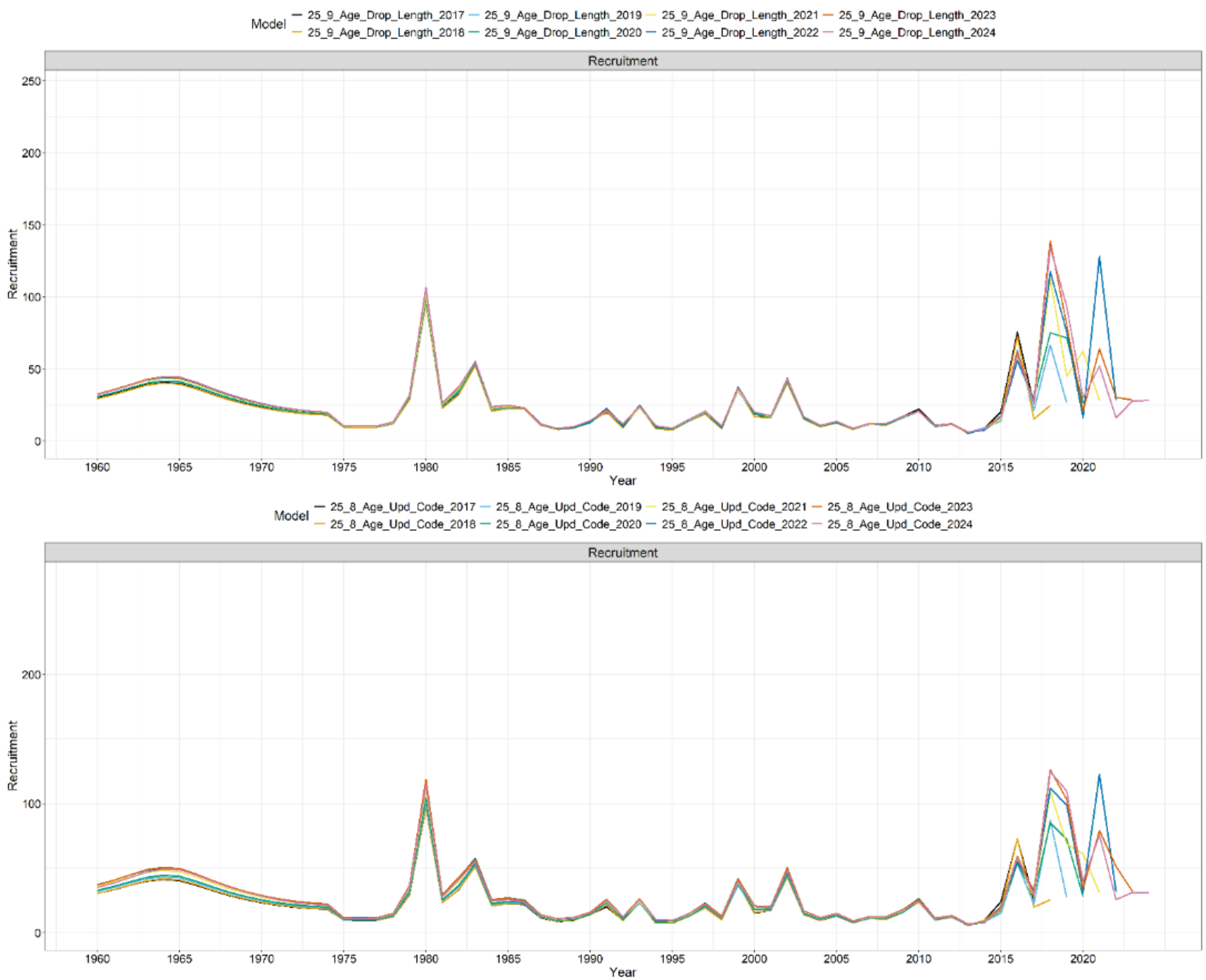


Figure 19. Comparison of the estimated recruitment (in millions of fish) time series from a seven year 'true' retrospective analysis for model *25.9_Age_Drop_Len* (top panel) and *25.8_Age_Upd_Code* (bottom panel). The 'true' retrospective analysis removes a year of data for each peel, but also accounts for the lag in availability of compositional data (i.e., fishery ages and lengths as well as longline survey ages become available on a one year lag from associated catch or index data).

4. Update Data and Model Assumptions

Introduction

The primary index of abundance for the sablefish assessment is the Alaska-wide dedicated sablefish NOAA longline survey that has been conducted annually since 1990 except for 2024, and which was designed to continue the Japanese longline survey time series that dates back to 1979. As such, the longline survey data is a high quality, reliable indicator of both the magnitude and trends in sablefish abundance. Conversely, the NOAA Gulf of Alaska trawl survey is a multispecies survey that imperfectly samples primarily juvenile sablefish only in a portion of their distribution (GOA only, currently representing ~ 50% of the population biomass) and habitat (< 500m). The trawl survey was integrated into the sablefish assessment in 2007 during a period of extremely low recruitment and abundance, because there were concerns that the longline survey may not provide reliable early indicators of incoming recruitment (Hanselman et al., 2007). Given that the trawl survey tended to sample small, young sablefish, it was determined that it could be a useful index of new recruits and juvenile sablefish:

> "Trawl survey abundance indices were not previously used in the sablefish assessment because they were not considered good indicators of the sablefish relative abundance. However, there is a long time series of data available and given the trawl survey's ability to sample smaller fish, it may be a better indicator of recruitment than the longline survey.… At this point we have experimented with using only the Gulf of Alaska trawl survey biomass estimates (<500 m) and length data (<500 m) as an index for the whole population (since the largest proportion of the population is located there)." (Hanselman, 2007).

After almost 10 years of above average recruitment, it is now apparent that the longline survey provides adequate indicators of incoming recruitment strength via age compositions despite selectivity at age-2 being relatively low. Moreover, signals regarding recent cohort strength in the longline survey data have recently countered those observed in the trawl survey, and the indices themselves strongly diverged in 2023 (Goethel and Cheng, 2024; Figure 20). Thus, there is potential that recent recruitment variability and retrospective patterns may be due to inclusion of the trawl survey data (as noted in Section 3).

Similarly, there is increasing reason to believe that the trawl survey data is not representative of the Alaska-wide population. As noted by Maunder and Piner (2017), "An example where data maybe complete, but not representative, is a survey that only partially covers the spatial distribution of the stock, and therefore trends in the index do not represent that of the entire stock." Non-representative data may be considered for removal from an assessment, especially if other data sources exist that better reflect the dynamics of the entire population. Therefore, we explore removal of the GOA trawl survey data (model *25.10_Drop_Trwl_Sur*), given that it is not deemed representative of the Alaska-wide population, it only partially surveys adult sablefish habitat, the associated selectivity is poorly estimated and specified (i.e., assuming an exponential decay power function), and it provides contradictory information on year class strength compared to the dedicated sablefish longline survey.

Natural mortality is a critical scaling parameter that can be extremely difficult to estimate within a stock assessment model, but directly impacts SSB, fishing mortality, and catch advice (Hamel and Cope, 2022). Because natural mortality is poorly estimated in the sablefish assessment, a prior of [$\sim N(\ln(0.1), 0.1^2)$] has been used, which was last revisited in 2018 and based on a meta-analysis and a weighted ensemble approach (Hanselman et al., 2018). Despite sablefish being a relatively long-lived species (maximum observed age in Alaska is 94 years), the natural mortality for Alaskan sablefish in the assessment is estimated to be 0.115 yr^{-1}, which roughly corresponds to a longevity of 45 years (Hamel and Cope, 2022). As noted in Section 3, sablefish natural mortality estimates are relatively unstable with conflicting information coming from the age compared to length data. As Wang and Maunder (2017) emphasize "much of the information about population scale contained in age-composition data is lost if natural mortality is estimated, which implies that misspecifying natural mortality can substantially bias the information about population scale contained in age-composition data." The current prior [$\sim N(\ln(0.1), 0.1^2)$] allows for natural mortality estimates that appear to be too high for sablefish given their known longevity. Moreover, Hanselman et al. (2018) concluded that this prior "…actually resulted in a slightly degraded fit. This may indicate that the age data in the sablefish assessment has informative data on natural mortality, and that information is more consistent with the mark-recapture estimate of *M*."

Thus, the natural mortality prior was updated to ensure that resulting model estimates better aligned with known biology (model *25.11_Upd_M_Pr*). We developed a new prior [$\sim N(\ln(0.085), 0.1^2)$] based on estimates of natural mortality from previous tagging models (0.0852 from Hanselman et al., 2018), while also considering realistic average longevity values and estimates of sablefish natural mortality from assessments on the US west coast and British Columbia. In terms of longevity, a mean of 0.085 corresponds to a maximum age of ~65 years, which was deemed appropriate, on average, because there are only 6 fish >70 years in the survey age compositions. The US west coast trawl survey observes slightly more older sablefish, and the associated longevity based prior in the west coast sablefish assessment was based on a maximum age of 75 resulting in a prior of $\sim N(0.072, 0.31^2)$. The final estimate of natural mortality from the west coast sablefish assessment was 0.088 (Wetzel et al., 2025). For British Columbia, sex-specific natural mortality is estimated as part of the management procedure process, where female *M* is ~0.094 and male *M* is ~0.05 (DFO, 2023).

The final model in this grouping explored implementing both updates (removing the trawl survey data and updating the natural mortality prior) simultaneously (model *25.12_Drop_TS_Upd_M*).

Methods

Three models were developed that explored removing the trawl survey data (model *25.10_Drop_Trwl_Sur*), using the updated *M* prior (model *25.11_Upd_M_Pr*), and performing both of these updates simultaneously (model *25.12_Drop_TS_Upd_M*). All models in this section built upon the terminal model in Section 3 (model *25.9_Age_Drop_Len*) and underwent Francis reweighting. For model *25.11_Upd_M_Pr*, exploratory runs with differing mean *M* values used for the prior were also run to demonstrate the impact of the choice of prior on the model outputs. As noted, the description of results focuses on the final model in the group (*25.12_Drop_TS_Upd_M*), primarily in comparison to continuity model (*25.1_SPoRC_Cont*), and a thorough comparison of model outputs and diagnostics between these two models is provided in Appendix 2.

Primary model updates included:

- *25.10_Drop_Trwl_Sur*:
    - Remove the trawl survey index and associated length compositions from the model.
- *25.11_Upd_M_Pr*:
    - Change the natural mortality prior from $\sim N(\ln(0.1), 0.1^2)$ to $\sim N(\ln(0.085), 0.1^2)$.
- *25.12_Drop_TS_Upd_M*:
    - Remove the trawl survey data and implement the new natural mortality prior.

Results

Removing the trawl survey index and length compositions (model *25.10_Drop_Trwl_Sur*) led to very modest improvements in the fit to the NOAA longline survey index along with age compositions from the fishery and longline survey (Figure 20, Table 10). Thus, further indications are provided of the contradictory nature of the trawl survey data compared to the other data sources. Updating the natural mortality prior (model *25.11_Upd_M_Pr*) had minimal impact on data fits, but highlighted the inherent tension among data components related to model scaling (Figure 21). However, the new natural mortality estimate did lead to slightly better alignment with the information contained in the age composition data.

Not surprisingly, the biggest impact from model runs in this grouping was a moderate decline in the natural mortality estimate (from ~0.115 to ~0.1) when the new prior was implemented (model *25.11_Upd_M_Pr*; Figure 24). Although the new natural mortality estimate better aligns with known biology and longevity estimates for sablefish, it is still slightly higher than might be expected for a long-lived species. Unfortunately, exploratory runs that examined loosening the prior variance were unable to adequately converge, which further supports previous inference that M is not freely estimable in the current model formulation. Given that natural mortality has a direct impact on model scaling, there is an associated increase in relative fishing mortality and a decline in SSB as natural mortality decreases (Figure 25, results are from exploratory runs with alternate M prior mean values). Concomitant with the decrease in the scale of SSB (Figure 26), the resulting catch advice is also reduced (Table 11). Because sablefish are now assumed to have an increased longevity (increased survivorship), $F_{40\%}$ must also decrease to ensure adequate survival to mature ages and sufficient reproductive potential.

Although removal of the trawl survey data had minimal impact on model scaling and SSB, there were important refinements to estimates of recent cohort strength (Figure 27). In particular, recruitment variability over the last decade is further reduced, while the magnitude of the 2016 year class increases yet again. Changes to the M prior had little impact on recruitment, though, the mean recruitment did decline in conjunction with rescaling of the SSB.

The final model in this grouping (model *25.12_Drop_TS_Upd_M*), which removed the trawl survey data and utilized the new M prior, generally had the same SSB rescaling of model *25.11_Upd_M_Pr* along with the refinement in recruitment of model *25.10_Drop_Trwl_Sur*

(Figures 26-27). The overall catch advice decreased by ~10kt from the continuity model (*25.1_SPoRC_Cont*), which was strongly driven by the new natural mortality estimate. Compared to the final model in Section 3 (model *25.9_Age_Drop_Len*), a critical improvement in model *25.12_Drop_TS_Upd_M* is a reduction in the recruitment retrospective (Figure 28). Removal of the trawl survey data appears to remove early signals of a strong 2017 and 2019 year class, which apparently contradict later signals from the NOAA longline survey and fixed gear fishery age compositions. Conversely, model *25.12_Drop_TS_Upd_M* does initially underestimate the large magnitude 2016 year class, but quickly rectifies initial estimates. Thus, it is likely that model *25.12_Drop_TS_Upd_M* is more precautionary (i.e., has less rapid quota increases) than previous models when extreme recruitment events initially enter the population. However, estimates of year class strength (and recommended quotas) generally stabilize by the time a cohort is about 50% mature (i.e., age-7/8).

Notable changes in model dynamics included:

- *25.10_Drop_Trwl_Sur* (full results here):
    - Minimal improvement in fit to the NOAA longline survey index as well as survey and fishery age compositions.
    - Refined and less variable estimates of recent year class strength, with increased estimates of the 2016 year class and reductions in the size of the 2017 and 2019 year classes.
    - General reduction in recent recruitment retrospective patterns where increased consistency is observed across recent retrospective peels, but with a tendency to initially underestimate the extreme 2016 year class until age-7/8.
- *25.11_Upd_M_Pr* (full results here):
    - Limited impact on data fits.
    - Moderate reduction in *M* (~0.1) resulting in associated increases in fishing mortality and decreases in the scale of SSB.
    - Reduction in catch advice
- *25.12_Drop_TS_Upd_M* (full results here):
    - Slight improvement in data fits as in model *25.10_Drop_Trwl_Sur.*
    - Reduction in SSB scale and catch advice reflecting those of model *25.11_Upd_M_Pr.*
    - Improved recruitment retrospective patterns and reduced recruitment variability for recent cohorts.

Overall, removing the trawl survey data appears to reduce spurious estimates of recent strong year classes (e.g., 2017 and 2019), which are not supported by age composition data from either the longline survey or the fixed gear fishery age compositions. Similarly, the new *M* prior results in a more biologically reasonable estimate of natural mortality (~0.1). Therefore, given improved model diagnostics and biological realism, model *25.12_Drop_TS_Upd_M* was chosen as the final model for the 'Update Data and Model Assumptions' group as well as the author recommended model for the 2025 Alaskan sablefish stock assessment.

A number of further model sensitivity runs (see Section 5) were undertaken to explore the handling of age and length composition data (e.g., estimation of time-varying processes such as growth and selectivity, along with exploration of alternative composition likelihoods). However, more work was deemed necessary before recommending any of these models as the basis of management advice, given the general increase in model complexity required. In the future, the lead authors also plan to further explore the handling of sex-specific dynamics, particularly estimation of sex-specific natural mortality and potentially recruitment sex ratio. Given results of exploratory runs with and without Francis reweighting, it appears that the downweighting of age composition information inherent in the Francis approach (to ensure adequate abundance index weights) may result in the loss of signal regarding natural mortality and, likely, sex-specific dynamics. Initial explorations with self-weighting compositional likelihoods (e.g., the logistic normal, model *25.16_2dLN*) that explicitly account for correlations among ages and sexes have demonstrated promise for more freely estimating sex-specific natural mortality (see Section 5). Moreover, revisiting aging error assumptions may be worthwhile for sablefish, given the intertwined nature of recruitment estimation, age composition observation error, and aging error. Similarly, methods for aggregating regional compositional data for use in the single region assessment (i.e., catch-weighted age and length compositions) may also warrant further exploration, given important differences in apparent year class strength across regions (e.g., in the BS and AI compared to the GOA for the 2016 and 2017 year classes, see Section 5, model *25.15_Spatial*) and plus group catch (e.g., the excessive AI plus group size in the early 2010s).

Conclusion

For the 2025 Alaska sablefish stock assessment, the author recommended model is *25.12_Drop_TS_Upd_M*. This final model integrates updates across four sets of model changes, including moving from ADMB to R-TMB (i.e., the SPoRC package; model *25.1_SPoRC_Cont*), resolving minor legacy coding issues and improving model stability via good practice implementation (model *25.6_Upd_Likes*), disaggregating age compositions and removing length compositions when ages are available (model *25.9_Age_Drop_Len*), and removing non-representative trawl survey data and updating the natural mortality prior to be more biologically reasonable (model *25.12_Drop_TS_Upd_M*). The resulting author recommended model better adheres to stock assessment good practices, generally improves fits to the available data, and is more stable leading to more consistent estimates. However, model tension among data components (e.g., between age compositions and survey indices) is not fully resolved, while important tradeoffs resulted among certain data sets (e.g., improved fits to age composition data at the expense of slightly reduced fits to the most recent NOAA longline survey index values).

Despite important model and data changes across the range of updates explored for 2025, SSB trajectories are generally consistent across model runs and fall within the 95% confidence intervals of the continuity model (Figure 29). The recruitment time series across model runs are slightly more variable, which is to be expected given important changes to data processing (i.e., disaggregating age compositions) and removal of data sets (e.g., length compositions for data sources that have age compositions along with the trawl survey index and length compositions) that directly influence estimates of year class strength and variability (Figure 30). In particular, recent year class dynamics have undergone important changes, which suggest that the 2016 year class is larger than previously estimated, but that subsequent recruitment events (e.g., the 2017 and 2019 year classes) were smaller.

Although stock status has not changed drastically across models, important model rescaling (e.g., in SSB) due to changes in estimates of natural mortality and associated changes in survivorship have led to moderate reductions in catch advice (Table 12). Moreover, the final model is likely slightly more precautionary than previous models due to moderate initial underestimation of the 2016 year class based on recruitment retrospective patterns (Figure 28). Overall, the authors feel that model *25.12_Drop_TS_Upd_M* fits the data slightly better, reduces retrospective bias, improves model stability, and provides a more biologically realistic assessment model, and should form the basis for management advice in 2025. However, as with all assessments, the model is not perfect and ongoing work will continue in the coming years to better inform sex-specific dynamics and address model tension among data inputs.

Tables

Table 10. Negative log-likelihood (nLL) components for each model run in the 'Data and Model Updates' grouping, including data fit components and parameter prior or penalty components. Note that the changes in data used and structure of priors make comparing nLL values across models in this table inadvisable.

Model	jnLL	M Prior	Recruitment Penalty	Catch nLL	Fishing Mortality Penalty	Survey Index nLL	Fishery Index nLL	Initial Age Penalty	Survey Age nLL	Fishery Age nLL	Survey Length nLL	Fishery Length nLL	Selectivity Penalty
25_1_SPoRC_Cont	768.8210	0.7838709	23.11463	3.688151	6.088752	64.49362	37.72432	0.1631310	197.0703	46.62799	169.03957	220.02667	NA
25_9_Age_Drop_Length	563.6117	-0.4228482	75.25805	-84.405651	145.615963	-37.66538	-11.53491	0.1953826	197.1234	49.75360	75.57116	82.03714	72.08573
25_10_Drop_Trwl_Sur	520.1376	-0.3317710	74.71358	-84.485935	145.654711	-43.49581	-10.91950	0.1927378	188.4512	48.43536	57.46386	82.53831	61.92076
25_11_Upd_M_Pr	515.6982	-0.3081874	75.25098	-136.167056	146.946437	-36.65904	-11.54096	0.2849454	198.4586	49.73668	74.45816	83.22979	72.00785
25_12_Drop_TS_Upd_M	471.8339	-0.2370536	74.67764	-136.218750	146.995480	-42.64644	-10.95465	0.2847093	189.5774	48.50805	56.40044	83.61324	61.83381

Table 11. Estimates of key assessment outputs for each model run in the 'Data and Model Updates' grouping.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_1_SPoRC_Cont	1	189.0804	0.04037044	49.68588	121.0229	0.08593	1.56235	0.46979
25_9_Age_Drop_Length	1	219.7952	0.04162539	50.00421	125.1969	0.09062	1.75560	0.45932
25_10_Drop_Trwl_Sur	1	241.1688	0.03880675	53.56695	129.4799	0.09093	1.86260	0.42678
25_11_Upd_M_Pr	1	178.3803	0.05147170	37.28203	123.4346	0.08281	1.44514	0.62155
25_12_Drop_TS_Upd_M	1	194.0907	0.04842486	39.55329	126.6058	0.08308	1.53303	0.58289

Table 12. Estimates of key assessment outputs for the final model runs in each model grouping.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_1_SPoRC_Cont	1	189.0804	0.04037044	49.68588	121.0229	0.08593	1.56235	0.46979
25_6_Upd_Likes	1	200.9191	0.03842778	52.95886	128.0622	0.08665	1.56892	0.44350
25_9_Age_Drop_Length	1	219.7952	0.04162539	50.00421	125.1969	0.09062	1.75560	0.45932
25_12_Drop_TS_Upd_M	1	194.0907	0.04842486	39.55329	126.6058	0.08308	1.53303	0.58289

Figures

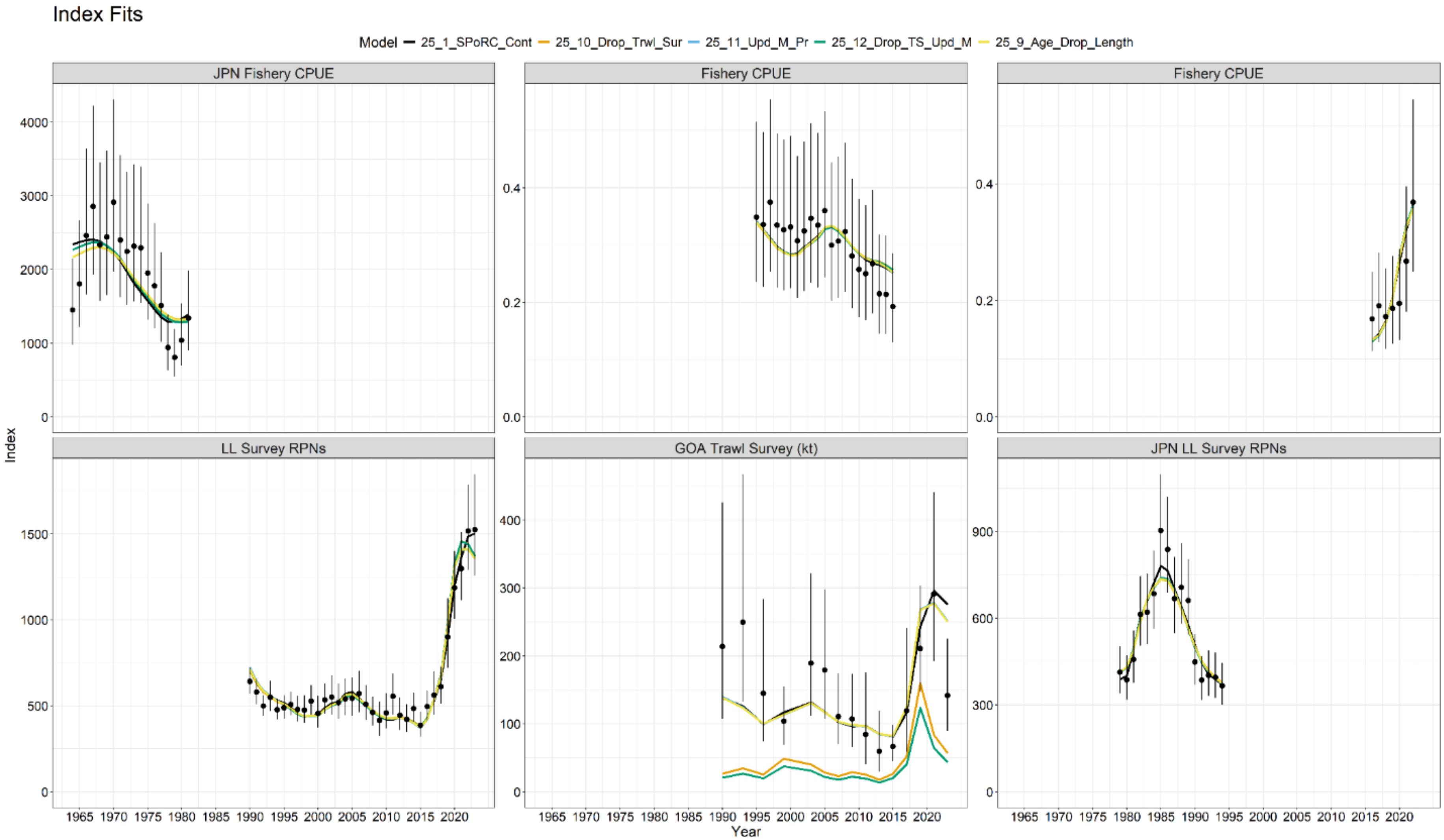


Figure 20. Fits (lines) to the various observed indices of abundance (dots; vertical lines represent 95% confidence intervals) used in the sablefish assessment for each model run in the 'Data and Model Updates' grouping. Note that models *25.10_Drop_Trwl_Sur* and *25.12_Drop_TS_Upd_M* do not fit the trawl survey index, so the expected values for the index should be ignored for these models.

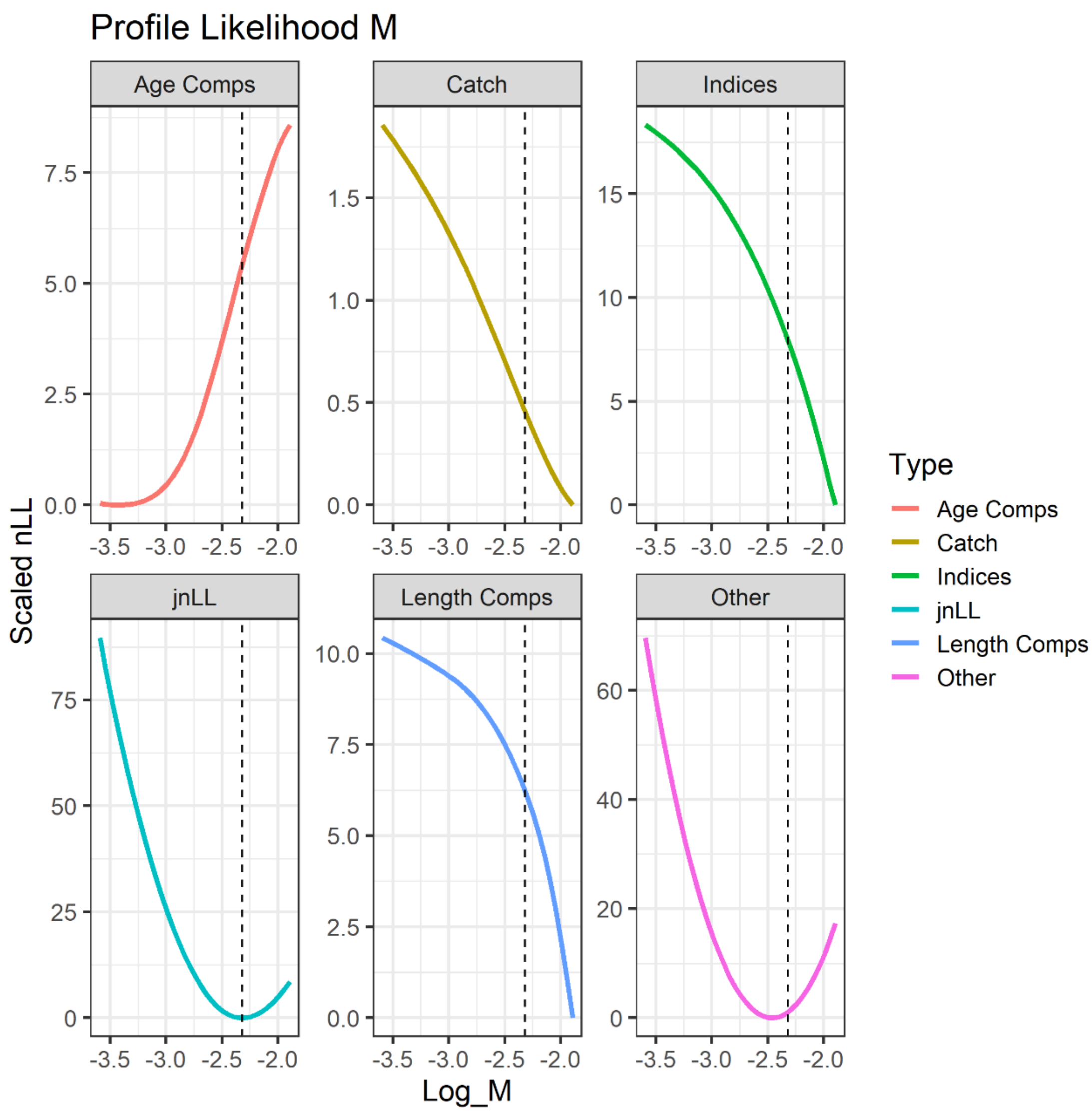


Figure 21. Likelihood profile for the natural mortality parameter (*M*) from model *25.11_Upd_M_Pr* where panels represent the profile for each data type used in the model.

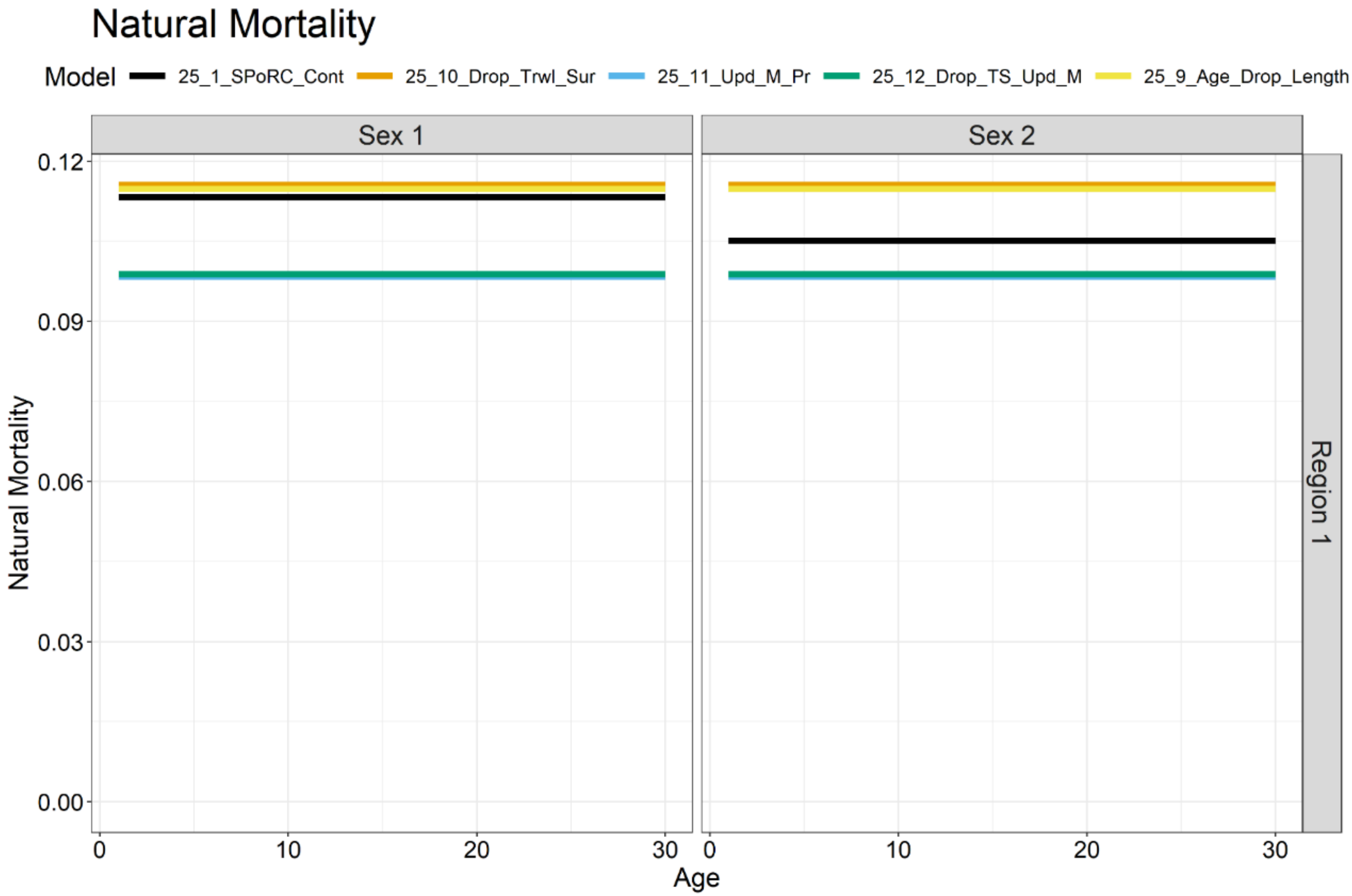


Figure 24. Comparison of natural mortality for each model run in the 'Data and Model Updates' grouping. Sex 1 is females and Sex 2 is males. Note that the male natural mortality offset value in model *25_1_SPoRC_Cont* was not estimated and was inadvertently fixed less than female natural mortality. All other models depicted here correct this issue, and the natural mortality values shown for both sexes are equivalent (i.e., sex-invariant).

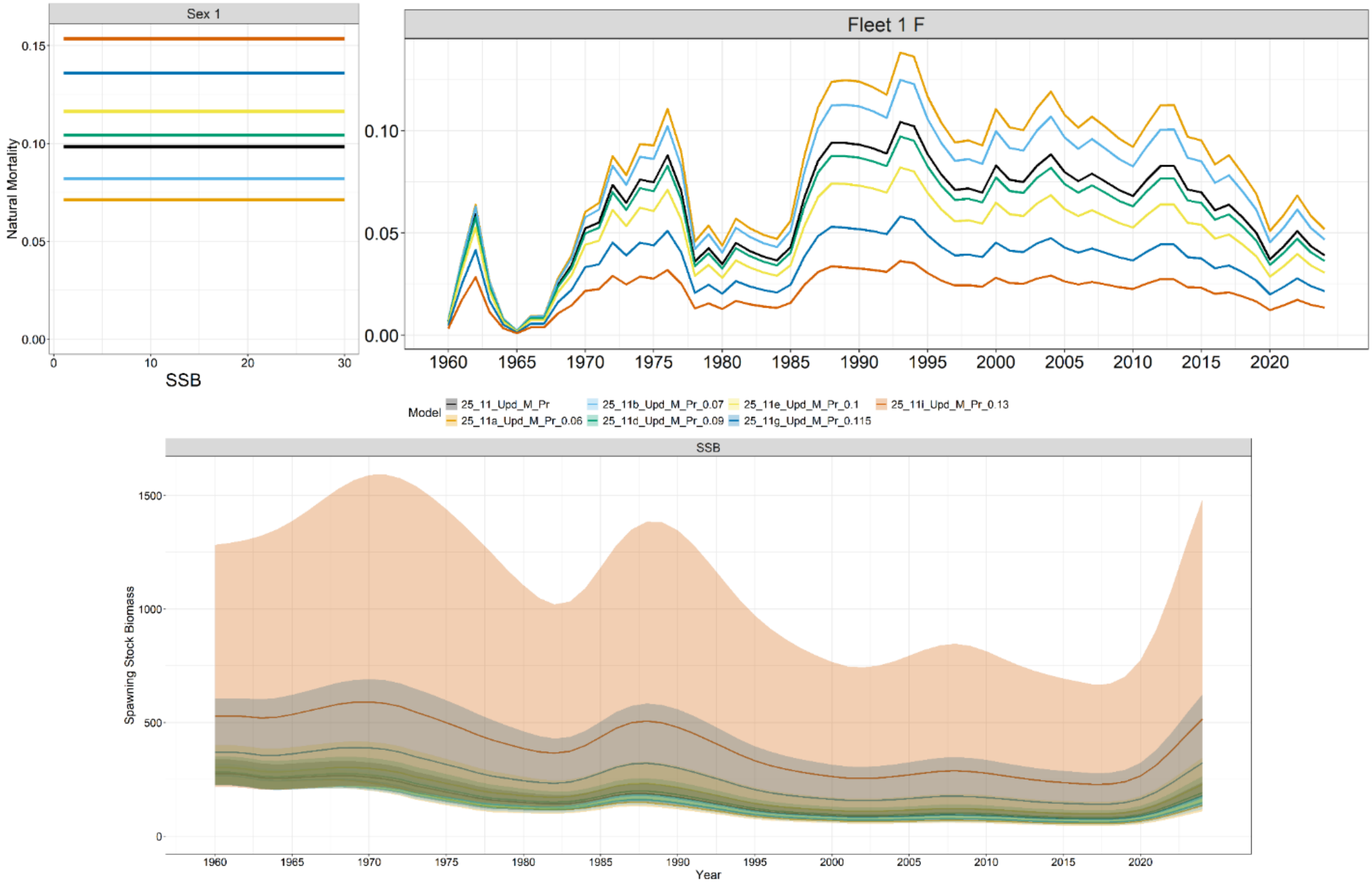


Figure 25. Comparison of estimates of female natural mortality (top left), fixed gear fishery fishing mortality (top right), and spawning stock biomass (kt; bottom) with associated 95% confidence intervals (shading) for exploratory runs using different natural mortality prior mean values based on the settings of model *25.11_Upd_M_Pr*.

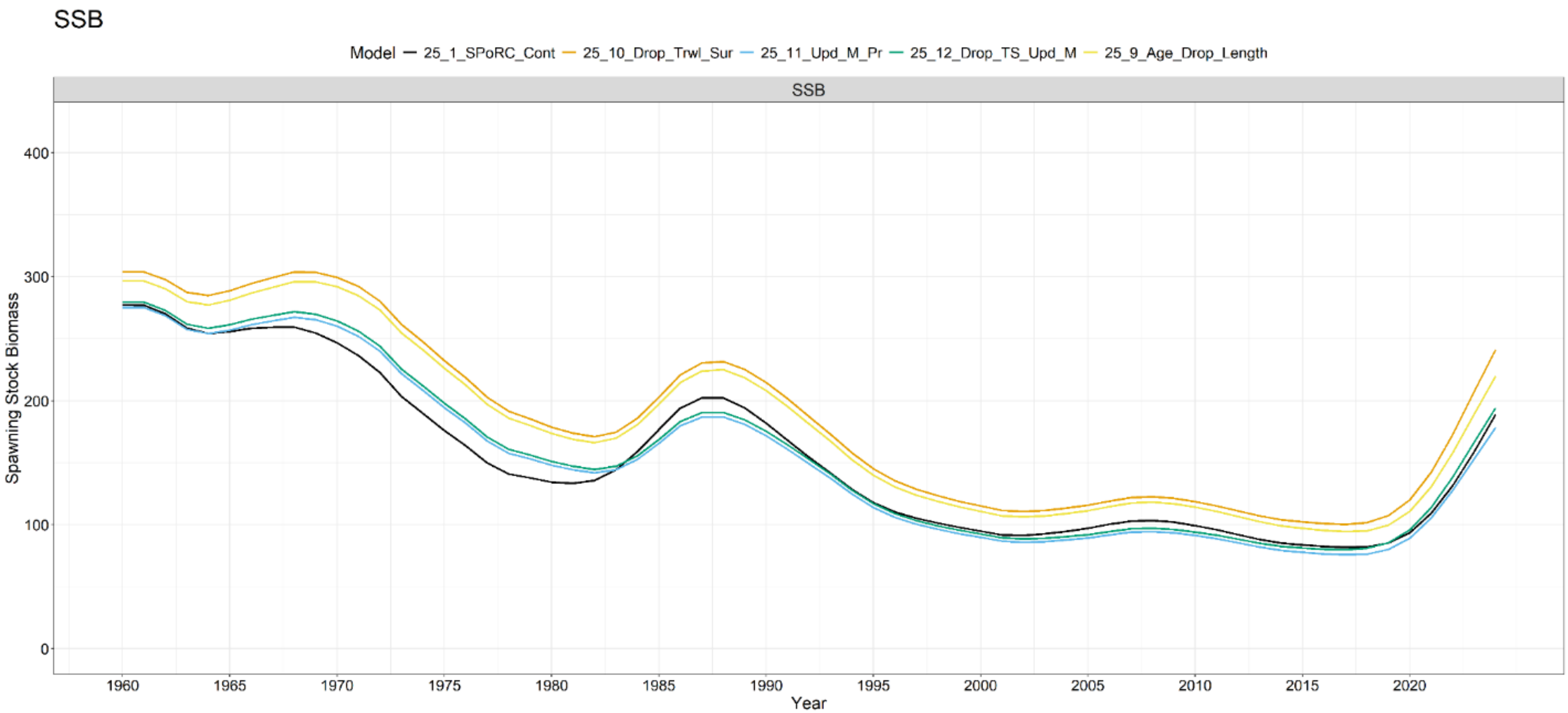


Figure 26. Comparison of spawning stock biomass (SSB, in kilotons) for each model run in the 'Data and Model Updates' grouping.

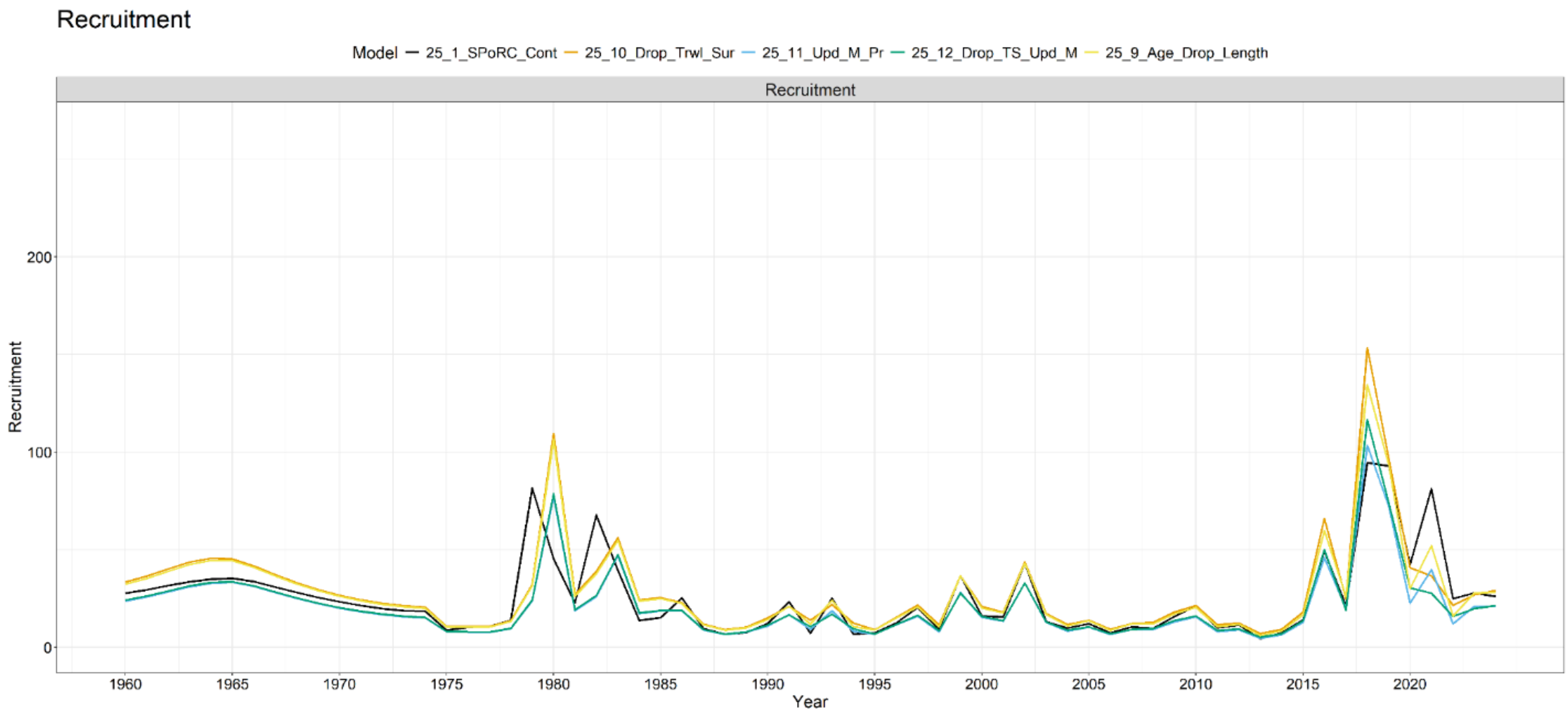


Figure 27. Comparison of recruitment (in millions of fish) for each model run in the 'Data and Model Updates' grouping.

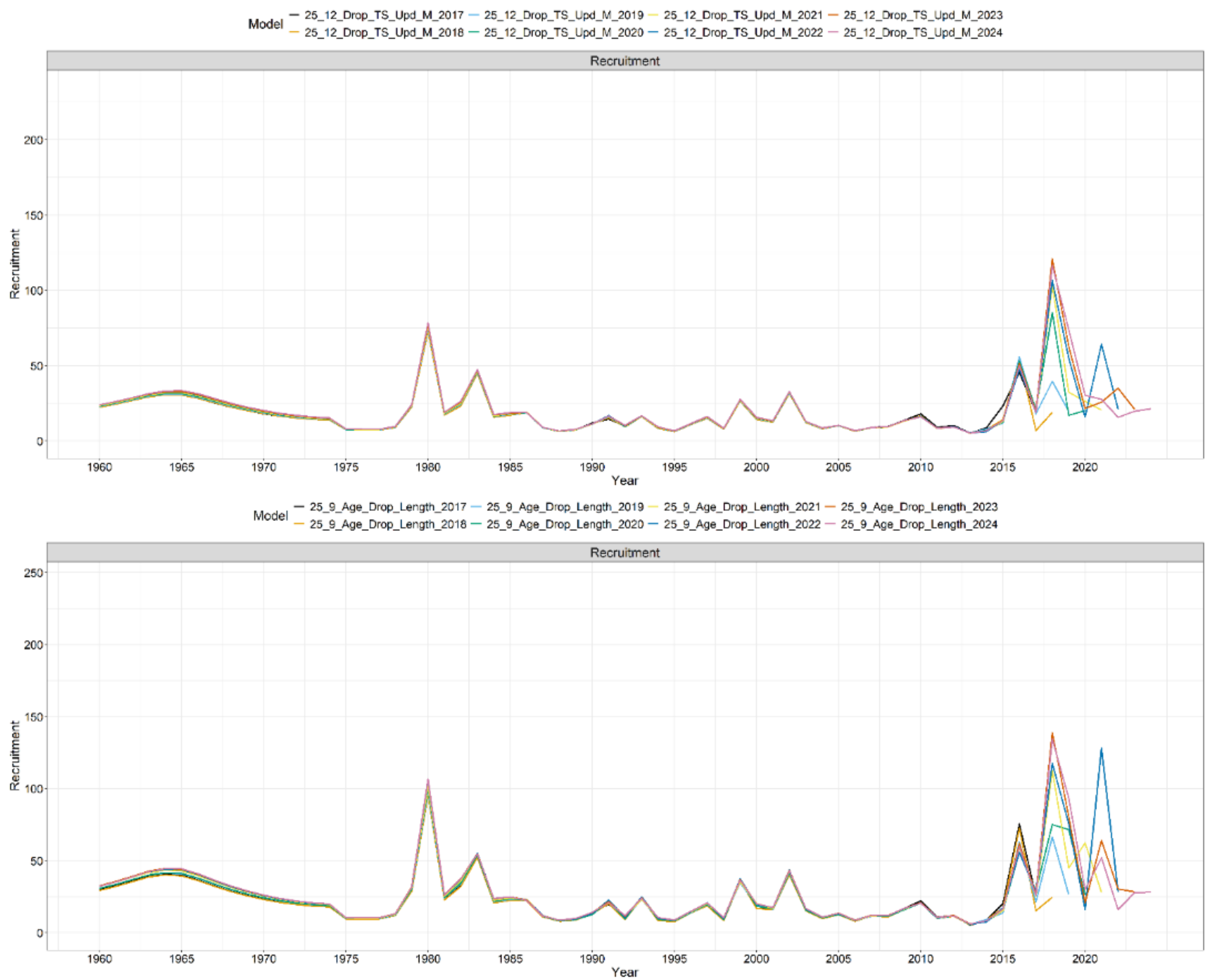


Figure 28. Comparison of the estimated recruitment (in millions of fish) time series from a seven year 'true' retrospective analysis for model *25.12_Drop_TS_Upd_M* (top panel) and *25.9_Age_Drop_Len* (bottom panel). The 'true' retrospective analysis removes a year of data for each peel, but also accounts for the lag in availability of compositional data (i.e., fishery ages and lengths as well as longline survey ages become available on a one year lag from associated catch or index data).

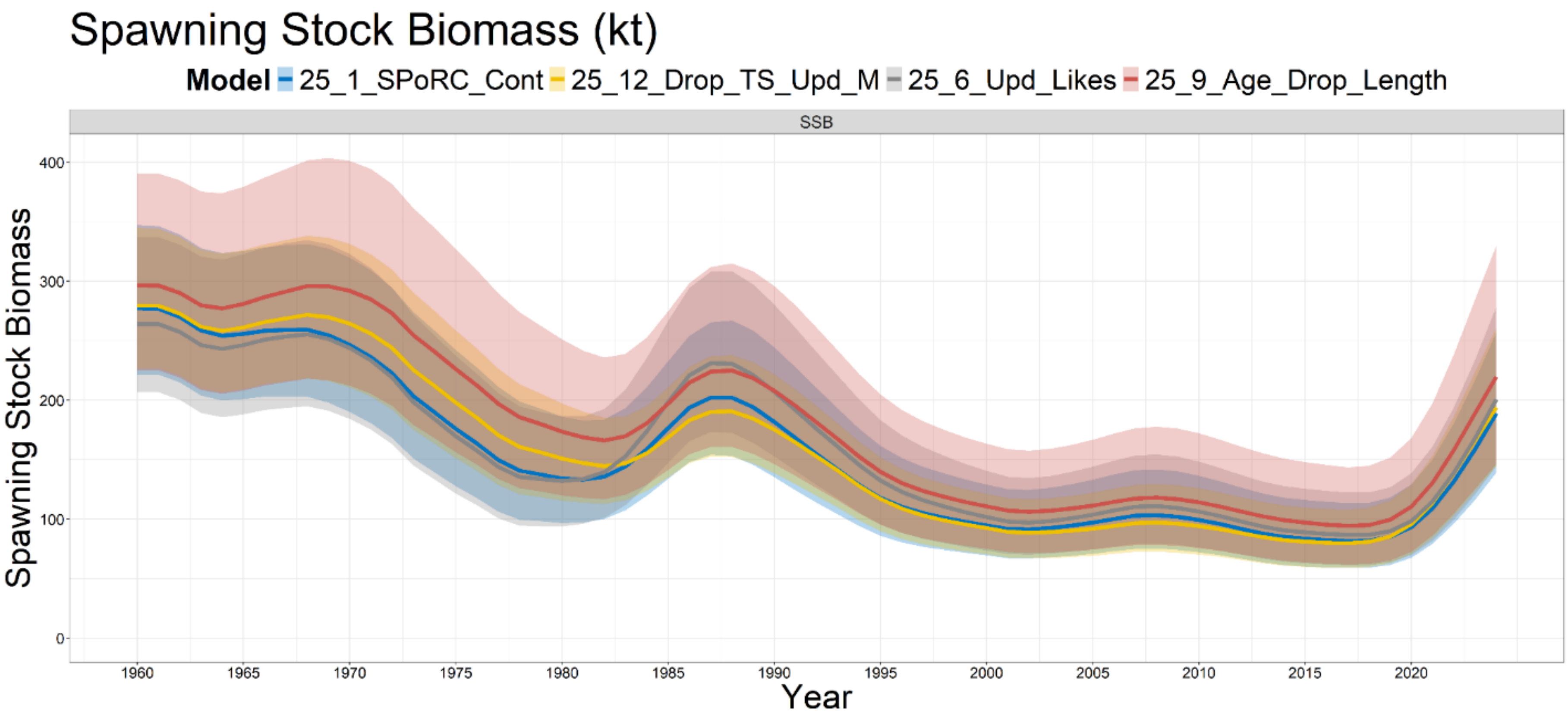


Figure 29. Comparison of spawning stock biomass (SSB, in kilotons) for the final model runs in each model grouping. The 95% confidence intervals for each model run are provided by the associated shading.

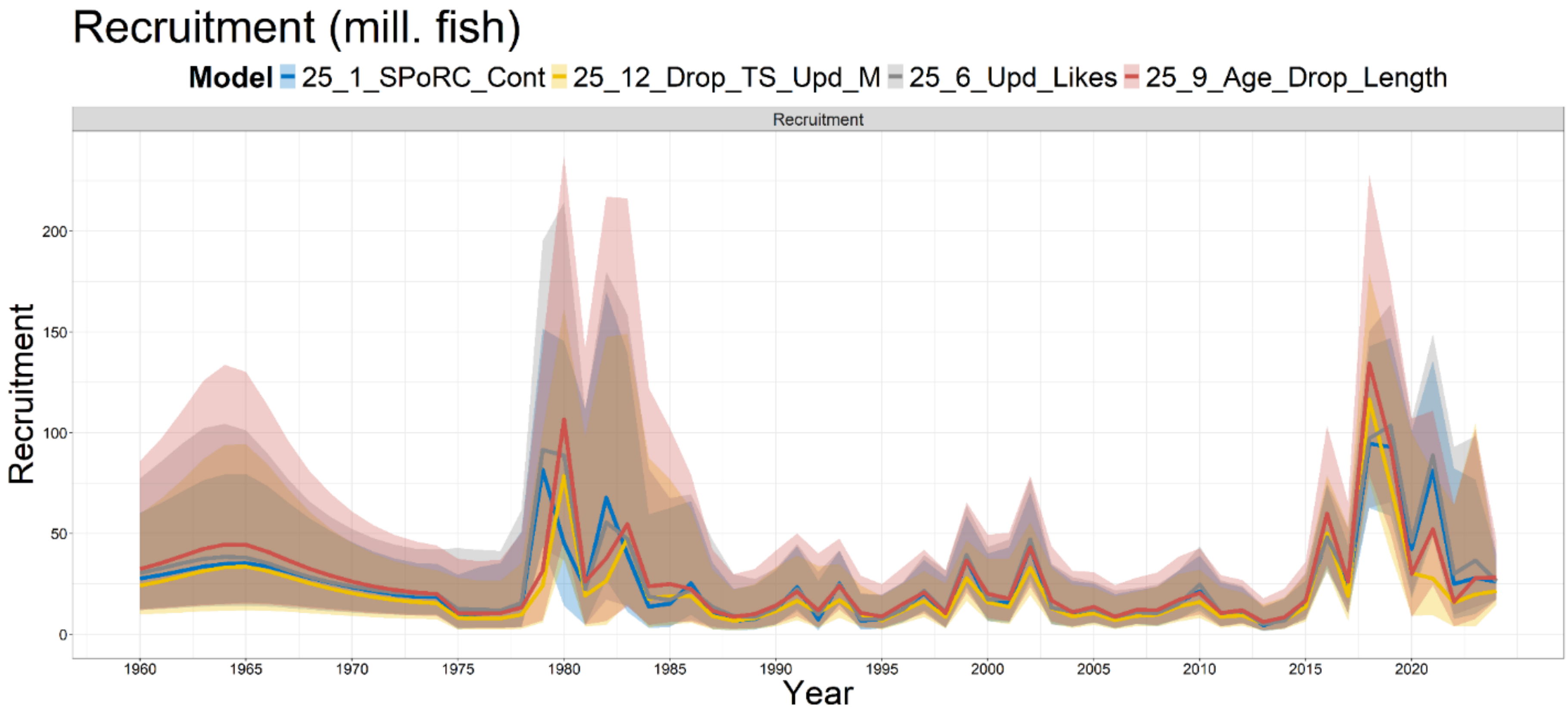


Figure 30. Comparison of recruitment (in millions of fish) for the final model runs in each model grouping. The 95% confidence intervals for each model run are provided by the associated shading.

5. Sensitivity Runs

Introduction

To address concerns raised by the NPFMC JPT and SSC, four sensitivity model runs were developed. The first sensitivity model (*25.13_TV_ALK*) explores whether accounting for density-dependent growth due to recent large year classes, through implementing time-varying growth (Cheng et al. 2024c), can reduce tension among compositional data (instead of assuming only two growth time blocks). Next, continuous time-varying selectivity for the fixed gear fishery was modeled (*25.14_TV_Selex*) to explore whether a more flexible selectivity approach can better capture changing fleet dynamics (e.g., the transition to pot gear) compared to the current time block method (Cheng et al., 2024a,b). While increased flexibility in selectivity parameterizations have the potential to improve model fits to composition data, appropriately weighting composition data may also achieve similar improvements. In general, the current Francis reweighting algorithm prioritizes fits to abundance indices by aligning mean observed and predicted ages or lengths, rather than ensuring internally consistent model weights. Thus, Francis reweighting has the potential to substantially downweight compositional data, reducing the associated information content for informing recruitment and natural mortality estimates. The third sensitivity model (*25.16_2dLN*) implemented a self-weighting logistic normal compositional likelihood for age compositions that allows for a two-dimensional autoregressive correlation structure, which was explored to see if sex-specific dynamics could be better estimated. The last sensitivity run (*25.15_Spatial*) provides an update, with data through 2024 and slight model refinements, to the spatially explicit sablefish assessment model developed by Cheng et al (2025c). Given changes in the spatial footprint of the longline survey starting in 2025, the regional structure of the spatial model may better handle data gaps in the index (i.e., only surveying a portion of the population in a given year) compared to the operational panmictic assessment. Taken together, these sensitivity runs provide a foundation for exploring future model alternatives.

Methods

The sensitivity model runs follow the parameterization of the final author recommended model (*25.12_Drop_TS_Upd_M*) with updates as noted below, and results are compared to the author recommended model. Francis reweighting was only utilized for model *25.13_TV_ALK*.

The configuration of sensitivity model runs was as follows:

- *25.13_TV_ALK*:
    - Time-varying growth and weight-at-age was incorporated into the model following the methods from Cheng et al. (2023, 2024c). Specifically, annual length-at-age and weight-at-age values were estimated external to the assessment model, then provided as fixed inputs.
- *25.14_TV_Selex*:
    - Fixed-gear fishery selectivity was parameterized to exhibit continuous time-varying logistic selectivity. Deviations were sex-and parameter-specific and were assumed to be independent and identically (iid) distributed random variables following a normal distribution. These deviations were estimated within a penalized likelihood framework, with the standard deviation fixed at 0.3 for selectivity parameters.

- *25.15_Spatial* (see the associated PT document for more details on this model):
    - Five regions were modeled, including the Bering Sea (BS), Aleutian Islands (AI), Western Gulf of Alaska (WGOA), Central Gulf of Alaska (CGOA), and the Eastern Gulf of Alaska (EGOA).
    - The fishery CPUE index was not fit.
    - Conventional mark-recapture data for sablefish in Alaska federal waters were directly fit through estimation of a tag reporting rate and an assumed tag loss and tag mortality rate.
    - Mean recruitment parameters were estimated for all five regions, along with region-specific recruitment deviations for each region.
    - All demographics were specified to be spatially-invariant, with natural mortality fixed at values estimated from *25.12_Drop_TS_Upd_M*.
    - Movement was estimated in three distinct age blocks: young individuals (ages 2-7), medium aged-individuals (ages 8 – 16), and old individuals (17 – 31).
    - The two fishery fleets were specified across all five regions and selectivity for each was assumed to be spatially invariant, but where the fixed-gear selectivity maintained three time blocks.
    - Tag reporting rates were estimated in time blocks, which aligned with those used for the fixed-gear selectivity.
    - Input sample sizes were specified at 60, 40, 20, and 20 for survey age compositions, fishery age compositions, fishery length compositions, and survey length compositions, respectively.

- *25.16_2dLN*:
    - For age-composition data, a logistic normal distribution with a two-dimensional autoregressive correlation structure was assumed. The correlation structure was formulated as the Kronecker product of a constant correlation matrix for sexes and a first-order autoregressive correlation matrix by ages (Figure 31). Because the logistic normal likelihood is unable to accommodate zeros, these values were not fit (similar approach as used in other common assessment packages).
    - Length composition data were fit assuming a multinomial distribution with an input sample size of 5.

Results

Model *25.13_TV_ALK* estimated a slightly higher population scale largely driven by a higher estimate of natural mortality, while recent SSB trends differ slightly from model *25.12_Drop_TS_Upd_M* due to decreasing time-varying weight-at-age (Figure 32). Trawl fishery selectivity increases for older ages, which required lower fishing mortality to adequately fit catch data (Figure 33). In general, model fits to the fixed-gear CPUE index and trawl fishery length compositions appeared to demonstrate slight improvements.

Estimating continuous time-varying selectivity for the fixed-gear fleet resulted in reduced population scale, while retaining similar population trends (Figure 34). Allowing greater flexibility in fixed-gear selectivity led to higher estimates of fishing mortality, which in turn required lower natural mortality estimates and a subsequent downward revision in population scale. Overall, estimates of fixed-gear selectivity appeared to be sufficiently constrained and relatively reasonable (Figure 35). Estimates of selectivity tended to shift towards selecting younger ages in recent years, a pattern also reflected in the current time-block approach. Despite the added flexibility, model fits to the fixed-gear CPUE index and associated age compositions did not demonstrate substantive improvements.

Moderate changes in internal model dynamics were detected when self-weighting likelihoods were implemented for the age composition data (*25.16_2dLN*). While trends in SSB remained similar, noticeable increases were evident in population scale (Figure 36), selectivity patterns (Figure 37), and recent recruitment trends (Figure 38). Notably, natural mortality was estimated to be much lower (~0.08), aligning more closely with known sablefish biology. Moreover, selectivity values were often more selective, requiring reductions in subsequent fishing mortality rates. Thus, it is expected that the combination of lower natural and fishing mortality contributed to an increase in population scale. With respect to recruitment trends, the *25.16_2dLN* model estimates a relatively large 2018 cohort, which was only slightly above average in the final author recommended model. Many of these changes likely emerged because model weights were more internally consistent, and that respective data sources were more appropriately weighted. For instance, there appears to be less tension among data sources, where fits to the both the LLS index in recent years and age-composition data are improved, while the model is also better able to fit the age composition plus group (Figure 39 - 41).

Application of the spatially-explicit, tag-integrated assessment model for Alaska sablefish (*25.15_Spatial*) revealed distinct regional population dynamics. Age-specific movement rates revealed a general counter-clockwise movement pattern, wherein young individuals remained in the BS and AI and moved eastwards as they matured (Figure 42). Moreover, residency in the AI increased with age, likely driven by an exceptionally large plus group that is periodically observed in this region (Figure 43). SSB was estimated to be lower in the BS, AI, and WGOA, and higher in the CGOA and EGOA (Figure 44). Conversely, age-2 recruitment was generally estimated to be highest in western regions (BS and AI; Figure 45). The spatial incoherence between regions of high SSB and those of high recruitment potentially results from the combination of ontogenetic movement patterns and larval dispersal away from the GOA. In comparison to the single region author recommended model (*25.12_Drop_TS_Upd_M*), region-wide SSB, stock status, and recruitment were relatively consistent (Figure 46 and 47). Differences between the single region and spatial models primarily emerged in early periods, where the spatial model lacked a spatially resolved CPUE index to anchor initial SSB estimates. Some differences in recent year class strength were also observed. For example, the 2017 cohort was estimated to be higher in the spatial model, likely because aggregation of compositions reduces regional signals on year class strength.

Notable changes in model dynamics included:

- *25.13_TV_ALK* (full results here):
    - Incorporating a time-varying size-age transition matrix improved fits to length compositions.
    - The trawl fishery was estimated to be more selective of older, larger sablefish.
    - Changes in the trawl fishery selectivity led to subsequent decrease in fishing mortality and higher natural mortality, resulting in a slight increase in population scale.

    - SSB and recruitment trends remained relatively similar to the final author recommended model, but terminal SSB is lower due to decreasing weight-at-age.
- *25.14_TV_Selex* (full results here):
    - Estimating continuous time-varying selectivity for the fixed-gear fleet led to slight improvements to fits of the fixed-gear CPUE and length compositions, but model fits to age-compositions were unchanged.
    - A higher estimate of fishing mortality resulted in lower values for natural mortality, and a small decrease in population scale.
- *25.15_Spatial* (full results here):
    - Movement estimates generally illustrated a counter-clockwise ontogenetic pattern. Young individuals tended to remain in the BS and AI and moved eastward as they matured.
    - Regional SSB was generally lower in the BS, AI, and WGOA, and was higher in the CGOA and EGOA.
    - Large age-2 recruitment events occurred mainly in western regions, and likely reflect the combined effects of movement and larval dispersal.
    - In general, single region and spatial model results were consistent. However, the magnitude of recruitment events differed slightly in recent years (i.e., the 2017 cohort), likely because utilizing spatially-aggregated composition data smears across the regional signals in year class strength.
- *25.16_2dLN* (full results here):
    - Modest improvements in fits to the LLS index and compositional data.
    - Lower estimates of natural mortality (~0.08) were more consistent with the biological dynamics of the species.
    - Estimates of population scale were moderately higher than the final author recommended model, due to the combination of lower natural and fishing mortality.
    - Many changes likely arose from improved internal consistency of the model, which stemmed from more appropriate model data weights.

The sensitivity runs provide a foundation for exploring alternative parameterizations in future assessments. Given the added complexity of many of these models, the authors felt that further investigations were warranted to better understand performance and stability before being recommended as the basis of operational management advice. For instance, more flexible time-varying selectivity alone did not substantially improve model performance, but evaluating its use in tandem with self-weighted likelihoods could be beneficial. Likewise, while time-varying growth had limited impact when considered in isolation, estimating these dynamics internally could help reduce tension among length and age data. Finally, given recent changes in the spatial coverage of the LLS, further consideration of a spatial stock assessment model may be warranted.

Ongoing Explorations for Future Sablefish Assessments

- Improving the characterization of sex-specific dynamics through estimation of sex-specific natural mortality and recruitment sex ratio.
- Implement time-varying selectivity in tandem with self-weighting compositional likelihoods.

- Evaluating the use of a spatial stock assessment model to address changes in LLS design (e.g., a three region model for the BS, AI, and GOA).
- Explore updates to the aging error matrix.
- Investigate treating recruitment deviations as random effects and directly estimate the recruitment variance term to better utilize the capabilities of TMB.

Figures

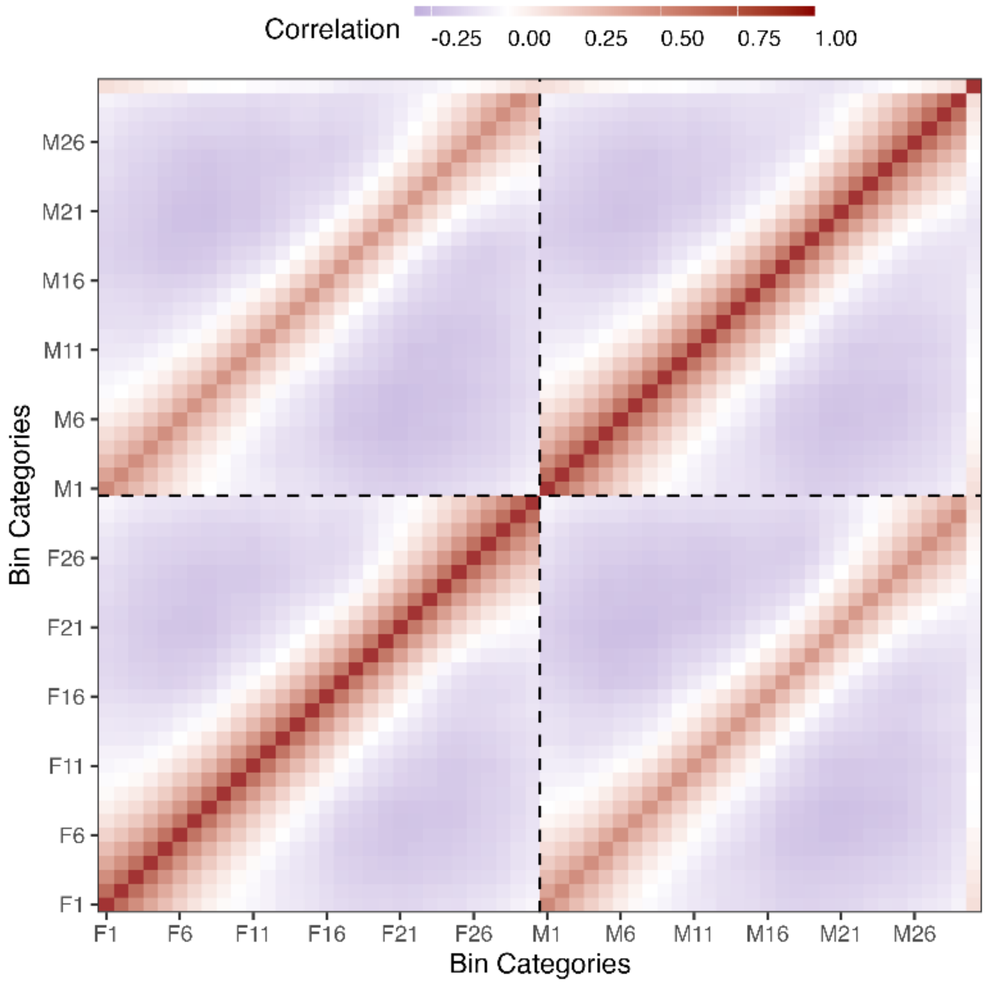


Figure 31. Expected residual correlations arising from a two-dimensional autoregressive process under the logistic normal likelihood assumed for model *25.16_2dLN***.** Dotted lines separate sexes, with F denoting females and M denoting males.

Figure 32. Comparison of spawning stock biomass (SSB, in kilotons) between models *25.12_Drop_TS_Upd_M* and *25.13_TV_ALK*.

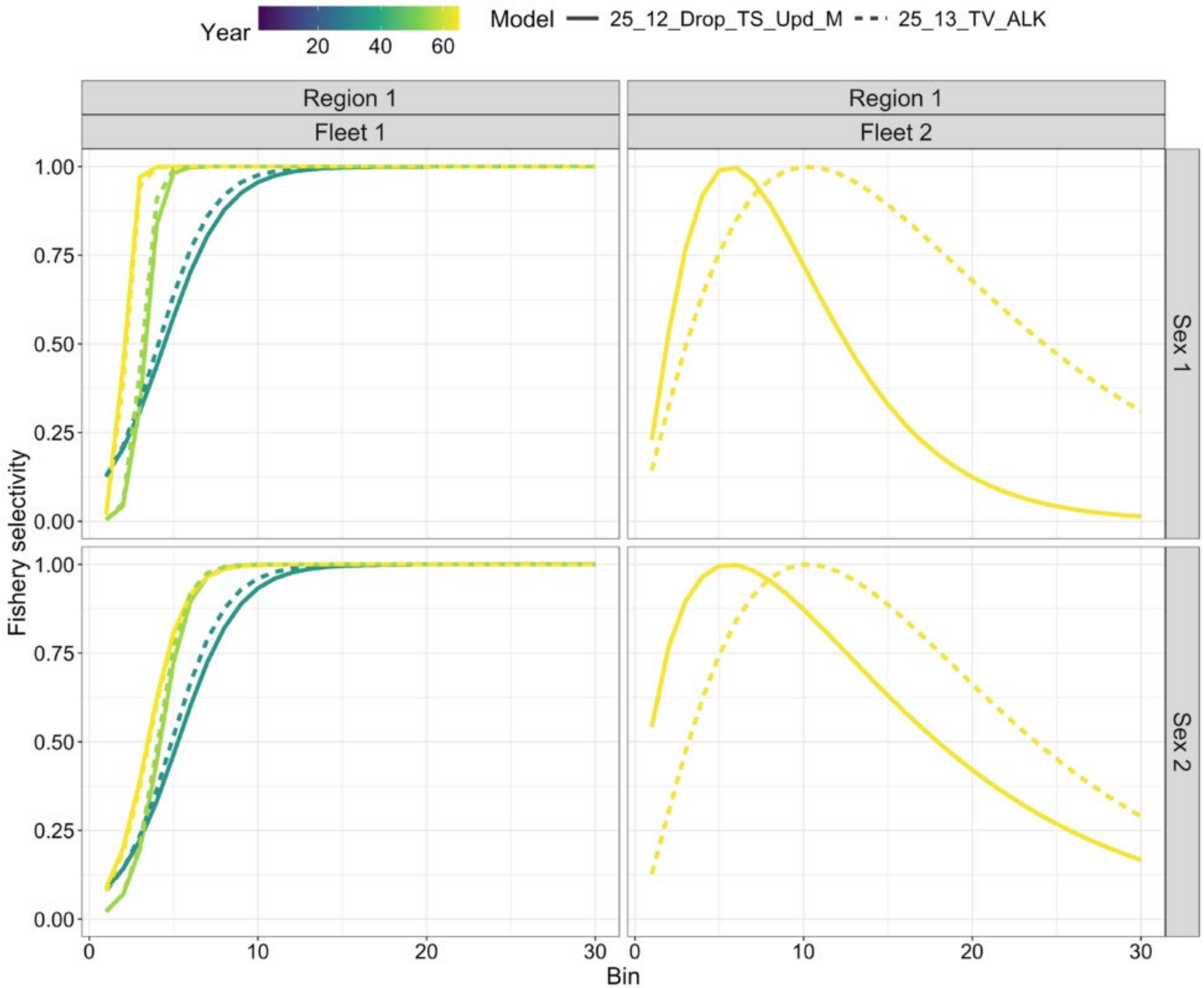


Figure 33. Comparison of selectivity estimates between models *25.12_Drop_TS_Upd_M* and *25.13_TV_ALK*. Fleet 1 indicates the fixed-gear fleet, while Fleet 2 indicates the trawl fishery. Sex 1 and Sex 2 denote females and males, respectively. Separate colored lines are depicted for the three fixed-gear fishery time blocks. Note that Year 1 corresponds to 1960 and Year 65 corresponds to 2024.

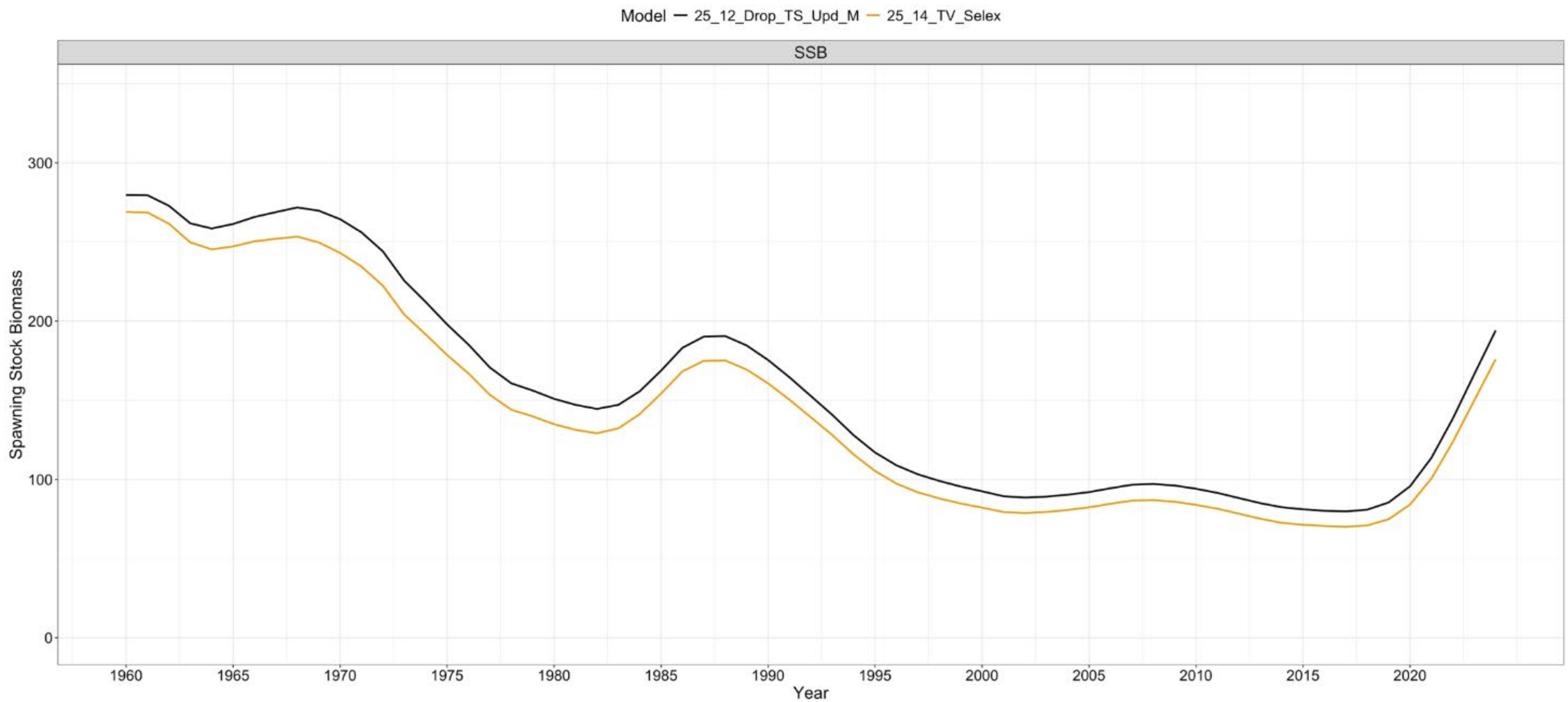


Figure 34. Comparison of spawning stock biomass (SSB, in kilotons) between models *25.12_Drop_TS_Upd_M* and *25.14_TV_Selex*.

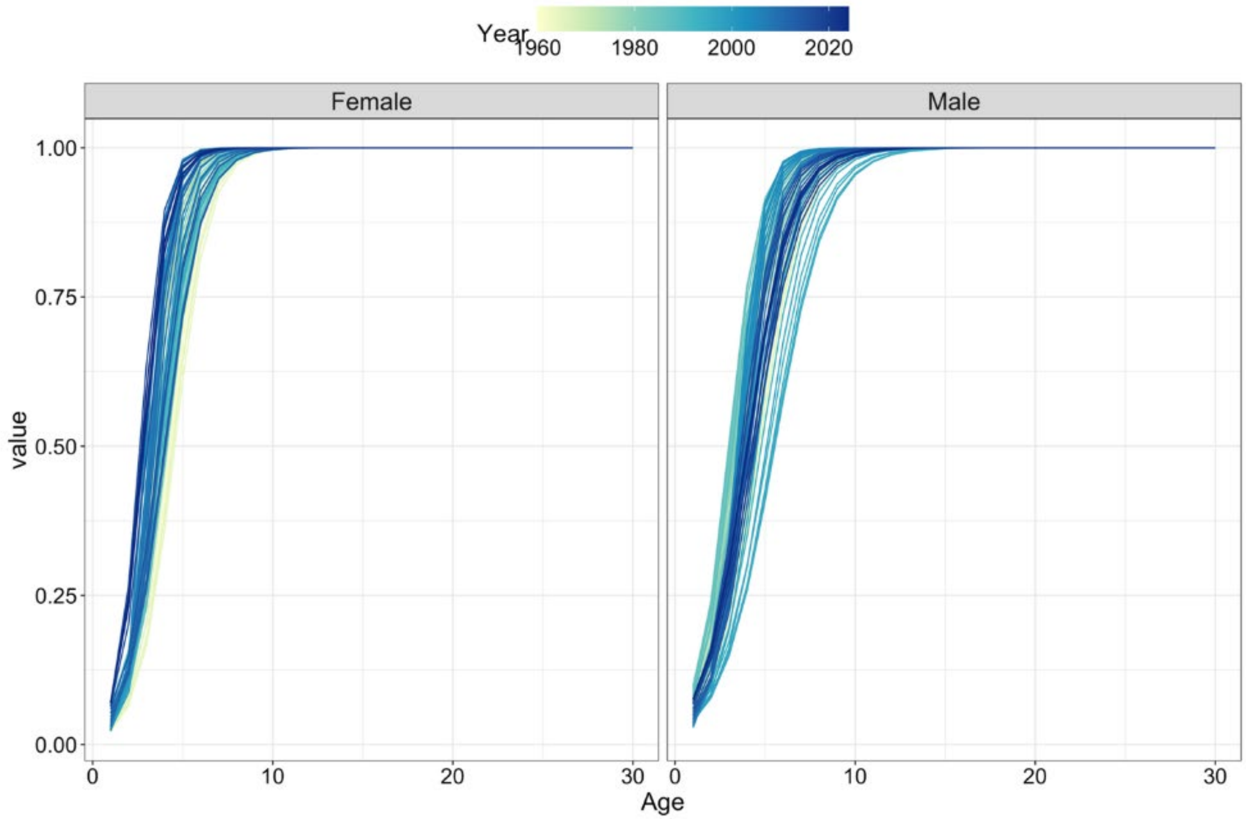


Figure 35. Estimates of continuous time-varying selectivity from model *25.14_TV_Selex*.

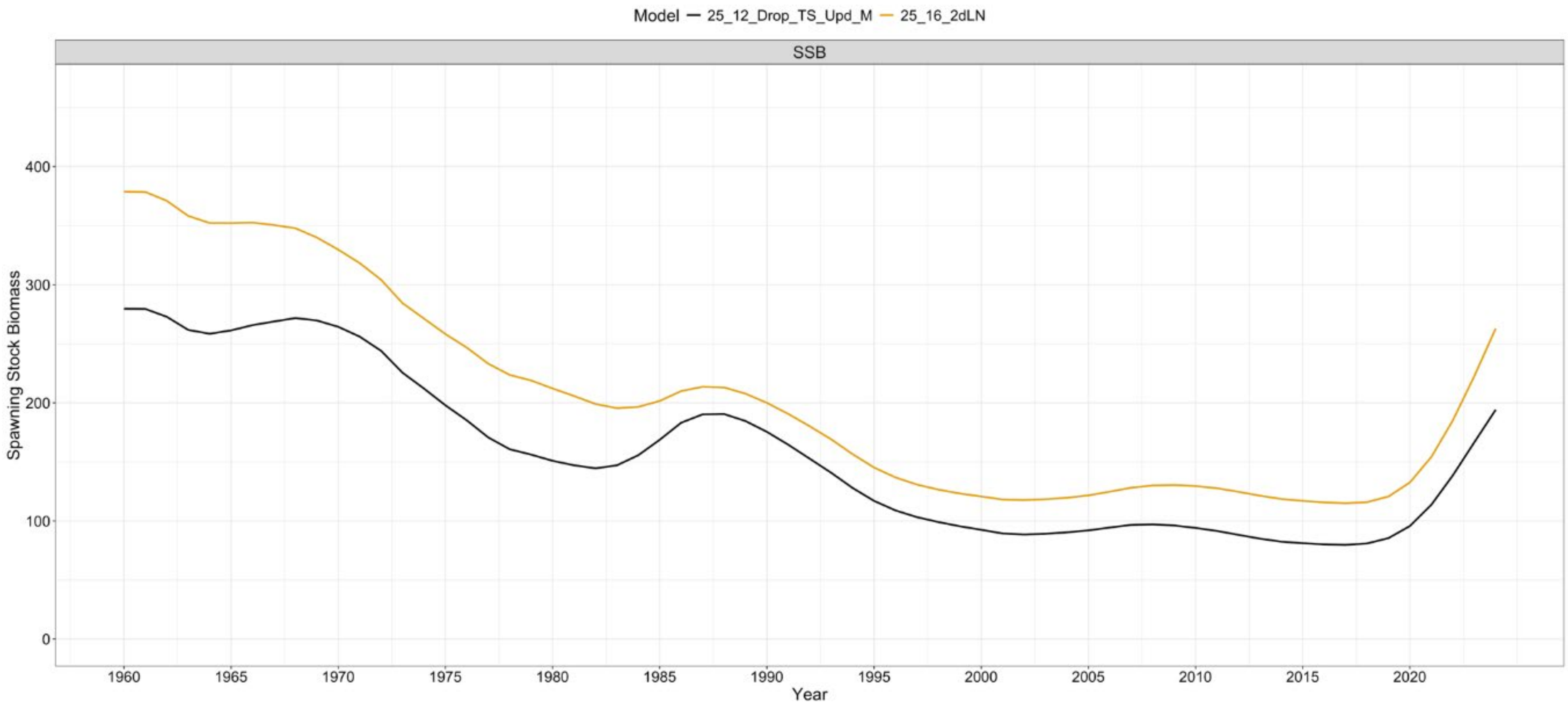


Figure 36. Comparison of spawning stock biomass (SSB, in kilotons) between models *25.12_Drop_TS_Upd_M* and *25.16_2dLN*.

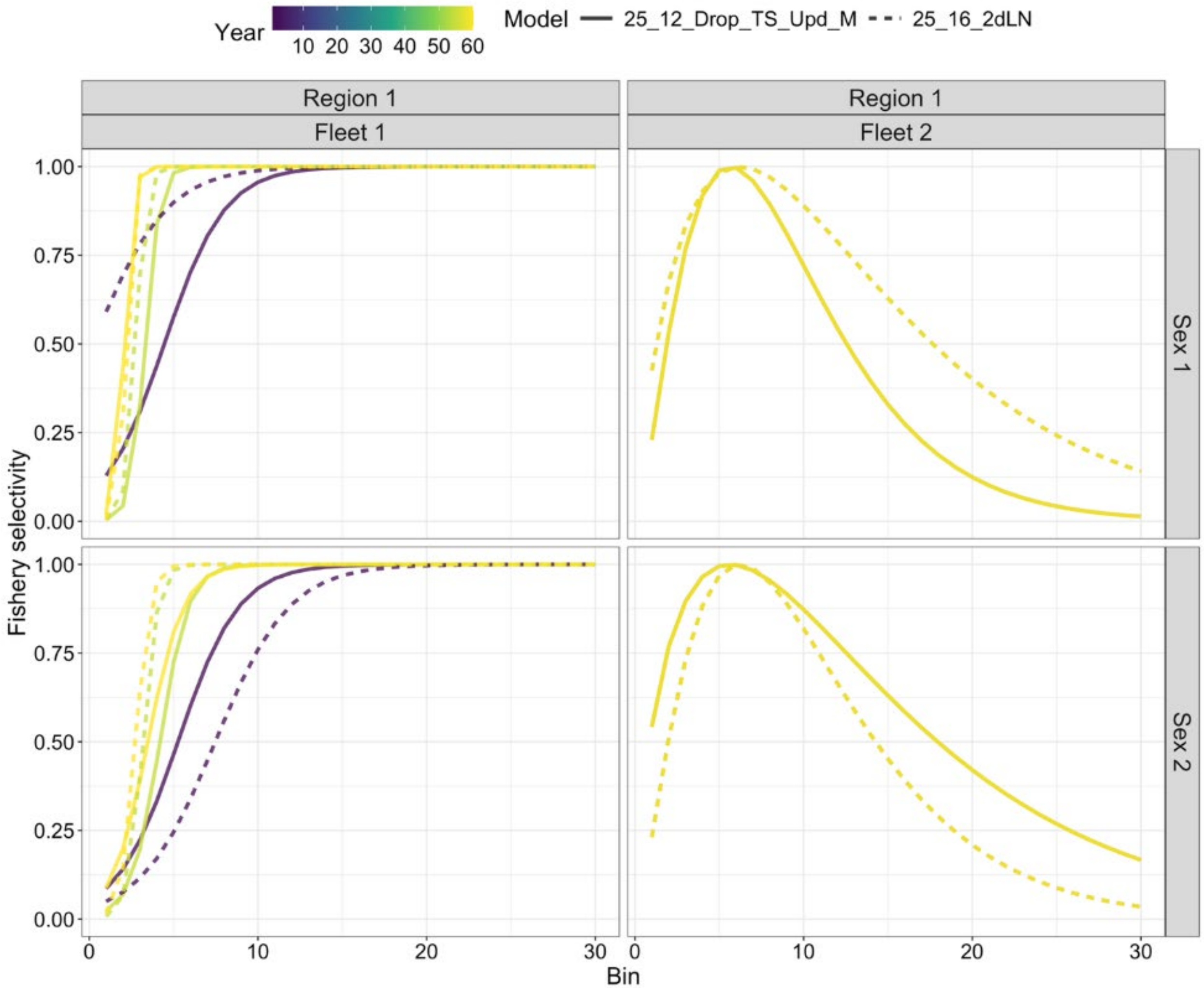


Figure 37. Comparison of selectivity estimates between models *25.12_Drop_TS_Upd_M* and *25.16_2dLN*. Fleet 1 indicates the fixed-gear fleet, while Fleet 2 indicates the trawl fishery. Sex 1 and Sex 2 denote females and males, respectively. Separate colored lines are depicted for the three fixed-gear fishery time blocks. Note that Year 1 corresponds to 1960 and Year 65 corresponds to 2024.

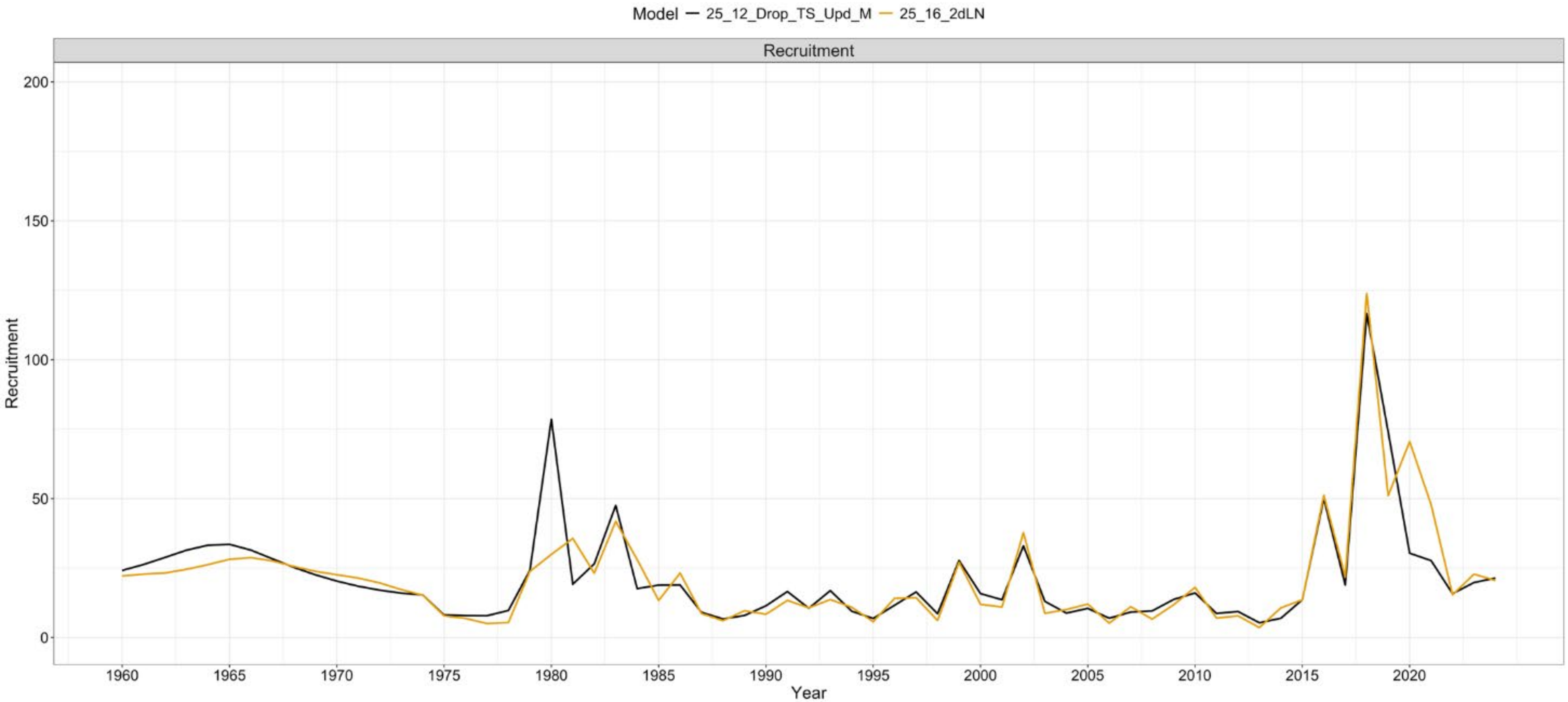


Figure 38. Comparison of age-2 recruitment (millions of fish) between models *25.12_Drop_TS_Upd_M* and *25.16_2dLN*.

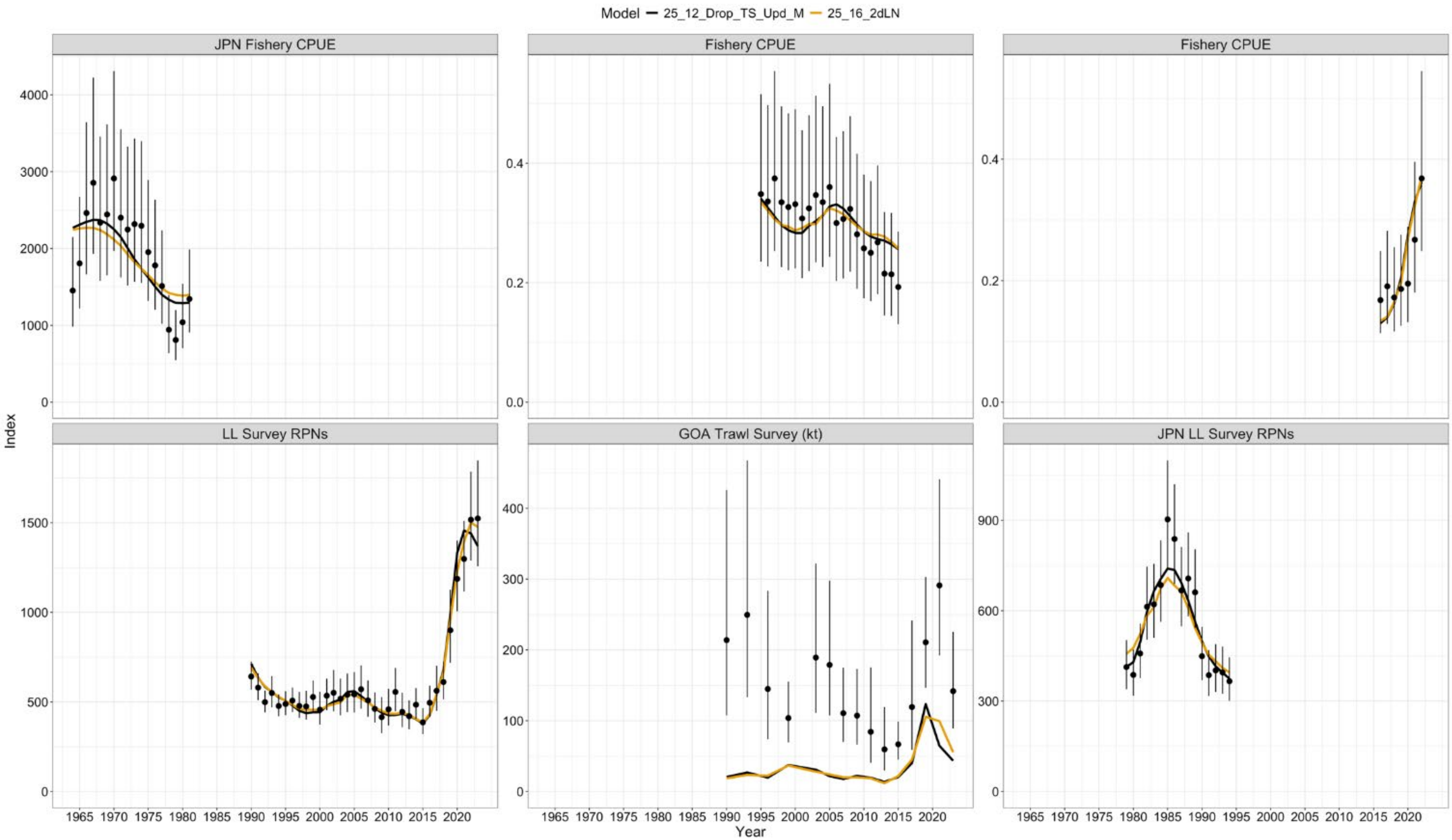


Figure 39. Fit to the fishery (top row) and survey (bottom panel) indices for models *25.12_Drop_TS_Upd_M* and *25.16_2dLN*. Note that the trawl survey is not fit in either model.

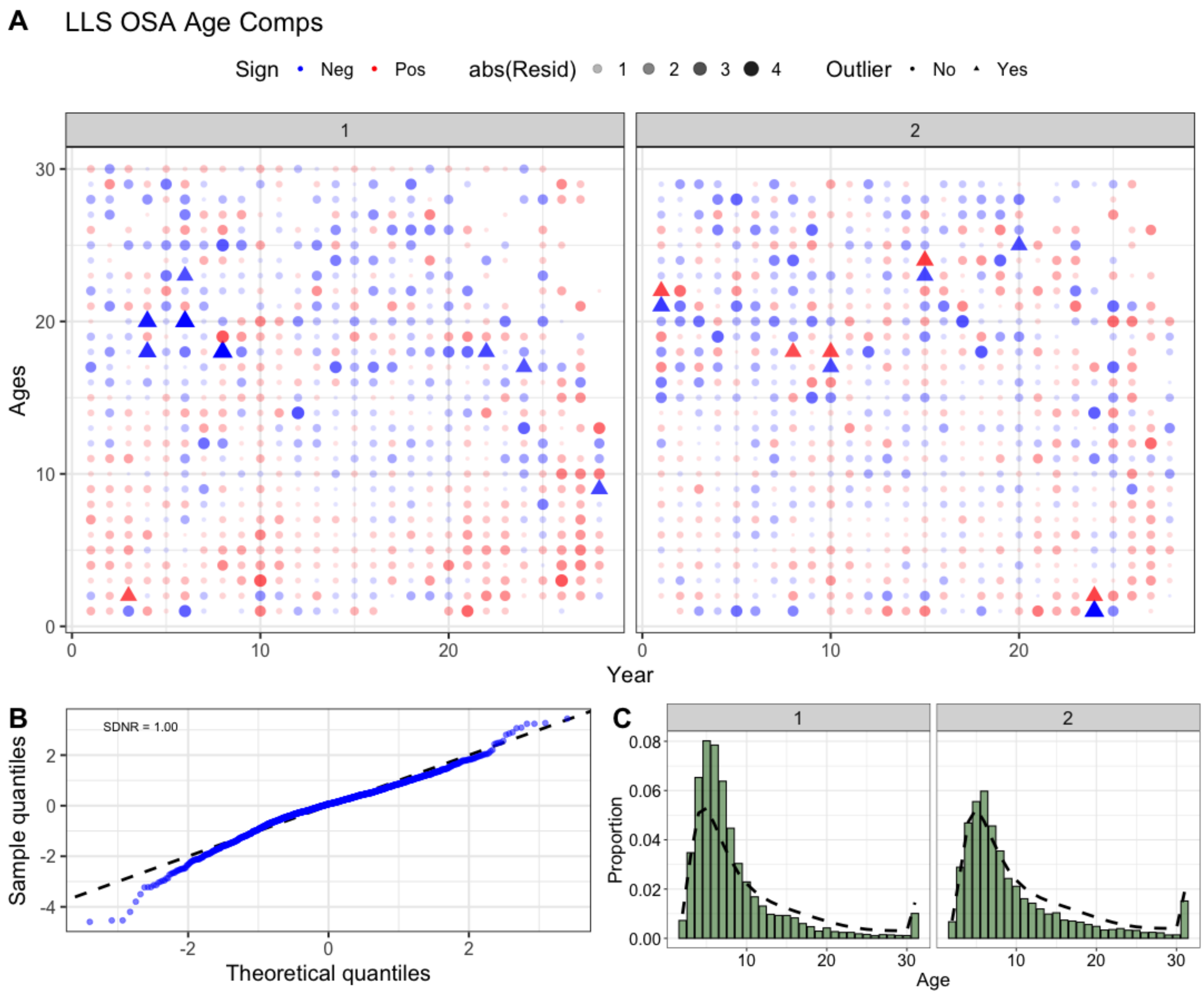


Figure 40. Fits to NOAA longline survey age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions) for model *25.16_2dLN.*

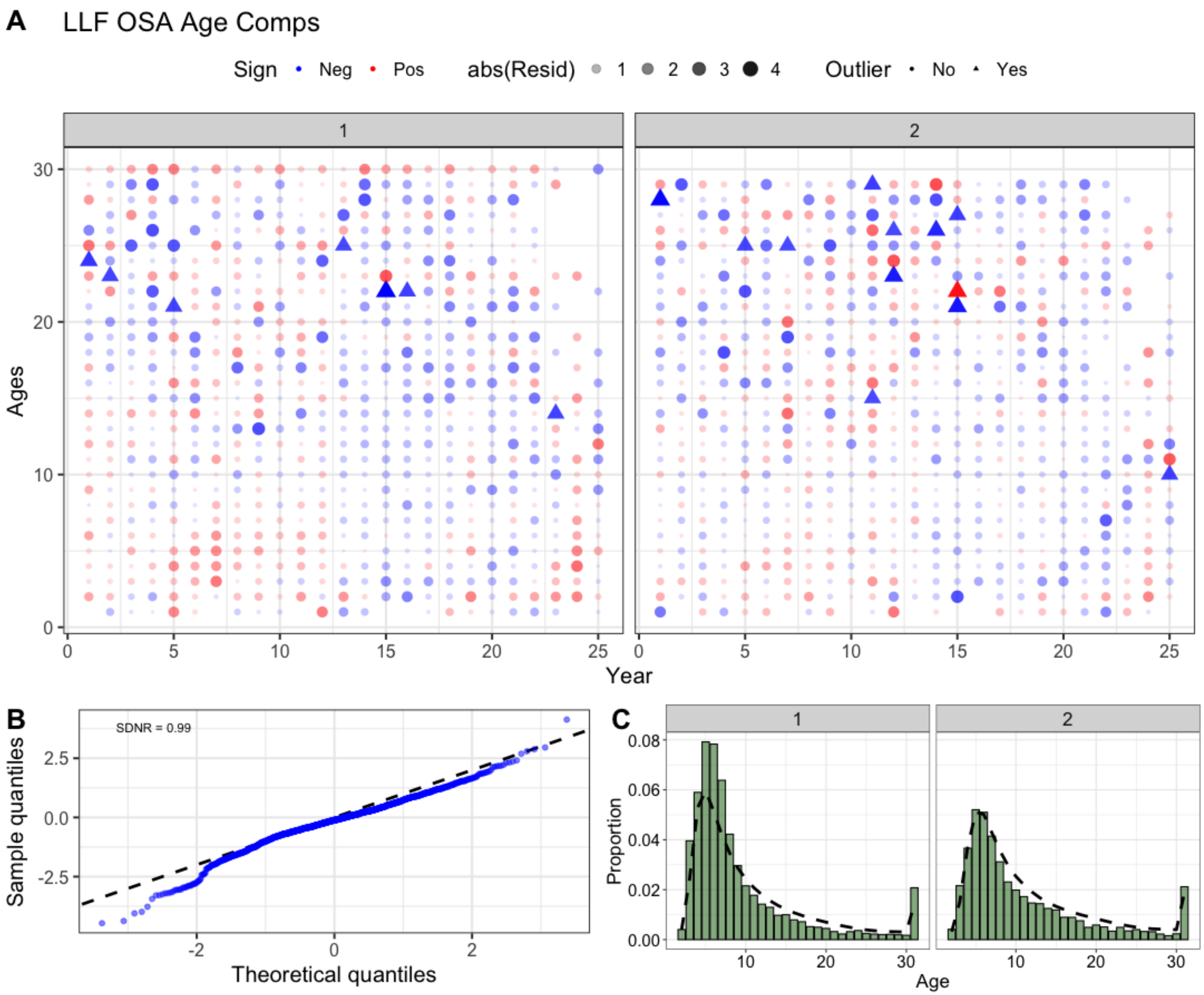


Figure 41. Fits to fixed gear fishery age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions) for model *25.16_2dLN*.

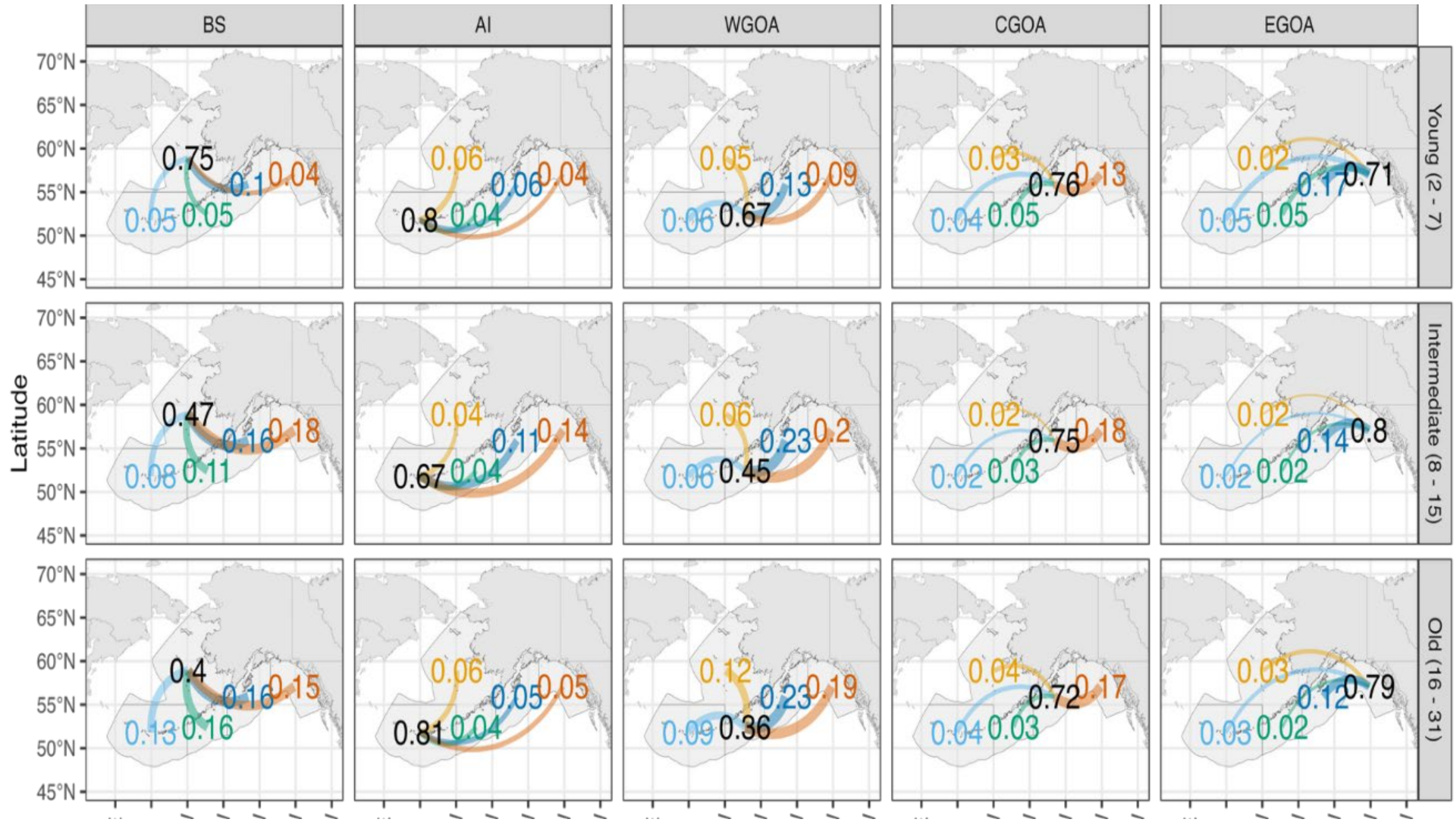


Figure 42. Estimated movement probabilities from the spatial assessment model (*25.15_Spatial*) by region and age groups. Row panels indicate the different age groups (young, intermediate, and old) and column panels represent movement originating from a given spatial region [Bering Sea (BS), Aleutian Islands (AI), Western GOA (WGOA), Central GOA (CGOA), Eastern GOA (EGOA)]. Values shown in black are residence probabilities (i.e., the probability of remaining within the same region), while colored values represent the probability of moving into a given spatial region. Line thickness is proportional to movement probabilities, with thicker lines corresponding to higher movement probabilities.

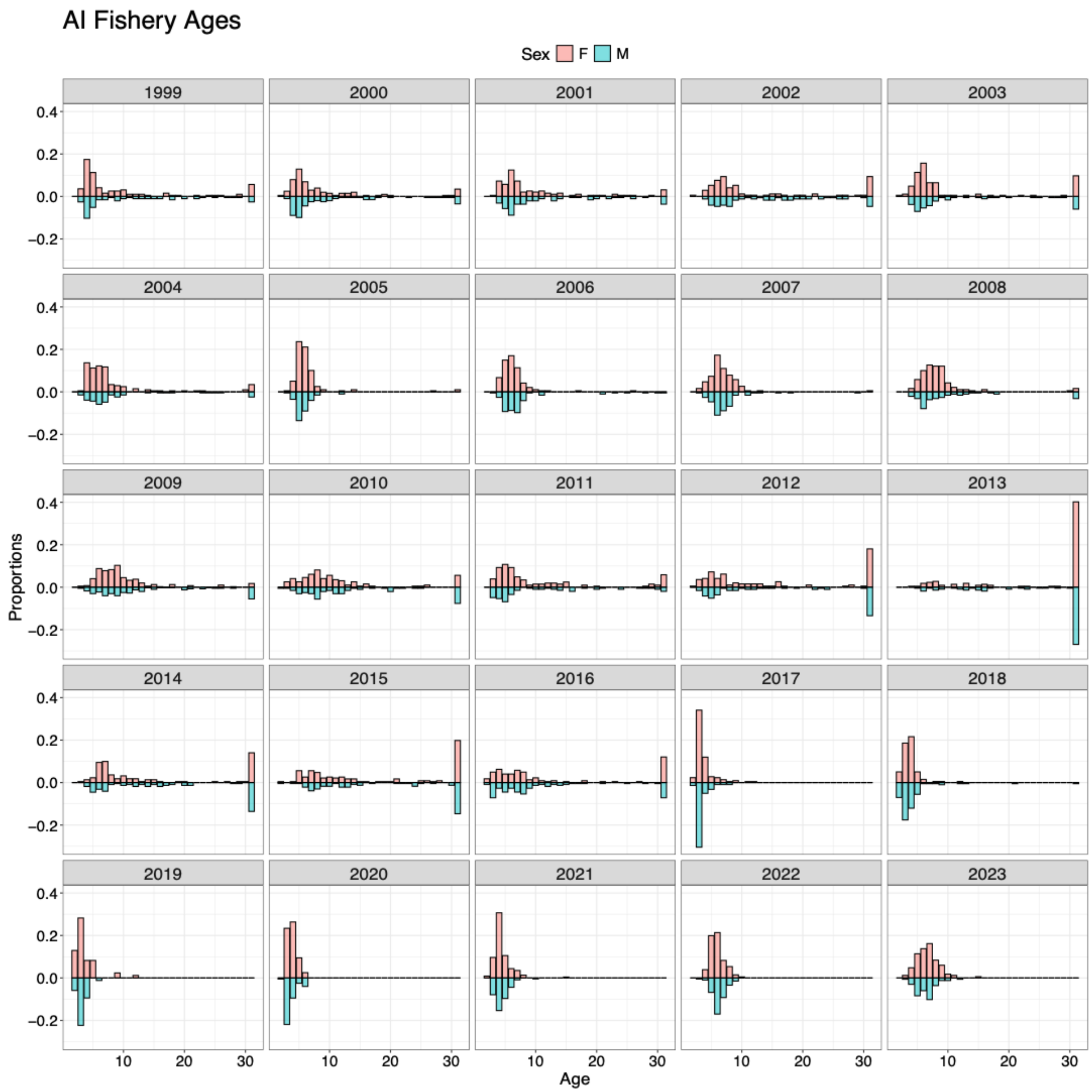


Figure 43. Observed fixed-gear fishery age compositions from the Aleutian Islands. Note that large, old individuals (age 31+) compose a significant portion of the composition data from 2012 - 2016.

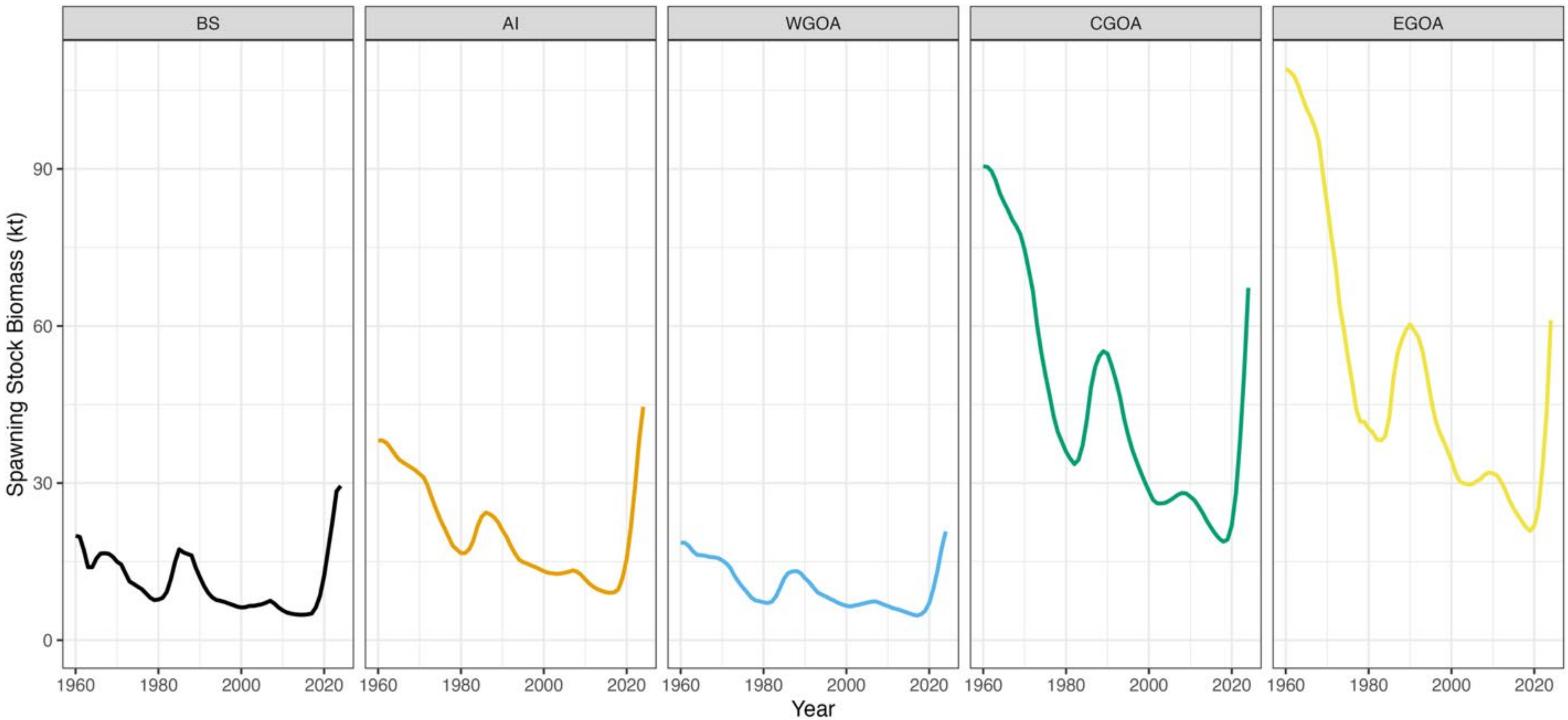


Figure 44. Spawning stock biomass (kt) estimates from model *25.15_Spatial* by region [Bering Sea (BS), Aleutian Islands (AI), Western GOA (WGOA), Central GOA (CGOA), Eastern GOA (EGOA)].

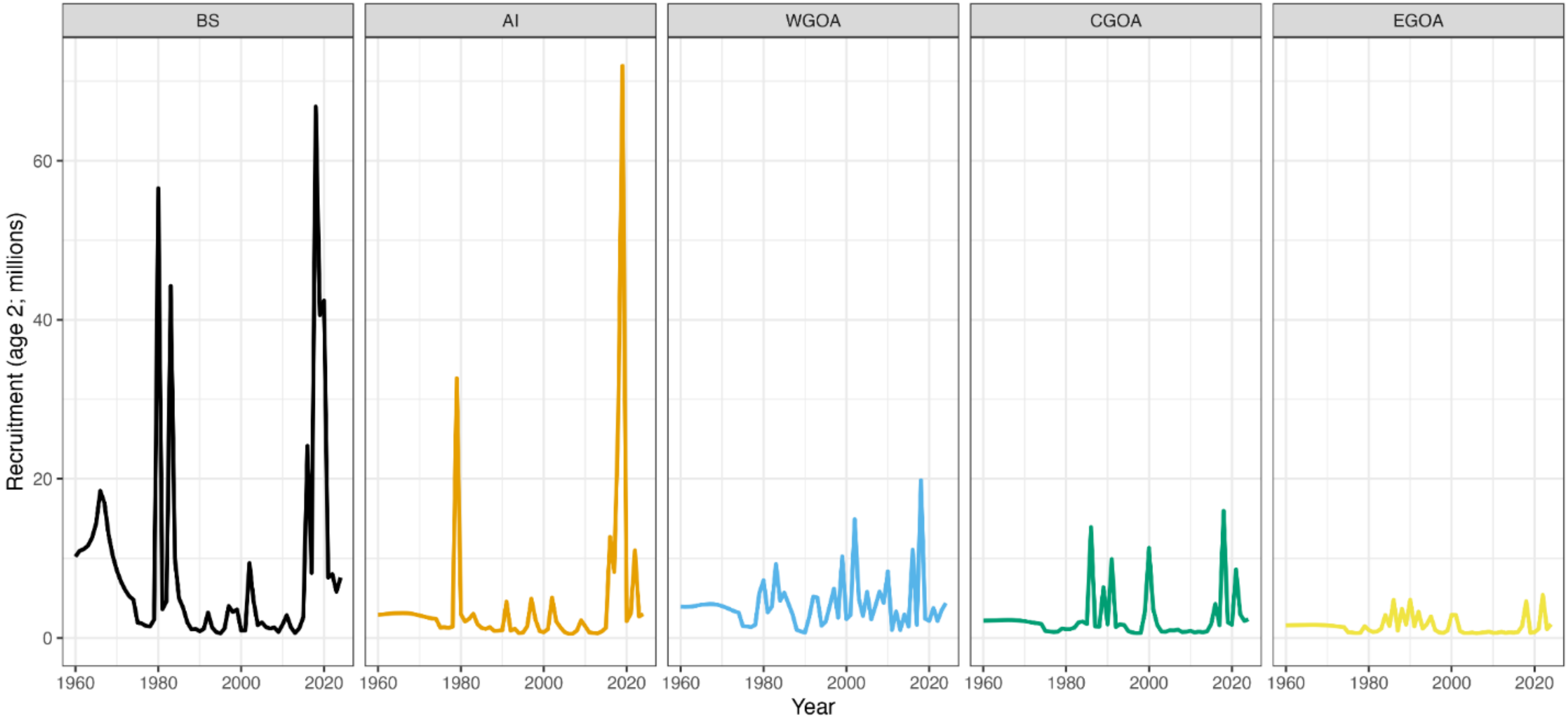


Figure 45: Age 2 recruitment estimates (millions of fish) from model *25.15_Spatial* by region [Bering Sea (BS), Aleutian Islands (AI), Western GOA (WGOA), Central GOA (CGOA), Eastern GOA (EGOA)].

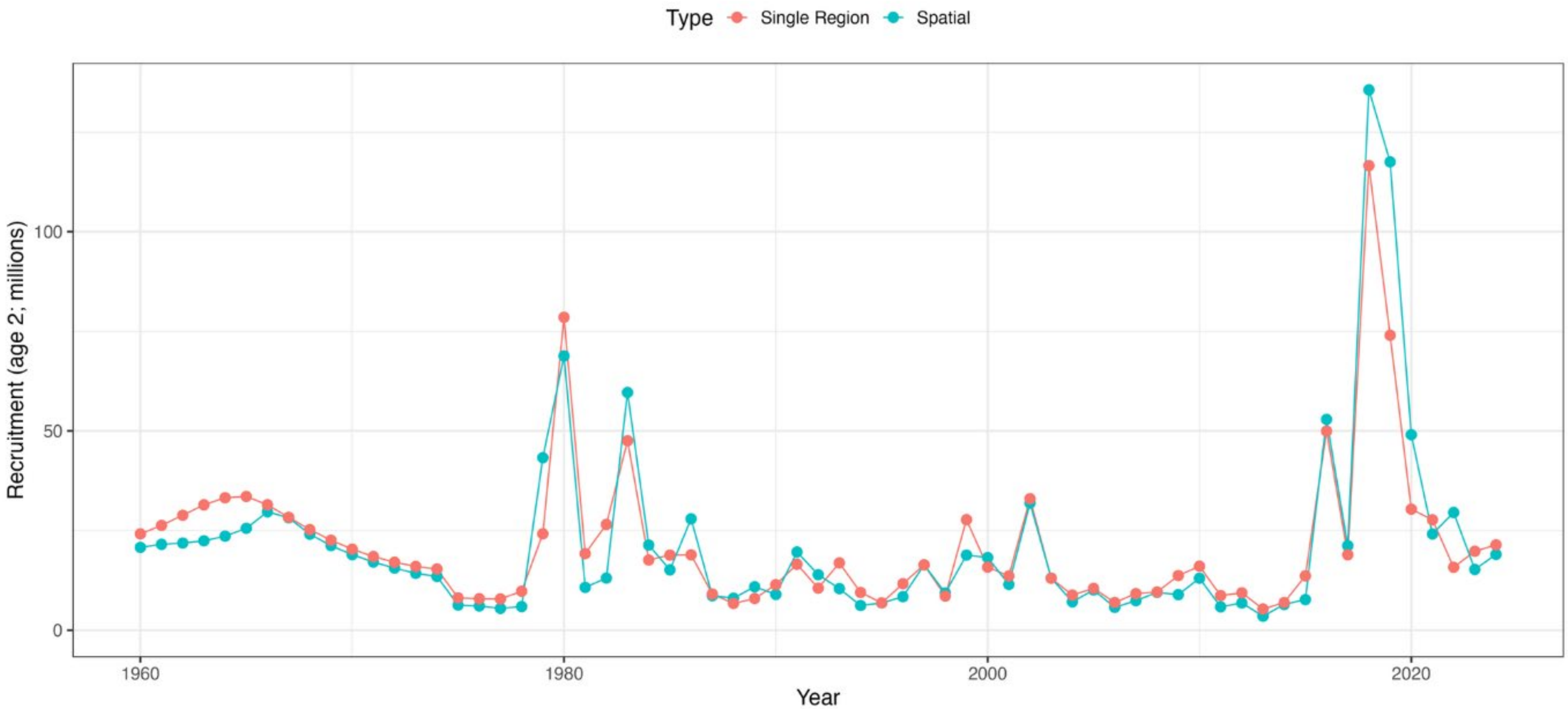


Figure 46. Comparison of Alaska-wide age 2 recruitment estimates (millions of fish) between the final author recommended model (Single Region: *25.12_Drop_TS_Upd_M*) and the sensitivity run of the spatial model (Five Regions: *25.15_Spatial*).

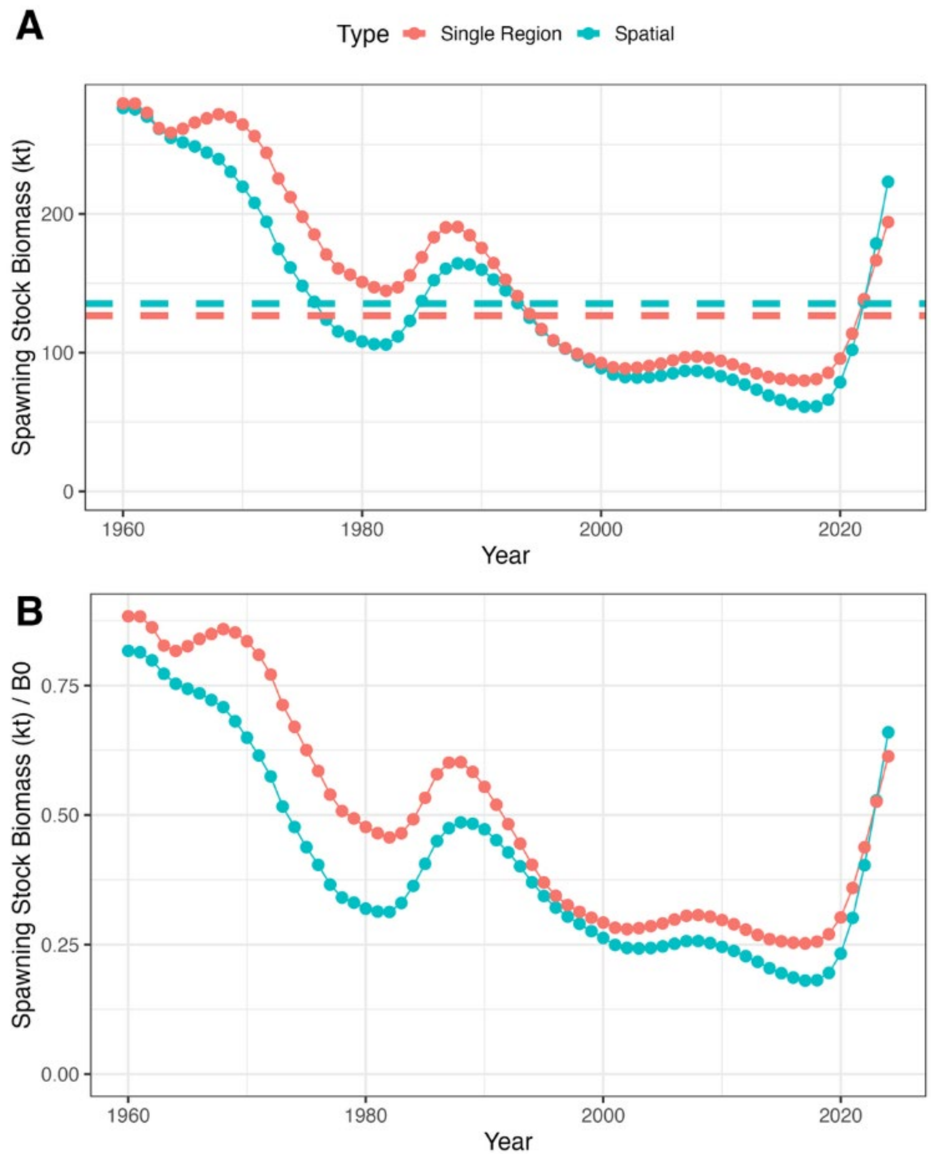


Figure 47. Comparison of Alaska-wide spawning stock biomass (panel A; dashed lines indicate the $B_{40\%}$ reference point, kt) and stock status (panel B) between the final author recommended model (Single Region: *25.12_Drop_TS_Upd_M*) and the sensitivity run of the spatial model (Five Region: *25.15_Spatial*).

6. Survey Retroactive Analysis

Introduction

Beginning in 2025 the NOAA longline survey will be undergoing a fundamental change to the survey design and annual survey footprint. As a cost-recovery survey, the ability to operationalize the survey in recent years has been severely impacted by rapidly declining sablefish markets. In 2024, the longline survey was not conducted for the first time since its inception in 1990. Thus, a reduction in survey footprint and length was required to ensure vessel participation. The new survey design involves sampling only the Gulf of Alaska in odd years and only the Bering Sea and Aleutian Islands in even years. Although mostly the same survey stations will be sampled within those regions, both the timing of when the stations will be sampled and the years in which they are sampled will change.

Historically, the longline survey sampled the GOA every year along with the Bering Sea in odd years and the Aleutian Islands in even years. To ensure that RPNs were reflective of the entire sablefish population (given that entire large regions were not sampled in a given year), an extrapolation process was implemented whereby the annual growth in the GOA was used to extrapolate the RPNs for a given unsurveyed area (either the BS or AI). Because the extent of that extrapolation will now be much larger (~50% of the RPNs would need to be extrapolated in any given year), it has been deemed more appropriate to use 'carry-over' RPNs for unsurveyed regions (i.e., the RPNs from the previous year will be used for years in which a given region is not surveyed). The carry-over approach would be expected to be precautionary during periods of population increase, but, perhaps, risk-prone in periods of population decline. However, any form of extrapolation carries the same risks, given that each region has moderately different dynamics (Figure 44; Cheng et al., 2025c). Yet, some form of extrapolation is required to ensure that the longline survey remains representative of the scale of the Alaska-wide population. Otherwise, time-varying catchability parameters would be required that would likely be inestimable and greatly increase model instability. In the future, a spatially explicit model might be pursued that can better fit survey data on a regional scale (e.g., model *25.15_Spatial*), thereby, allowing for years without a survey (instead of requiring extrapolation).

A 'retroactive' analysis was developed to provide insight into how the new survey design and approach for extrapolating unsurveyed regions (i.e., 'carry-over' of RPNs) might impact estimates from the stock assessment. A relatively straightforward method was implemented using the existing data to explore how the survey data might have changed had certain regions not been surveyed over the last 10 years. Mainly, what would the RPN index and associated age compositions have looked like had the BS (odd years) or GOA (even years) not been surveyed, which generally represents the survey design moving forward. Two forms of the retroactive analysis were developed to create two separate longline survey data sets, which represent the 2025 situation (only missing the BS; *Extrap_BSAI_Only*) and the long-term situation (missing the BS in odd years and missing the GOA in even years; *Extrap_BSAI_GOA*). Although the analysis is limited, it provides reasonable expectations of what the expected impact will be for 2025 (*Extrap_BSAI_Only*) and what the 'worst case' impact might be (*Extrap_BSAI_GOA*). Specifically, the purpose was not to test different extrapolation approaches or survey design options, while this analysis was deemed the most parsimonious given time and data constraints. Future simulation testing will be pursued using the spatially explicit management strategy evaluation (MSE) tool (Zahner et al., 2025), especially as new assessment models are explored to better handle the new survey design (e.g., the spatial model).

Methods

Two analyses were implemented to emulate the new survey design, where a new longline survey time series was developed for each, including recalculation of the RPN index and the RPN-weighted age compositions. The longline survey data was recalculated going back ~10 years to represent a period dating back to before the longline survey RPN index began to increase rapidly. The final author recommended model (*25.12_Drop_TS_Upd_M*) was then applied to the new data sets with Francis reweighting applied in each new model run, and key model outputs (e.g., SSB, recruitment, and catch advice) were compared across models.

The two analyses were:

- *Extrap_BSAI_Only* (demonstrate the potential impact for the 2025 assessment of not having a BS survey in an odd year):
    1. Remove only BS data in odd years (2015, 2017, 2019, 2021, 2023).
        - Because the AI is not surveyed in odd years, there is no need to account for loss of AI data.
        - Remove the BS age compositions in odd years.
        - Remove BS RPNs in odd years.
    2. Recalculate the RPN index in odd years based on 'carry-over' (constant) values from the AI in the previous even year and from the BS in the previous odd year (2 years previous).
        - Note that the implicit assumption of the BS 'carry-over' is that there would have been a BS survey in a preceding year from which to carry-over the RPN value.
        - The Alaska-wide RPN is calculated as the sum of the regional RPNs, including the carry-over values from the BS and AI.
    3. Recalculate the age compositions in odd years using the RPN-weighted regional compositions from the GOA only.
    4. Rerun the assessment model with the new time series of longline survey RPNs and age compositions.

- *Extrap_BSAI_GOA* (emulate the new long-term survey design where only the GOA is surveyed in odd years then the BS and AI in even years; this represents a 'worst-case' scenario given that only AI age composition data are now available and used in even years):
    1. For odd years (2015, 2017, 2019, 2021, 2023), follow the same methods as *Extrap_BSAI_Only* to represent only having GOA data in odd years.
    2. For even years (2014, 2016, 2018, 2020, 2022), remove the GOA data to represent having only BS and AI data in even years.
        - Remove the GOA age compositions in even years.
        - Remove the GOA RPNs in even years.
    3. Recalculate the RPN index in even years based on 'carry-over' (constant) values from the GOA in the previous odd year.

- The Alaska-wide RPN is calculated as the sum of the regional RPNs, including the carry-over values from the GOA.

4. Recalculate the age compositions in even years using compositions from the AI only (since there are no historical data from the BS in even years).
5. Rerun the model with the new time series of longline survey RPNs and age compositions for both even and odd years going back to 2014.

Results

In most years (except 2015 and 2018), the newly calculated index values are lower than the observed index, which is not surprising given the rapid population growth that has occurred since 2015 (Figure 48). The corresponding recalculated age compositions differ modestly from the full set of age compositions, particularly in even years when only the AI age compositions are used to represent the entire Alaska-wide population for the *Extrap_BSAI_GOA* analysis (Figures 49-50). Overall, the model fits do not differ drastically between models, though, as expected, predicted index values in recent years are smaller (Figure 48). Given that the terminal longline survey index value remains relatively large along with the long-lived nature of sablefish, the model is relatively inelastic to the changes in data.

SSB trends are only modestly impacted (Figure 51), primarily due to a slight decline in the catchability estimate in the *Extrap_BSAI_GOA* analysis. However, the alteration to the age compositions does have a moderate impact on the estimates of recent year class strength (Figure 52). In particular, the reliance on only AI age compositions in even years in the *Extrap_BSAI_GOA* analysis leads to larger magnitude 2017 and 2019 year classes, both of which are primarily only observed in the AI. Ultimately, the resulting management advice is minimally impacted (Table 13), suggesting that the change in the survey design is unlikely to have a large impact on model outputs under current population conditions (i.e., population growth). Thus, the impacts for 2025, where only the BS survey stations will be missing from the normal expected survey data, will likely be minimal (*Extrap_BSAI_Only* analysis; full results here). However, future impacts will be larger, especially when the GOA is not surveyed, given the more extreme differences in age compositions between the GOA and BS or AI (*Extrap_BSAI_GOA* analysis; full results here).

Conclusion

For 2025, using the 'carry-over' RPN approach is unlikely to have major impacts on assessment estimates. However, the analysis was only applied to a period of population growth, while the AI region has not been surveyed since 2022 (due to the 2024 survey being cancelled). The impacts of the carry-over method may become more detrimental if the population reverts to a period of more average recruitment and the population declines (i.e., using carry-over RPNs will then lead to overly optimistic Alaska-wide RPNs). The carry-over RPN approach will also lead to an implicit extrapolation of ~50% of the RPNs in a given year, because the regional RPNs are now almost evenly split between the GOA and BS/AI. The loss of age composition data is likely to be more detrimental to the assessment than the RPNs themselves, and there are likely to be noticeable impacts on interannual variability in recruitment estimates and increased recruitment retrospective patterns. In the future, the spatial assessment might be beneficial to better handle unsurveyed regions, but further work with the spatial MSE tool is warranted.

Tables

Table 13. Estimates of key assessment outputs for the *Extrap_BSAI_Only* and *Extrap_BSAI_GOA* analyses along with the author recommended model *25.12_Drop_TS_Upd_M*.

Model	Region	Terminal_SSB	Terminal_F	Catch_Advice	B_Ref_Pt	F_Ref_Pt	B_over_B_Ref	F_over_F_Ref
25_12_Drop_TS_Upd_M	1	194.0907	0.04842486	39.55329	126.6058	0.08308	1.53303	0.58289
5yr_Extrap_BSAI_Only	1	195.8136	0.04863375	39.38514	127.0446	0.08309	1.54130	0.58534
5yr_Extrap_BSAI_GOA	1	183.8196	0.05030035	38.69233	128.0174	0.08415	1.43590	0.59776

Figures

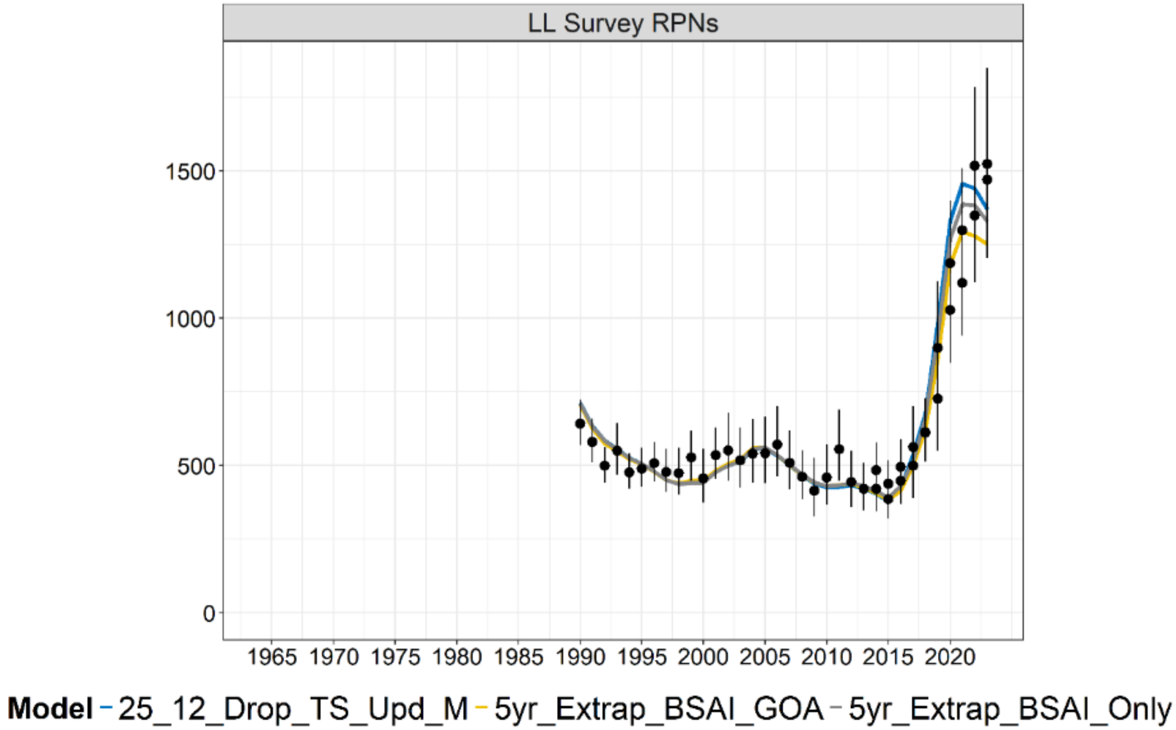


Figure 48. Observed (black points) and predicted (lines) longline survey relative populations numbers (RPNs). Note that the recalculated index is represented by the lower point (smaller index value) for each year dating back to 2014, except 2015 and 2018 (the larger index value is the new index value in the latter years, and the 2018 values are nearly identical between the original and new index). For the *Extrap_BSAI_Only* analysis only the odd year index values are recalculated. For the *Extrap_BSAI_GOA* analysis both the even and odd year index values are recalculated.

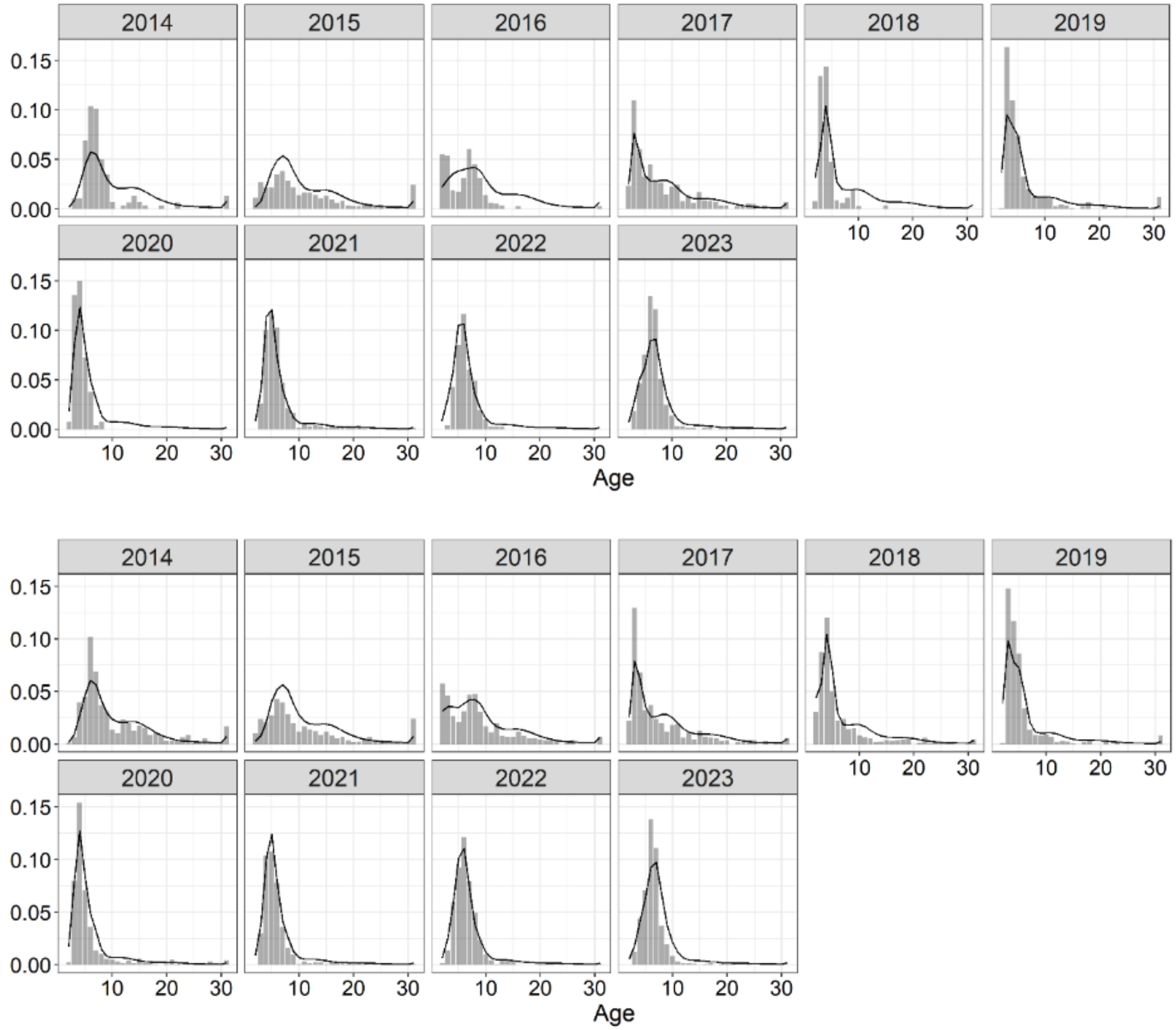


Figure 49. Observed (grey bars) and predicted (black lines) NOAA longline survey male age compositions after removing observations from non-surveyed areas (top panel; predicted values are from the *Extrap_BSAI_GOA* analysis) and with the full complement of age observations (bottom panel; predicted values are from model *25.12_Drop_TS_Upd_M*). For the *Extrap_BSAI_Only* analysis only the odd year compositions are recalculated. For the *Extrap_BSAI_GOA* analysis both the even and odd year compositions are recalculated based solely on Aleutian Island observations in even years.

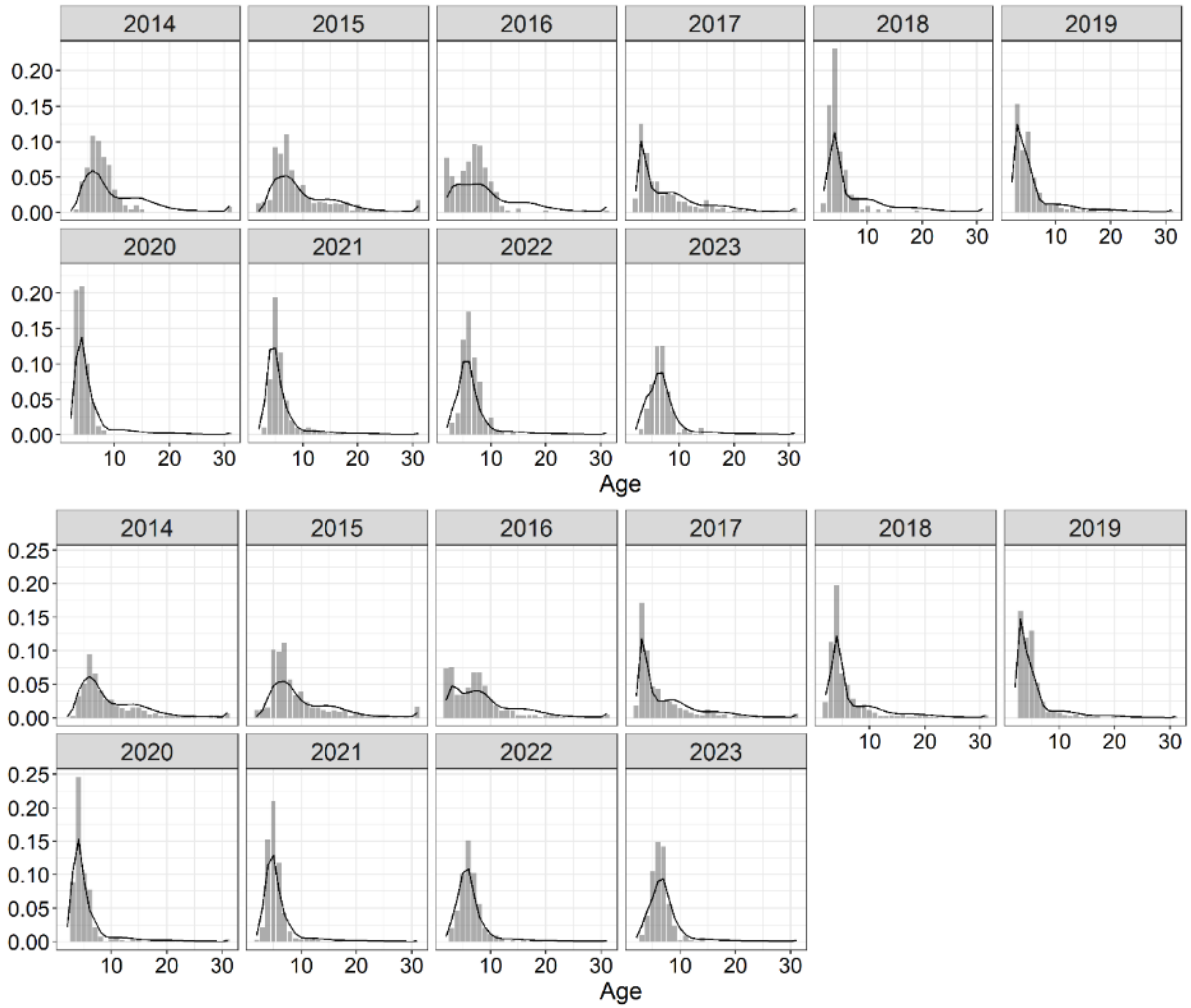


Figure 50. Observed (grey bars) and predicted (black lines) NOAA longline survey female age compositions after removing observations from non-surveyed areas (top panel; predicted values are from the *Extrap_BSAI_GOA* analysis) and with the full complement of age observations (bottom panel; predicted values are from model *25.12_Drop_TS_Upd_M*). For the *Extrap_BSAI_Only* analysis only the odd year compositions are recalculated. For the *Extrap_BSAI_GOA* analysis both the even and odd year compositions are recalculated based solely on Aleutian Island observations in even years.

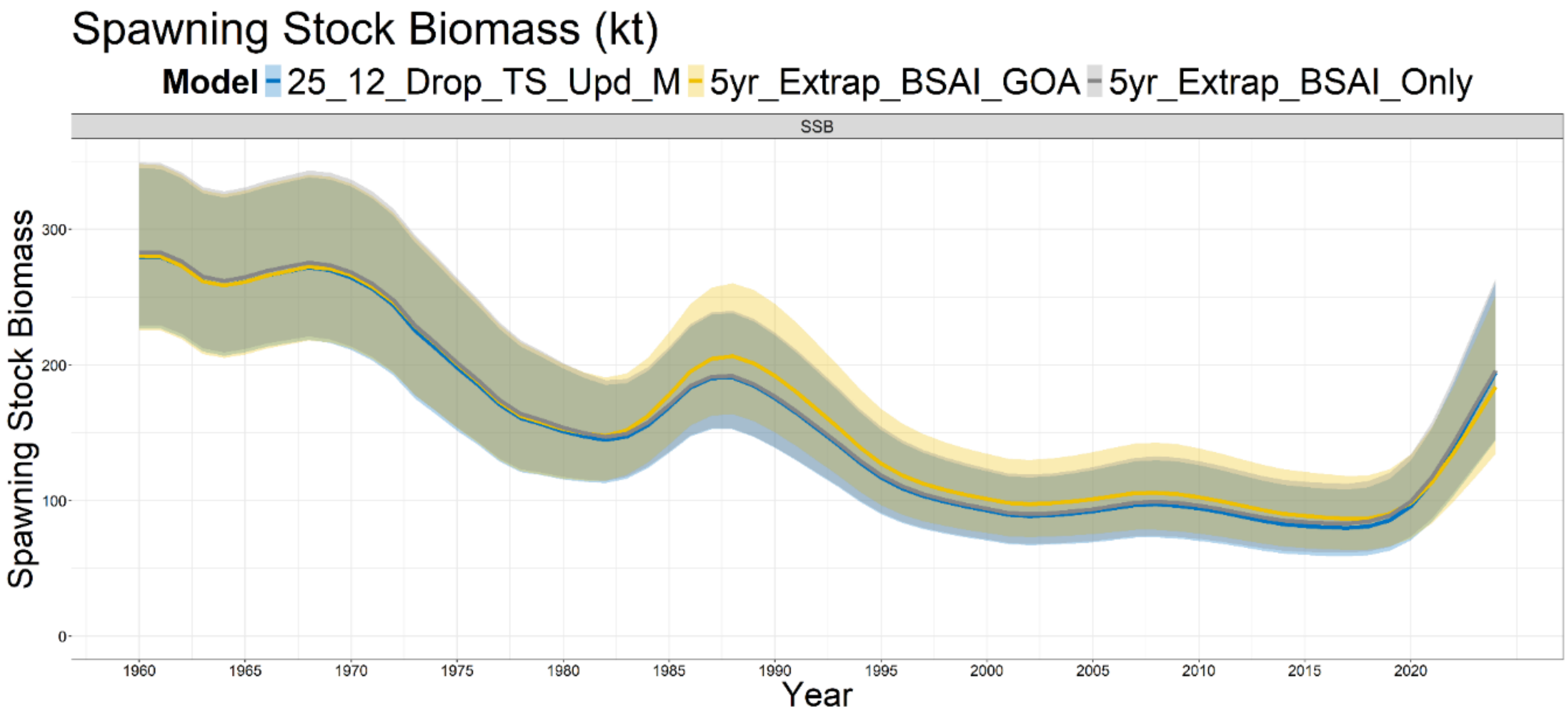


Figure 51. Comparison of spawning stock biomass (SSB, in kilotons) for the *Extrap_BSAI_Only* and *Extrap_BSAI_GOA* analyses. The 95% confidence intervals for each model run are provided by the associated shading.

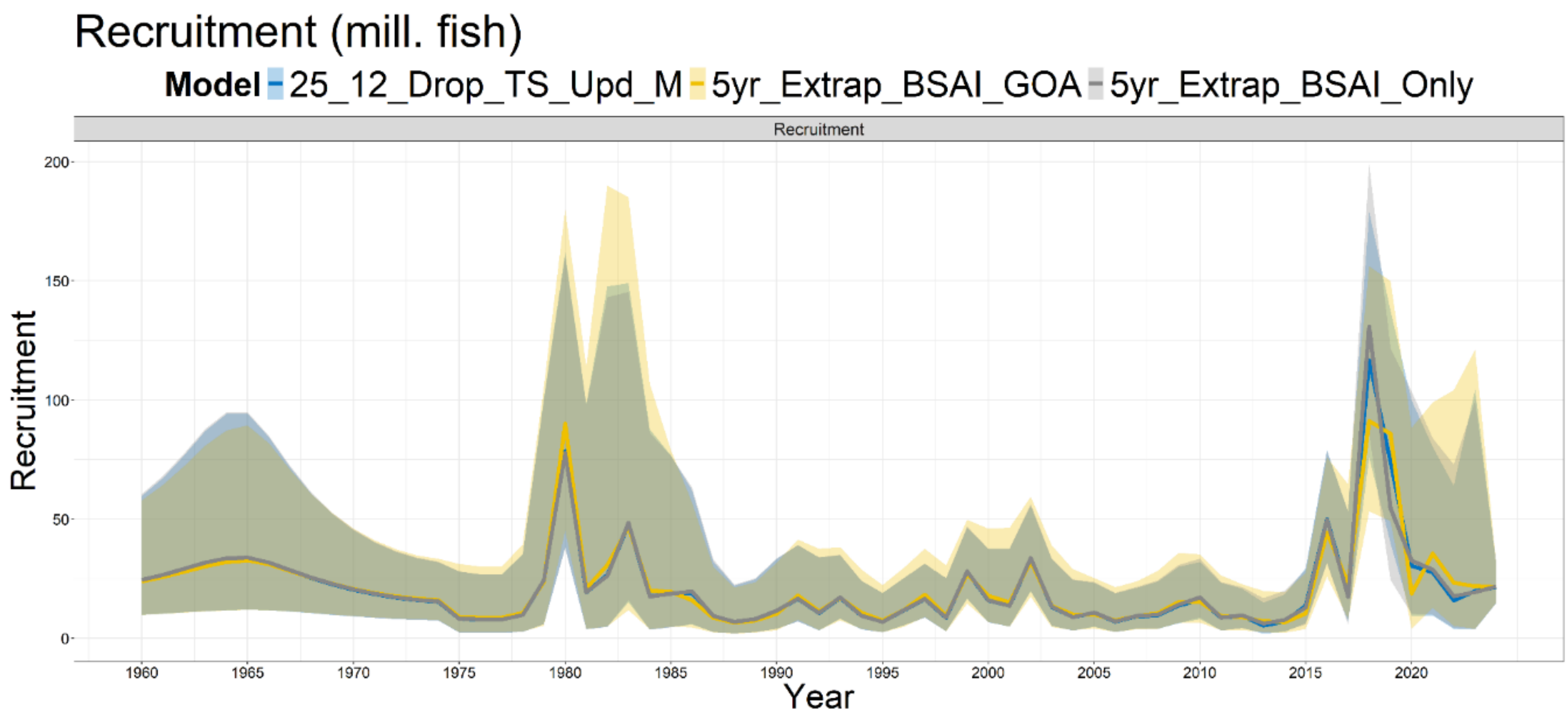


Figure 52. Comparison of recruitment (in millions of fish) for the *Extrap_BSAI_Only* and *Extrap_BSAI_GOA* analyses. The 95% confidence intervals for each model run are provided by the associated shading.

7. References

Cheng, M.L.H., Goethel, D.R., Cunningham, C.J., 2024a. Addressing complex fleet structure in fishery stock assessment models: Accounting for a rapidly developing pot fishery for Alaska sablefish (Anoplopoma fimbria). Fisheries Research 271, 106917. https://doi.org/10.1016/j.fishres.2023.106917

Cheng, M.LH., Goethel, D.R., Hulson, P.-J.F., Cunningham, C.J., 2024b. Confronting transitions in fishery fleet structure and selectivity: practical recommendations for integrated age-structured stock assessments based on simulation analysis. Can. J. Fish. Aquat. Sci. https://doi.org/10.1139/cjfas-2024-0129

Cheng, M.L.H., Goethel, D.R., Hulson, P.-J.F., Echave, K.B., and Cunningham, C.J. 2024c. ''Slim pickings?'': Extreme large recruitment events may induce density-dependent reductions in growth for Alaska sablefish (*Anoplopoma fimbria*) with implications for stock assessment. Canadian Journal of Fisheries and Aquatic Sciences. 82: 1-13. http://dx.doi.org/10.1139/cjfas-2024-0228

Cheng, M.L.H., Goethel, D.R., Hulson, P.-J.H., Wilberg, M.J., Marsh, C., and Cunningham, C.J. 2025a. Misspecifying sex-structured dynamics in stock assessment models. Fish and Fisheries. 26: 454-472. https://doi.org/10.1111/faf.12891

Cheng, M.L.H., Hulson, P.-J.H., Goethel, D.R., and Cunningham, C.J. 2025b. A mathematical proof comparing the statistical properties between two common approaches for parameterizing sex-composition likelihoods in fishery stock assessments. Fisheries Research. 281: 107231. https://doi.org/10.1016/j.fishres.2024.107231

Cheng, M.L.H., Marsh, C.A., Goethel, D.R., Hulson, P.-J.F., Echave, K.B., Williams, B.C., Berger, A.M., and Cunningham, C.J. 2025c. Panmictic Panacea? Demonstrating Good Practices for Developing Spatial Stock Assessments Through Application to Alaska Sablefish (*Anoplopoma fimbria*). Fish and Fisheries. 26: 825-847. https://doi.org/10.1111/faf.70002

DFO (Department of Fisheries and Oceans Canada). 2023. A revised operating model for sablefish in British Columbia in 2022. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2023/010. Accessed from: https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/41102587.pdf

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138. https://doi.org/10.1139/f2011-025

Francis, R.I.C.C. 2017. Revisiting data weighting in fisheries stock assessment. Fisheries Research. 192: 5-115. https://doi.org/10.1016/j.fishres.2016.06.006

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

Goethel, D. R., and Cheng, M. L. 2024. 3. Assessment of the Sablefish Stock in Alaska. NPFMC SAFE Reports. Accessed from: https://www.npfmc.org/wp-content/PDFdocuments/SAFE/2024/Sablefish.pdf

Hamel, O.S., and Cope, J.M. 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research. 256: 106477. https://doi.org/10.1016/j.fishres.2022.106477

Hanselman, D., Lunsford, C.R., Fujioka, J.T., and Rodgveller, C.J. 2007. Alaska sablefish assessment for 2007. NPFMC SAFE Reports. Accessed from: https://apps-afsc.fisheries.noaa.gov/refm/docs/2006/sablefish.pdf

Hanselman, D., Fenske, K., Rodgveller, C., and Lunsford, C. 2018. Alaska sablefish model update. NPFMC Groundfish Joint Plan Team Working Paper. Accessed from: https://meetings.npfmc.org/CommentReview/DownloadFile?p=c8cf7786-bc4d-485d-b431-fc709b0f8cc8.pdf&fileName=Sablefish_Modeling_0918_Hanselman.pdf

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. 2016. TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software, 70.

Maunder, M., and Piner, K.R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research. 192: 16-27. https://doi.org/10.1016/j.fishres.2016.04.022

Minte-Vera, C.V., Maunder, M., Aires-da-Silva, A.M., Satoh, K, and Uosaki, K. 2017. Get the biology right, or use size-composition data at your own risk. Fisheries Research. 192: 114-125. https://doi.org/10.1016/j.fishres.2017.01.014

Monnahan, C. C., Adams, G. D., Ferriss, B. E., Shotwell, S. K., McKelvey, D. R., and McGowan, D. W. 2023. 1. Assessment of the Walleye Pollock Stock in the Gulf of Alaska: 128. Accessed from: https://www.npfmc.org/wp-content/PDFdocuments/SAFE/2023/GOApollock.pdf

Ono, K., Licandeo, R., Muradian, M.L., et al. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES Journal of Marine Science. Volume 72: 31–43. https://doi.org/10.1093/icesjms/fsu007

Timm, L.E., Larson, W.A., Jasonowicz, A.J., Nichols, K.M., 2024. Whole genome resequencing of sablefish at the northern end of their range reveals genetic panmixia and large putative inversions. ICES Journal of Marine Science fsae070. https://doi.org/10.1093/icesjms/fsae070

Wang, S.-P., and Maunder, M. 2017. Is down-weighting composition data adequate for dealing with model misspecification, or do we need to fix the model? Fisheries Research. 192: 41-51. https://doi.org/10.1016/j.fishres.2016.12.005

Wetzel, C.R>, Berger, A.M., Barnes, C., Zahner, J.A., Tolimieri, N., Ward, E.J., and Head, M. 2025. Status of the sablefish (*Anoplopoma fimbria*) off the U.S. West Coast in 2025. Pacific Fisheries Management Council, Portland, Oregon.

Williams, B. C., Hulson, P.-J. F., and Ferris, B. E. 2024. 10. Assessment of the Northern Rockfish Stock in the Gulf of Alaska. Accessed from: https://www.npfmc.org/wp-content/PDFdocuments/SAFE/2024/GOAnork.pdf

Zahner, J.A., Goethel, D., Cunningham, C., et al. 2025. Sablefish MSE Project Update. NPFMC Groundfish Joint Plan Team.

Appendix 1: Technical Description of Assessment Model

The 2025 SAFE for Alaskan sablefish uses the SPoRC package, which supports single-region and spatial models. The operational assessment is a single, Alaska-wide model, whereas a spatially explicit model is also developed as a sensitivity run. The models are age- and sex-structured, consisting of 30 ages (2–31), 2 sexes (females and males), and 2 fishery fleets (fixed-gear and trawl). The spatial model includes five regions: the Bering Sea, Aleutian Islands, Western Gulf of Alaska, Central Gulf of Alaska, and Eastern Gulf of Alaska. The sections below outline the generalized model structure for these models.

Population dynamics follow an annual time step, which proceed in the following sequence:

1. Recruitment and tag releases initially occur (tag releases only occur if tagging data are used),
2. Markovian movement of individuals then follow (movement only occurs in the spatial model), and
3. Total mortality, chronic tag loss (if applicable), and ageing processes occur.

These processes are modeled across four main partitions: region (r), year (y), age (a and a_+, where a_+ is the plus group), and sex (s). In single-region models, most equations reduce by setting $r = 1$.

Process Equations

Population Initialization

Initialization of equilibrium abundance differs between the single-region and spatial models due to the need to incorporate movement dynamics in the latter. In the single-region model, the initial equilibrium population is computed as:

$$N'_{r=1,a,s} = \mu^{\text{Rec}} \exp\left(-(a-1) \cdot \left(Z'_{r=1,a,s}\right)\right) \psi_s, \quad \text{for } 2 \leq a < a_+$$

$$Z'_{r=1,a,s} = \text{Natmort}_{r=1,y=1,a,s} + \mu^{\text{Fsh}}_{r=1,f=1} \text{Fmort}^{\text{InitProp}} \text{Sel}^{\text{Fsh}}_{r=1,y=1,a,s,f=1}$$

where $N'_{r=1,a,s}$ are the equilibrium numbers-at-age, μ^{Rec} is a global recruitment parameter, $Z'_{r=1,a,s}$ is the initial instantaneous total mortality rate, $\text{Natmort}_{r=1,y=1,a,s}$ is the instantaneous natural mortality rate, $\mu^{\text{Fsh}}_{r=1,f=1}$ is the mean fishing mortality rate from the first fleet (f denotes fishery fleet), $\text{Fmort}^{\text{InitProp}}$ is the proportion of the mean fishing mortality from the first fleet applied during the initialization stage, $\text{Sel}^{\text{Fsh}}_{r=1,y=1,a,s,f=1}$ is the fishery selectivity-at-age for the first fleet, ψ_s is the sex ratio of the equilibrium population (fixed at 50:50).

The plus group (a_+) of the initial population is then derived by solving for a geometric series:

$$N'_{r,a,s} = \mu^{\text{Rec}} \frac{\exp\left(-(a_+ - 1) \cdot \left(Z'_{r=1,a=a_+,s}\right)\right)}{1 - \exp\left(Z'_{r=1,a=a_+,s}\right)} \psi_s$$

By contrast, the initial population abundance for the spatial model is derived by numerically iterating the population to equilibrium, since a closed-form solution for the plus group cannot easily be obtained. In this model structure, the total instantaneous mortality in each region is defined similarly to the single-region case. An exponential decay model is used to initialize the age structure at the first iteration:

$$N'_{r,a,s} = \begin{cases} \mu^{\text{Rec}}\,\psi_s\,\zeta_r, & \text{if } a = 2 \\ \mu^{\text{Rec}}\,\psi_s\,\zeta_r \exp\left(-\sum_{j=2}^{a} Z'_{r,j,s}\right), & \text{if } a > 2 \end{cases}$$

where ζ_r apportions the global mean recruitment across regions, and is derived using a multinomial logit transform to ensure proportions sum to one. The initialized age structure is then iterated forward to equilibrium by applying recruitment, movement, and mortality and ageing processes in order (see Population Projection section for equations).

Following the definition of equilibrium age structure for both the single-region and spatial model, initial age deviations are applied:

$$N_{r,y=1,a\notin 2,s} = N'_{r,a\notin 2,s}\exp\left(\epsilon_{r,i}^{\text{Init}}\right)$$

$N_{r,y=1,a\notin 2,s}$ are the stochastic numbers-at-age in the first year, and initial age deviations ($\epsilon_{r,i}$) are constrained by a lognormal distribution.

Recruitment Processes

In both model types, recruitment is estimated based on lognormal deviations about a mean recruitment parameter:

$$N_{r,y,a=2,s} = \mu^{\text{Rec}}\exp\left(\epsilon_{r,y}^{\text{Rec}} - \frac{\sigma_{\text{Rec}}^2}{2}b_y\right)\psi_s\zeta_r$$

Here, $\epsilon_{r,y}$ are annual, lognormally distributed recruitment deviations with a lognormal bias correction term ($\frac{\sigma_{\text{Rec}}^2}{2}b_y$).

Population Projection

Following recruitment processes, the population can then be projected forward. In the context of the spatial model, Markovian movement dynamics are first applied:

$$\mathbf{N}_{y,a,s} = \left(\mathbf{N}_{y,a,s}\right)^{\text{T}}\mathbf{M}_{y,a,s}$$

where $\mathbf{M}_{y,a,s}$ is a first-order Markov matrix representing movement.

After the population undergoes movement, individuals experience mortality and ageing processes via an exponential mortality model:

$$N_{r,y+1,a+1,s} = N_{r,y,a,s}\exp\left(-Z_{r,y,a,s}\right), \quad \text{for } 2 \leq a < a_{+}$$
$$N_{r,y+1,a,s} = N_{r,y,a-1,s}\exp\left(-Z_{r,y,a-1,s}\right) + N_{r,y,a,s}\exp\left(-Z_{r,y,a,s}\right), \quad \text{for } a = a_{+}$$

$Z_{r,y,a,s}$ denotes the annual total instantaneous mortality rate and is defined as:

$$Z_{r,y,a,s} = \text{Natmort}_{r,y,a,s} + \sum_{f} \text{Sel}_{r,y,a,s,f}^{\text{Fsh}}\, \text{Fmort}_{r,y,f}$$

$\text{Fmort}_{r,y,f}$ is the annual instantaneous fishing mortality rate parameterized based on lognormal deviations about a mean fishing mortality parameter for a given fishery fleet:

$$\text{Fmort}_{r,y,f} = \mu_{r,f}^{\text{Fsh}}\exp\left(\epsilon_{r,y,f}^{\text{Fsh}}\right).$$

Observation Equations

Fishery Observation Model

The fishery observation model describes the expected catch-at-age, catch-at-length, catch (in units of biomass), and fishery indices. Expected catch-at-age ($C_{r,y,a,s,f}^{a}$) is computed using Baranov's catch equation:

$$C_{r,y,a,s,f}^{a} = \frac{\text{Fmort}_{r,y,f}\text{Sel}_{r,y,a,s,f}^{\text{Fsh}}}{Z_{r,y,a,s}} N_{r,y,a,s}\left[1 - \exp\left(-Z_{r,y,a,s}\right)\right]$$

Catch-at-length ($C_{r,y,l,s,f}^{l}$) can then be derived using the following equation:

$$C_{r,y,l,s,f}^{l} = \left(\mathbf{A}_{r,y,s}^{l}\right)^{\mathrm{T}}\mathbf{C^{a}}_{r,y,s,f}$$

where $\mathbf{A}_{r,y,s}^{l}$ is the size-age transition matrix. Expected catch ($\text{Catch}_{r,y,f}$) is then computed summation of the expected catch-at-age across ages and sexes, multiplied by their respective weights-at-age ($W_{r,y,a,s}$):

$$\text{Catch}_{r,y,f} = \sum_{a}^{a_{+}}\sum_{s}^{n_{s}} C_{r,y,a,s,f}^{a}\, W_{r,y,a,s}$$

The expected fishery index ($\text{FshIdx}_{r,y,f}$) is computed as:

$$\text{FshIdx}_{r,y,f} = q_{r,y,f}^{\text{Fsh}} \sum_{a}^{a_+} \sum_{s}^{n_s} N_{r,y,a,s} \exp(-Z_{r,y,a,s} 0.5) \text{Sel}_{r,y,a,s,f}^{\text{Fsh}} W_{r,y,a,s}$$

where $q_{r,y,f}^{\text{Fsh}}$ is the catchability coefficient for a given fishery fleet.

Survey Observation Model

Similarly, the survey observation model describes the expected survey catch-at-age, survey catch-at-length, and survey indices. Expected survey catch-at-age ($I_{r,y,a,s,sf}^{a}$) is calculated as follows:

$$I_{r,y,a,s,sf}^{a} = N_{r,y,a,s} \exp(-Z_{r,y,a,s} 0.5) \text{Sel}_{r,y,a,s,sf}^{\text{Srv}}$$

where subscript sf denotes a given survey fleet and $\text{Sel}_{r,y,a,s,sf}^{\text{Srv}}$ is the survey selectivity-at-age pattern. Expected survey catch-at-length ($I_{r,y,l,s,sf}^{l}$) is given by:

$$I_{r,y,l,s,sf}^{l} = \left(\mathbf{A}^{l}_{r,y,s}\right)^{\text{T}} \mathbf{I^a}_{r,y,s,f}$$

Survey indices ($\text{SrvIdx}_{r,y,sf}$) is computed as either abundance-based or biomass-based. Abundance-based survey indices are calculated as:

$$\text{SrvIdx}_{r,y,sf} = q_{r,y,sf}^{\text{Srv}} \sum_{a}^{a_+} \sum_{s}^{n_s} I_{r,y,a,s,sf}^{a}$$

while biomass-based indices are computed as:

$$\text{SrvIdx}_{r,y,sf} = q_{r,y,sf}^{\text{Srv}} \sum_{a}^{a_+} \sum_{s}^{n_s} I_{r,y,a,s,sf}^{a} W_{r,y,a,s}$$

In both equations, $q_{r,y,sf}^{\text{Srv}}$ represents the survey catchability coefficient.

Tagging Observation Model

The tagging observation model tracks tag cohorts ($T_{r,y,a,s}^{k}$) by the combination of release region and release year (k) and follows a Brownie tag attrition framework. Tag cohorts are tracked for a pre-defined maximum duration (maximum tag liberty; n_L), after which calculations for the tag cohort are no longer computed. This is done to help with computation time. Additionally, tag cohorts also have the option to undergo a fraction of mortality in the year of release, such that if a given cohort is released during the middle of the year, mortality processes are discounted. If mortality discounting is specified, movement does not occur in the year of release. In general, the process dynamics for the tagged cohort mimics those specified for the overall population.

Immediately following release, tag cohorts are decremented by an initial tag-induced mortality rate:

$$T^{k}_{r,y,a,s} = T^{k}_{r,y,a,s}\exp(-\tau)$$

where τ is the initial tag-induced mortality rate. If tagged cohorts are released at the beginning of a calendar year, Markovian movement occurs (movement does not occur otherwise):

$$\mathbf{T}^{k}_{y,a,s} = \left(\mathbf{T}^{k}_{y,a,s}\right)^{\mathrm{T}}\mathbf{M}_{y,a,s}$$

Mortality and ageing of the tagged cohort then occur:

$$\begin{aligned} T^{k}_{r,y+1,a+1,s} &= T^{k}_{r,y,a,s}\exp\left(-Z^{\mathrm{Tag}}_{r,y,a,s}\right), \quad \text{for } 2 \le a < a_{+} \\ T^{k}_{r,y+1,a,s} &= T^{k}_{r,y,a-1,s}\exp\left(-Z^{\mathrm{Tag}}_{r,y,a-1,s}\right) + T^{k}_{r,y,a,s}\exp\left(-Z^{\mathrm{Tag}}_{r,y,a,s}\right), \quad \text{for } a = a_{+} \end{aligned}$$

where tagged cohorts follow an exponential mortality model, with accumulation of individuals in the plus-group. Total mortality for the tagged cohort ($Z^{\mathrm{Tag}}_{r,y,a,s}$) is specified as:

$$Z^{\mathrm{Tag}}_{r,y,a,s} = \left(\kappa + \mathrm{NatMort}_{r,y,a,s} + \sum_{f} \mathrm{Sel}^{\mathrm{Fsh}}_{r,y,a,s,f}\ \mathrm{Fmort}_{r,y,f}\right)$$

where κ is a parameter describing chronic tag loss (i.e., annual tag shedding). Similar to computations for catch-at-age, tag recaptures are calculated using a modified version of Baranov's catch equation:

$$\mathrm{Recap}^{k}_{r,y,a,s} = \beta_{r,y}\frac{\sum_{f}\mathrm{Sel}^{\mathrm{Fsh}}_{r,y,a,s,f}}{Z^{\mathrm{Tag}}_{r,y,a,s}}T^{k}_{r,y,a,s}\left[1-\exp\left(-Z^{\mathrm{Tag}}_{r,y,a,s}\right)\right]$$

$\beta_{r,y}$ represents a tag reporting rate parameter that can vary across space and time, and is estimated in logit space such that it is constrained between $[0,1]$.

Likelihoods

The model incorporates likelihood components for the following data sources:

1. fishery catches,
2. fishery indices,
3. fishery age compositions,
4. fishery length compositions,
5. survey indices,
6. survey age compositions,
7. survey length compositions, and
8. tagging data.

The total likelihood is the sum of the individual likelihood contributions from these sources, along with penalties and priors. This total is minimized using a non-linear optimization algorithm to estimate model parameters.

Observation Likelihoods

Fishery catches are fit using a lognormal likelihood. The log-likelihood for observed catch, $L\left(\log(\text{ObsCatch}_{r,y,f})\right)$, is defined as:

$$L\left(\log(\text{ObsCatch}_{r,y,f})\right) = \lambda_{\text{ObsCatch}_{r,y,f}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_{\text{ObsCatch}_{r,y,f}}} \exp\left(-\frac{\left[\log(\text{ObsCatch}_{r,y,f}) - \log(\text{Catch}_{r,y,f})\right]^2}{2\sigma^2_{\text{ObsCatch}_{r,y,f}}}\right)$$

Here, $\lambda_{\text{ObsCatch}_{r,y,f}}$ is the likelihood weight, $\text{ObsCatch}_{r,y,f}$ is the observed catch, $\text{Catch}_{r,y,f}$ is the predicted catch, and $\sigma^2_{\text{ObsCatch}_{r,y,f}}$ is the variance on the log scale.

Fishery indices are also fit using a lognormal likelihood. The log-likelihood for observed indices is:

$$L\left(\log(\text{ObsFshIdx}_{r,y,f})\right) = \lambda_{\text{ObsFshIdx}_{r,y,f}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_{\text{ObsFshIdx}_{r,y,f}}} \exp\left(-\frac{\left[\log(\text{ObsFshIdx}_{r,y,f}) - \log(\text{FshIdx}_{r,y,f})\right]^2}{2\sigma^2_{\text{ObsFshIdx}_{r,y,f}}}\right)$$

The parameter $\lambda_{\text{ObsFshIdx}_{r,y,f}}$ controls the weight of fishery indices to the objective function. $\text{ObsFshIdx}_{r,y,f}$ represents the observed fishery indices, and $\sigma^2_{\text{ObsFshIdx}_{r,y,f}}$ denotes the variance of the index.

Similarly, survey indices are fit to assuming a lognormal likelihood, which is given by:

$$L\left(\log(\text{ObsSrvIdx}_{r,y,sf})\right) = \lambda_{\text{ObsSrvIdx}_{r,y,sf}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_{\text{ObsSrvIdx}_{r,y,sf}}} \exp\left(-\frac{\left[\log(\text{ObsSrvIdx}_{r,y,sf}) - \log(\text{SrvIdx}_{r,y,sf})\right]^2}{2\sigma^2_{\text{ObsSrvIdx}_{r,y,sf}}}\right)$$

Again, $\lambda_{\text{ObsSrvIdx}_{r,y,sf}}$ is the likelihood weight applied to survey indices, $\text{ObsSrvIdx}_{r,y,sf}$ are the observed survey indices, and $\sigma^2_{\text{ObsSrvIdx}_{r,y,sf}}$ indicates the variance of the index.

Both fishery and survey compositions are fitted assuming a multinomial likelihood, which is given by:

$$L\left(\left(\text{ObsCompositionData}_{r,y,j}\right) = \lambda_{\text{ObsCompositionData}_{r,y,j}} \cdot \text{ISS}_{r,y,j} \cdot \prod_{b=1}^{n_b} E_{r,y,b,j}^{O_{r,y,b,j}}\right.$$

where subscript j is used to indicate a fishery or survey fleet. $\lambda_{\text{ObsCompositionData}_{r,y,j}}$ are likelihood weights applied to composition data (can be derived using iterative weighting methods), $ISS_{r,y,j}$ is the input sample size, the b subscript generically indicates a bin number, $E_{r,y,b,j}$ denotes the expected composition proportions, and $O_{r,y,b,j}$ are the observed composition proportions.

Within the SPoRC framework, compositions can be arranged to be split by sex or joint by sex. For brevity, we will describe the joint by sex approach, which is the author recommended approach (i.e., models 25.7 +). Here, composition proportions are assumed to sum to one jointly across both ages and sexes. Thus, for the expected fishery age compositions, this is computed as:

$$E_{r,y,b,j} = \frac{\left[C_{r,y,a,s=1,f}^{a},\ C_{r,y,a,s=2,f}^{a}\right]}{\sum_a \sum_s C_{r,y,a,s,f}^{a}}$$

with $b \in \{1, \dots, n_a \cdot n_s\}$. The expected survey age compositions are computed similarly as:

$$E_{r,y,b,j} = \frac{\left[I_{r,y,a,s=1,sf}^{a},\ I_{r,y,a,s=2,sf}^{a}\right]}{\sum_a \sum_s I_{r,y,a,s,sf}^{a}}$$

For expected fishery length compositions, this is given by:

$$E_{r,y,b,j} = \frac{\left[C_{r,y,l,s=1,f}^{l},\ C_{r,y,l,s=2,f}^{l}\right]}{\sum_l \sum_s C_{r,y,l,s,f}^{a}}$$

where $b \in \{1, \dots, n_l \cdot n_s\}$. Expected survey length compositions follow in a similar fashion:

$$E_{r,y,b,j} = \frac{\left[I_{r,y,l,s=1,f}^{l},\ I_{r,y,l,s=2,f}^{l}\right]}{\sum_l \sum_s I_{r,y,l,s,f}^{a}}$$

To fit tagging data, a multinomial release-conditioned likelihood is assumed, which describes both recaptured and non-recaptured states. The observed and expected recapture proportions are defined as:

$$\begin{aligned} \text{PropRecap}^k_{r,y,a,s} &= \frac{\text{Recap}^k_{r,y,a,s}}{\text{InitTag}^k} \\ \text{ObsPropRecap}^k_{r,y,a,s} &= \frac{\text{ObsRecap}^k_{r,y,a,s}}{\text{InitTag}^k} \end{aligned}$$

where InitTag^k is the total number of tags released for tag cohort k.

The corresponding non-recapture proportions are computed as:

$$\begin{aligned} \text{PropNonRecap}^k &= 1 - \sum_r \sum_y \sum_a \sum_s \text{PropRecap}^k_{r,y,a,s} \\ \text{ObsPropNonRecap}^k &= 1 - \sum_r \sum_y \sum_a \sum_s \text{ObsPropRecap}^k_{r,y,a,s} \end{aligned}$$

Expected and observed tagging outcomes are represented by the following vectors:

$$\begin{aligned} \mathbf{O}^k_{\text{Tagging}} &= \left\{\text{ObsPropRecap}^k_{r,y,a,s},\ \text{ObsPropNonRecap}^k\right\} \\ \mathbf{E}^k_{\text{Tagging}} &= \left\{\text{PropRecap}^k_{r,y,a,s},\ \text{PropNonRecap}^k\right\} \end{aligned}$$

A multinomial likelihood is then used to evaluate tagging data:

$$L^k_{\text{Tagging}} = \lambda_{\text{Tagging}} \cdot \text{InitTag}^k \prod_i \left(E^k_{i,\text{Tagging}}\right)^{O^k_{i,\text{Tagging}}}$$

Here, λ_{Tagging} is the likelihood weight, and i indexes elements of the observed and expected tagging proportion vectors.

Several priors and penalties are applied to constrain, stabilize, and inform parameter estimation. Specifically, priors are used for natural mortality, selectivity, movement, recruitment proportions, and tag reporting rates. The last three parameter types are only applicable for the spatial model. Additionally, penalties are used to constrain the estimation of parameters associated with deviations, which include: initial ages, recruitment, and fishing mortality.

In the case of natural mortality, a lognormal prior is utilized:

$$P\big(\log(\text{NatMort}_{s=1})\big) = \frac{1}{\sqrt{2\pi}\,\sigma_{\text{NatMort}}} \exp\left(-\frac{[\log(\text{NatMort}_{s=1}) - \log(\mu^{\text{NatMort}})]^2}{2\sigma^2_{\text{NatMort}}}\right)$$

This prior is time-invariant and applied only to the first sex. The variance of the prior is given by $\sigma^2_{\text{NatMort}}$, and μ^{NatMort} denotes the prior mean.

Selectivity priors are also utilized which serve as regularizing priors to facilitate stable parameter estimation. These priors are assumed to be lognormal and are applied to the selectivity parameters themselves:

$$P\left(\theta_{r,p,t,s,j}\right) = \frac{1}{\theta_{r,p,t,s,j}\sqrt{2\pi}\,\sigma^{\mathrm{Sel}}_{r,p,t,s,j}} \exp\left(-\frac{\left[\ln\left(\theta_{r,p,t,s,j}\right) - \ln\left(\mu^{\mathrm{Sel}}_{r,p,t,s,j}\right)\right]^2}{2\left(\sigma^{\mathrm{Sel}}_{r,p,t,s,j}\right)^2}\right)$$

where $\theta_{r,p,t,s,j}$ is the selectivity parameter for region r, parameter p, time period t, sex s, and fleet j (fishery or survey). $\mu^{\mathrm{Sel}}_{r,p,t,s,j}$ is the prior mean, while $\left(\sigma^{\mathrm{Sel}}_{r,p,t,s,j}\right)^2$ is the prior variance.

Priors on movement values are assumed to arise from a Dirichlet process:

$$P\left(\mathbf{M}_{k,y,a,s}\right) = \frac{\Gamma\left(n_r \cdot \alpha_{y,a,s}\right)}{\left[\Gamma\left(\alpha_{y,a,s}\right)\right]^{n_r}} \prod_{r=1}^{n_r} M_{k,r,y,a,s}^{\alpha_{y,a,s}-1}$$

where $\mathbf{M}_{k,y,a,s} = \{M_{k,1,y,a,s}, \dots, M_{k,n_r,y,a,s}\}$ is the vector of movement probabilities from region k to all regions, and $\alpha_{y,a,s}$ is the concentration parameter controlling the strength of the prior for year y, age a, and sex s.

Regional recruitment is derived by apportioning a global recruitment parameter using regional recruitment proportions (i.e., $\mu^{\mathrm{Rec}} \cdot \zeta_r$). Here, ζ_r is derived via a multinomial logit transformation and Dirichlet priors are used to help constrain estimation:

$$P(\boldsymbol{\zeta}) = \frac{\Gamma(n_r \cdot c)}{[\Gamma(c)]^{n_r}} \prod_{r=1}^{n_r} \zeta_r^{c-1}$$

$\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \dots, \zeta_{n_r}\}$ are the estimated recruitment proportions across regions, and c is a shared concentration parameter governing the spread of the Dirichlet distribution.

To facilitate the estimation of tag reporting rates, a symmetric beta distribution is applied:

$$P\left(\beta_{r,y}\right) = \left(\beta_{r,y} + 1e - 4\right)^{\beta_{\mathrm{PriorScale}}} \left(1 - \beta_{r,y} + 1e - 4\right)^{\beta_{\mathrm{PriorScale}}}$$

$\beta_{\mathrm{PriorScale}}$ determines the scale of the tag reporting parameter and determines how strongly to penalize estimates when they approach the bounds of $[0,1]$.

Estimation of initial age deviations $\epsilon^{\mathrm{Init}}_{r,i}$ is constrained by the following normal likelihood:

$$L\left(\epsilon^{\mathrm{Init}}_{r,i}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{\mathrm{Init}}} \exp\left(-\frac{\left(\epsilon^{\mathrm{Init}}_{r,i}\right)^2}{2\sigma^2_{\mathrm{Init}}}\right)$$

with a mean of 0 and a variance determined by σ^2_{Init}.

Similarly, estimation of recruitment deviations is also constrained by a normal likelihood with a mean of 0 and a variance of σ^2_{Rec}. This is given by:

$$L\left(\epsilon^{\text{Rec}}_{r,y}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{\text{Rec}}} \exp\left(-\frac{\left(\epsilon^{\text{Rec}}_{r,y}\right)^2}{2\sigma^2_{\text{Rec}}}\right)$$

Estimation of fishing mortality deviations is also constrained by a normal likelihood. Fishing mortality deviations are then assumed to have a mean of 0 and a variance of σ^2_{Fsh}:

$$L\left(\epsilon^{\text{Fsh}}_{r,y,f}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{\text{Fsh}}} \exp\left(-\frac{\left(\epsilon^{\text{Fsh}}_{r,y,f}\right)^2}{2\sigma^2_{\text{Fsh}}}\right)$$

Lastly, the joint likelihood to be minimized represents the sum of all observational likelihood components, priors, and penalties defined above:

$$\text{Joint Likelihood} = \sum \text{Observation Likelihoods} + \sum \text{Priors} + \sum \text{Penalties}$$

Note that some of these components may be zero depending on the configuration of the model (e.g., tagging likelihood would equal 0 in a single-region model).

Table A1. Description of symbols used in the model

Symbol	Description
r	Region index
y	Year index
a	Age index
a_+	Plus group age
s	Sex index
f	Fishery fleet index
sf	Survey fleet index
k	Tag release cohort index (region + year)
$N'_{r,a,s}$	Initial equilibrium numbers-at-age
$N_{r,y,a,s}$	Numbers-at-age by region, year, age, sex
μ^{Rec}	Global recruitment parameter
ψ_s	Sex ratio of population
ζ_r	Regional recruitment proportion
b_y	Bias correction term for recruitment
$Z'_{r,a,s}$, $Z_{r,y,a,s}$	Total instantaneous mortality rate
$\mathrm{Natmort}_{r,y,a,s}$	Natural mortality rate
$\mu^{\mathrm{Fsh}}_{r,f}$	Mean fishing mortality for fleet f
$\mathrm{Fmort}^{\mathrm{InitProp}}$	Proportion of fishing mortality in initialization
$\mathrm{Fmort}_{r,y,f}$	Fishing mortality for fleet f
$\mathbf{M}_{y,a,s}$	Movement matrix (Markov transition)
$C^{a}_{r,y,a,s,f}$	Expected catch-at-age
$\mathbf{A}^{l}_{r,y,s}$	Size-age transition matrix
$C^{l}_{r,y,l,s,f}$	Expected catch-at-length
$W_{r,y,a,s}$	Weight-at-age
$\mathrm{Catch}_{r,y,f}$	Total expected catch biomass
$q^{\mathrm{Fsh}}_{r,y,f}$	Fishery catchability coefficient
$I^{a}_{r,y,a,s,sf}$	Expected survey catch-at-age
$\mathrm{Sel}^{\mathrm{Srv}}_{r,y,a,s,sf}$	Survey selectivity-at-age
$I^{l}_{r,y,l,s,sf}$	Expected survey catch-at-length
$\mathrm{SrvIdx}_{r,y,sf}$	Survey index (abundance or biomass)
$q^{\mathrm{Srv}}_{r,y,sf}$	Survey catchability coefficient

Symbol	Description
$T_{r,y,a,s}^{k}$	Tagged cohort abundance
τ	Initial tag-induced mortality rate
$Z_{r,y,a,s}^{\mathrm{Tag}}$	Tagged cohort total mortality
κ	Chronic tag loss parameter
$\mathrm{Recap}_{r,y,a,s}^{k}$	Tag recaptures
$\beta_{r,y}$	Tag reporting rate
$\mathrm{InitTag}^{k}$	Initial number of tags released
$\mathbf{O}_{\mathrm{Tagging}}^{k}$	Observed tagging proportions
$\mathbf{E}_{\mathrm{Tagging}}^{k}$	Expected tagging proportions
$\epsilon_{r,y}^{\mathrm{Rec}}$	Recruitment deviation
$\epsilon_{r,y,f}^{\mathrm{Fsh}}$	Fishing mortality deviation
$\epsilon_{r,i}^{\mathrm{Init}}$	Initial age deviations
$\lambda_{\mathrm{ObsCatch}_{r,y,f}}$	Likelihood weight for catch data
$\lambda_{\mathrm{ObsFshIdx}_{r,y,f}}$	Likelihood weight for fishery indices
$\lambda_{\mathrm{ObsSrvIdx}_{r,y,sf}}$	Likelihood weight for survey indices
$\lambda_{\mathrm{ObsCompositionData}_{r,y,j}}$	Likelihood weight for composition data
$\lambda_{\mathrm{Tagging}}$	Likelihood weight for tagging data
σ_{Init}	Standard deviation for initial age deviations
σ_{Fsh}	Standard deviation for fishing mortality deviations
$\theta_{r,p,t,s,j}$	Selectivity parameters
$\mu_{r,p,t,s,j}^{\mathrm{Sel}}$	Prior mean selectivity parameter

Appendix 2: Additional Figures

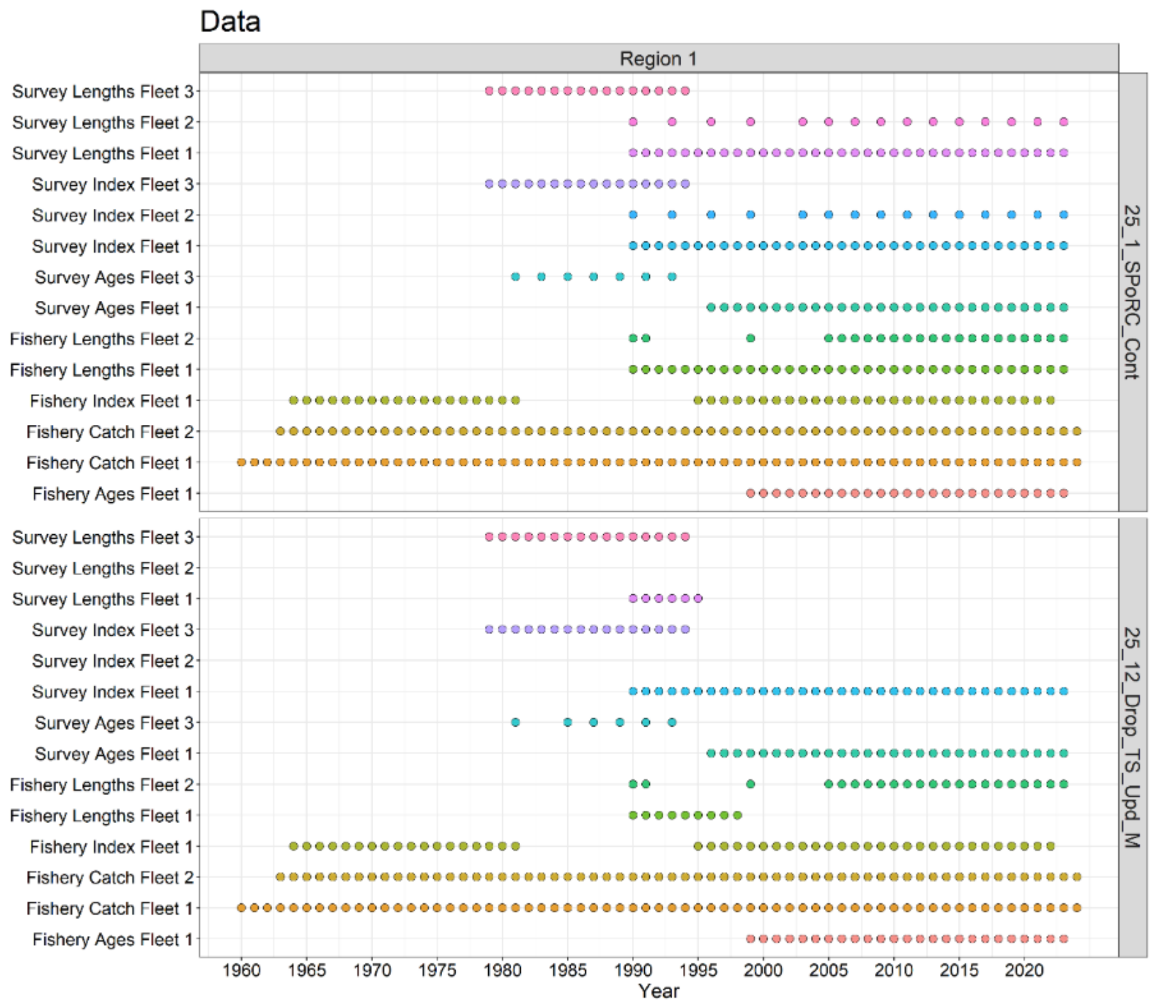


Figure A2.1. Data inputs used by the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*).

Comp_Type	Init_Weight	Final_Weight
LLF_Age_F	0.8261073	0.8120319
LLF_Age_M	NA	NA
LLS_Age_F	3.7922454	3.4281336
LLS_Age_M	NA	NA
JPN_LLS_Age_F	1.3168111	0.6409265
JPN_LLS_Age_M	NA	NA
LLF_Len_F	4.1837057	6.0265426
LLF_Len_M	4.2696935	NA
TF_Len_F	0.3164859	0.2651007
TF_Len_M	0.2293966	NA
LLS_Len_F	1.4379202	2.5287035
LLS_Len_M	1.0705376	NA
JPN_LLS_Len_F	1.2777281	1.0117184
JPN_LLS_Len_M	0.8575195	NA

Figure A2.2. Compositional data weights (λ) after Francis reweighting for model *25.1_SPoRC_Cont* (middle column, labelled 'Init_Weight') and model *25.12_Drop_TS_Upd_M* (last column, labelled 'Final_Weight'). The NA values indicate that no weight was needed because the data set was either sex-aggregated for model *25.1_SPoRC_Cont* or the joint by sex approach was utilized such that a single weight was required (as noted in the female row for each data set) for model *25.12_Drop_TS_Upd_M.* Abbreviations include: LLF—longline (fixed gear) fishery; LLS—longline survey; JPN—Japanese; TF—trawl fishery; TS—trawl survey; len—length; F—female; M—male.

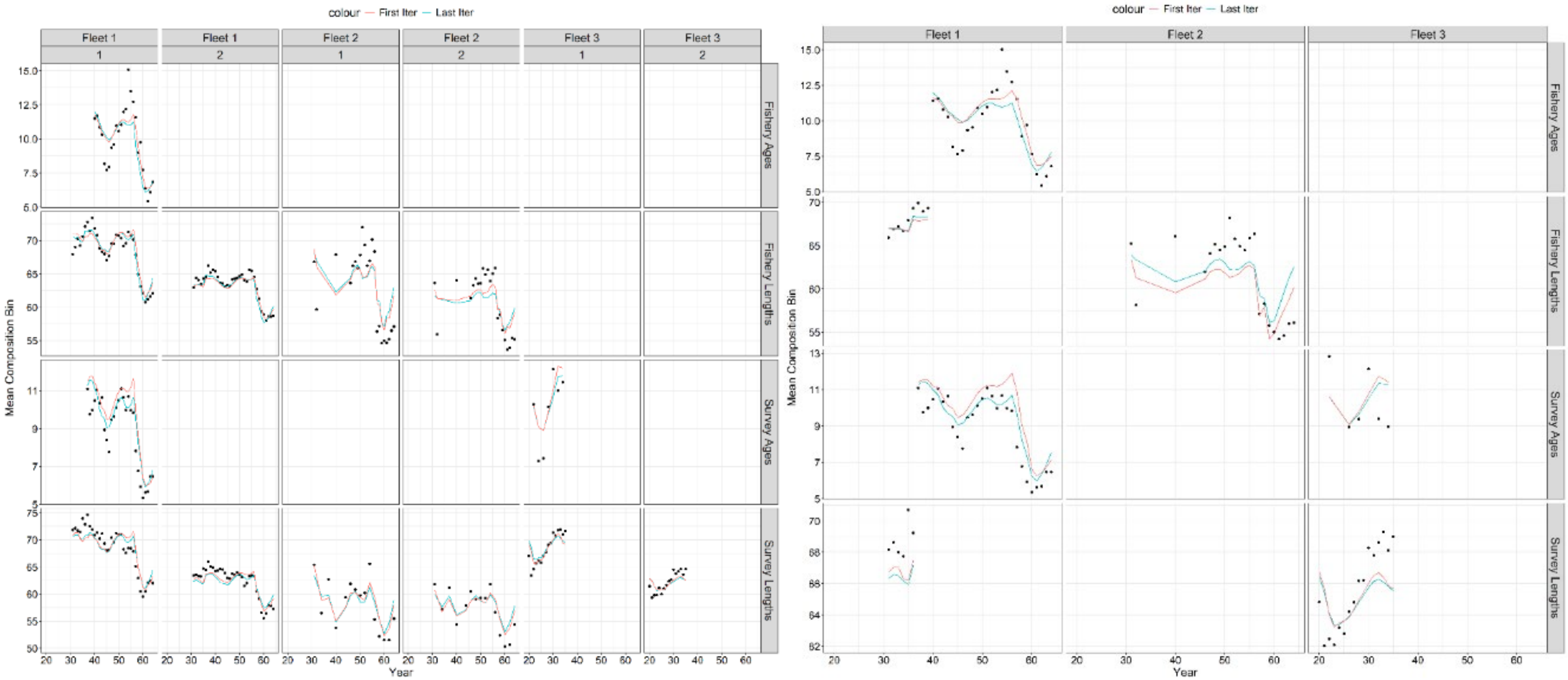


Figure A2.3. Fits to the mean age or mean length by year for each compositional data source before (red line) and after (blue lines) Francis reweighting for mode *25.1_SPoRC_Cont* (left panel) and the author recommended model *25.12_Drop_TS_Upd_M* (right panel). Fishery fleet 1 is the fixed gear fishery and fleet 2 is the trawl fishery. Survey fleet 1 is the NOAA longline survey, fleet 2 is the trawl survey, and fleet 3 is the Japanese longline survey. Sex 1 is female and sex 2 is male.

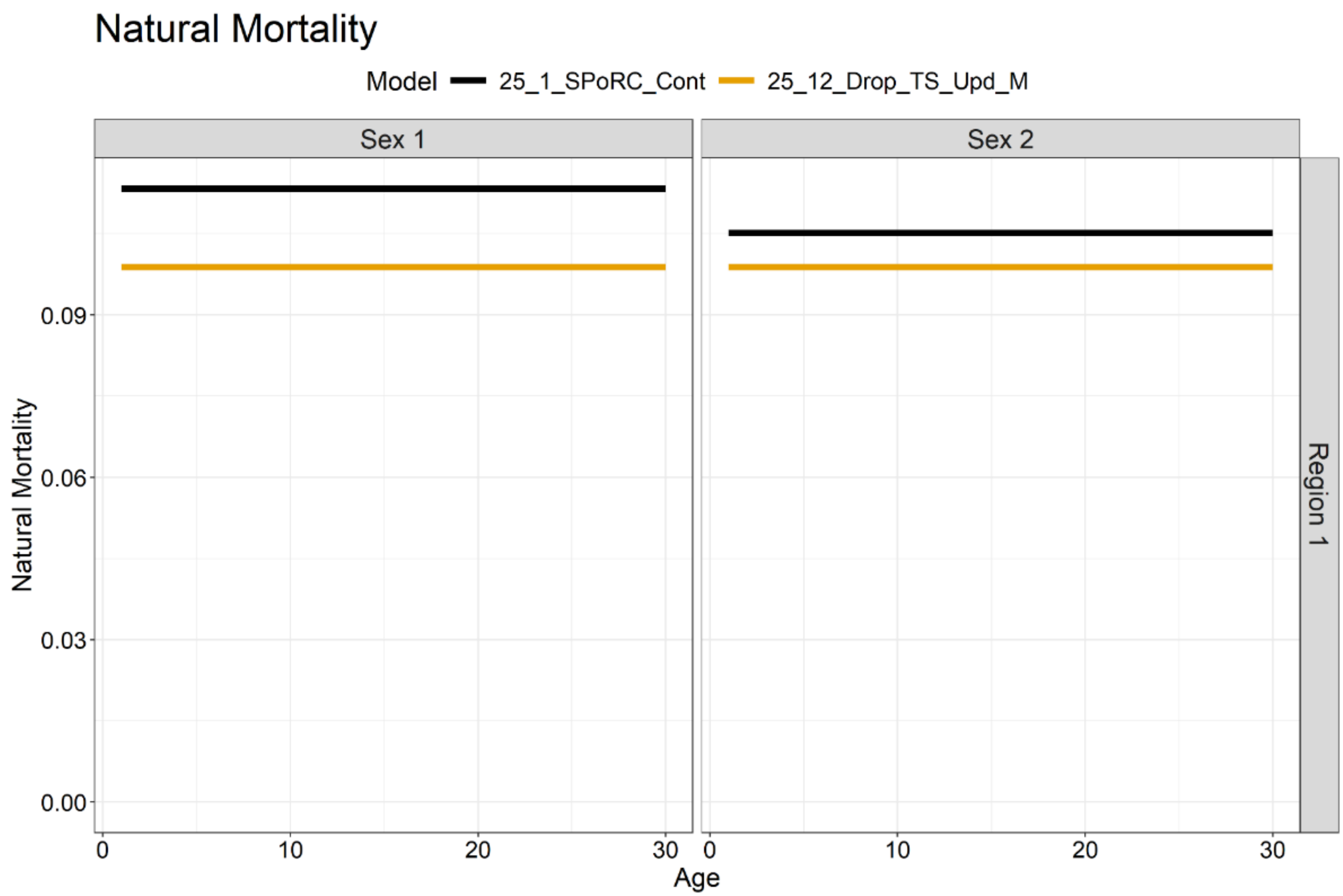


Figure A2.4. Natural mortality estimates from the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*). Note that the male natural mortality offset value in models *25_1_SPoRC_Cont* was not estimated and was inadvertently fixed less than female natural mortality. Model *25_12_Drop_TS_Upd_M* corrects this issue, and the natural mortality values shown for both sexes are equivalent (i.e., sex-invariant).

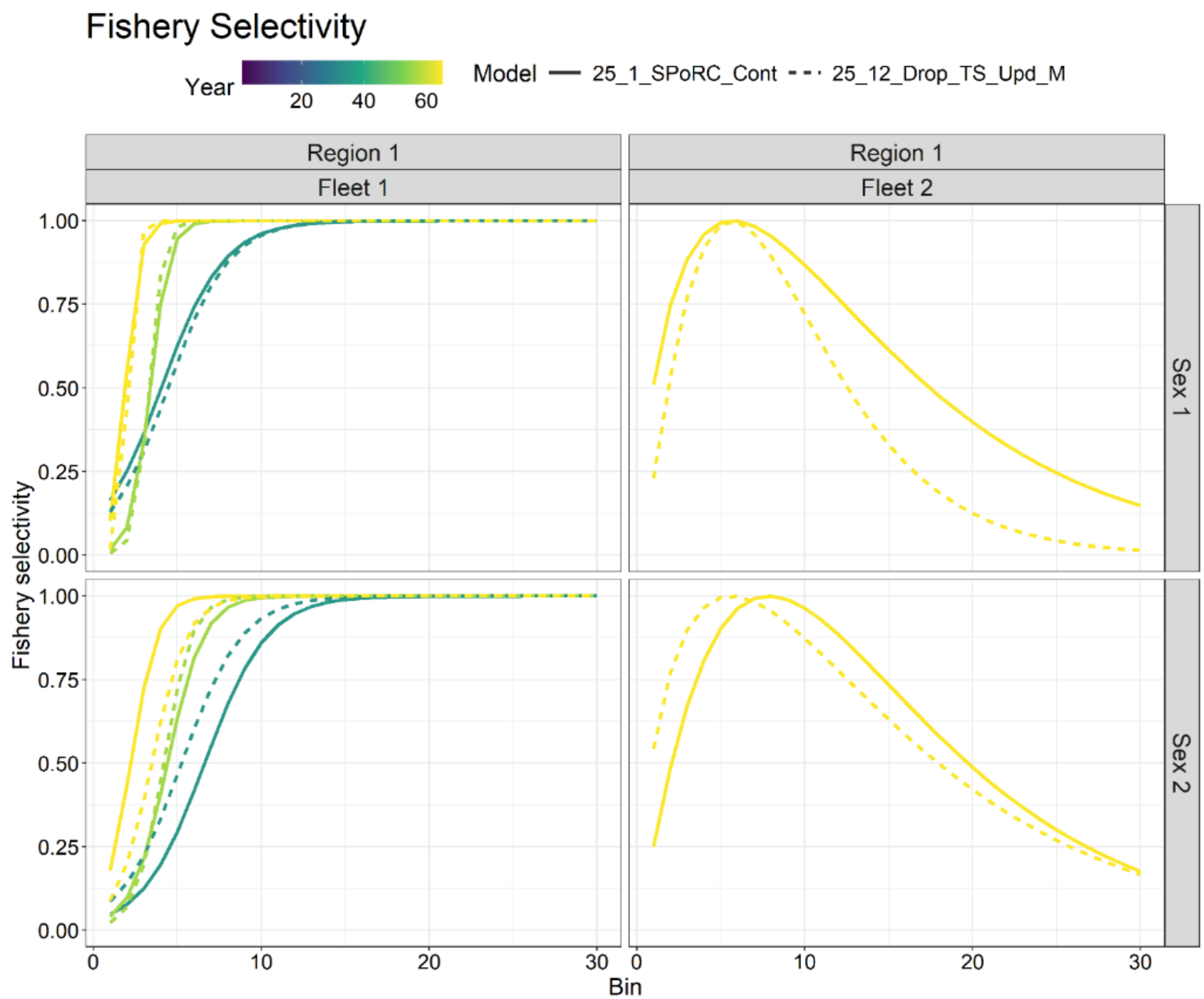


Figure A2.5. Fishery selectivity estimates from the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*), where different colors represent the different time blocks.

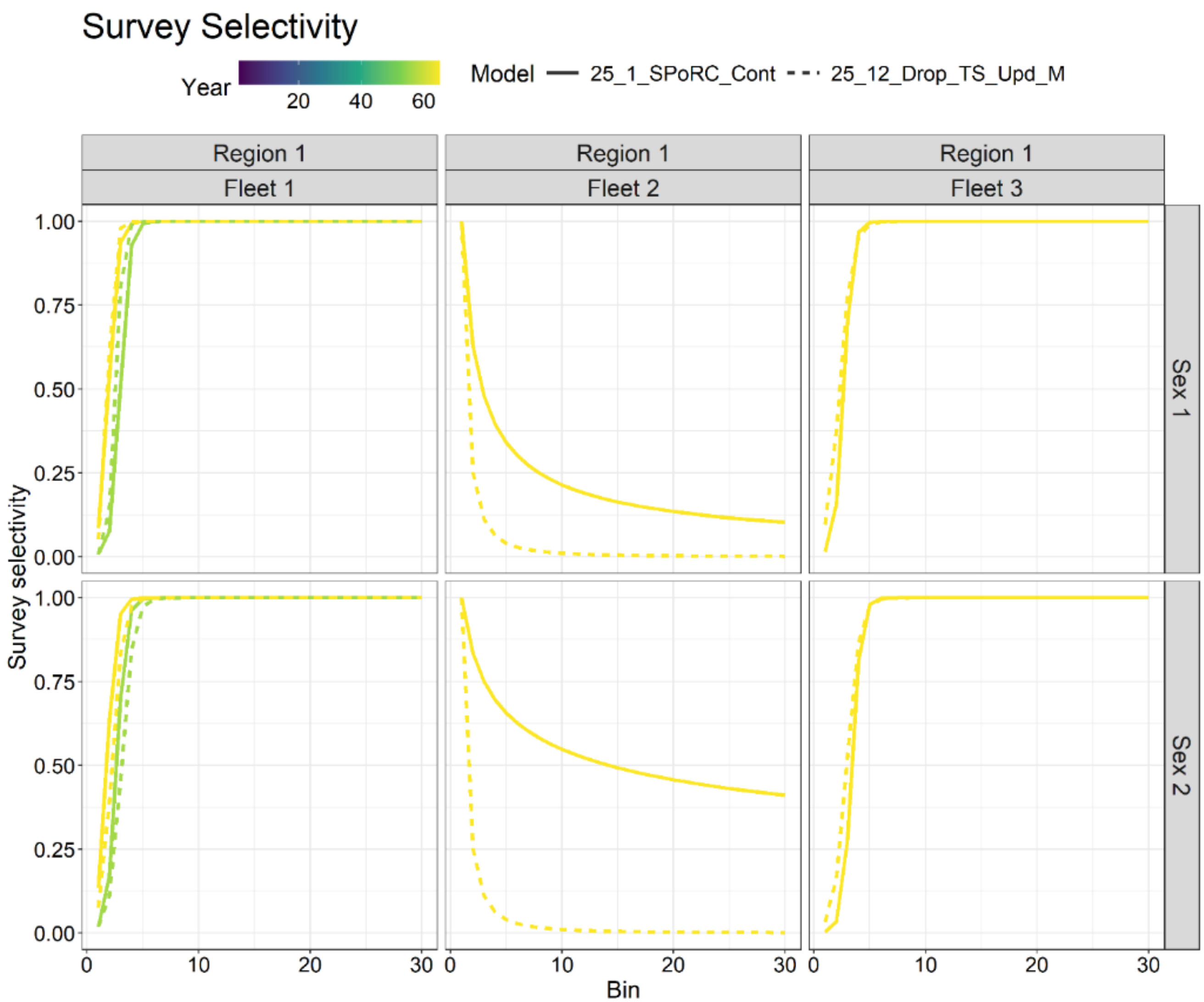


Figure A2.6. Survey selectivity estimates from the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*), where different colors represent the different time blocks. Note that the trawl survey is not used in the author recommended model and these selectivity parameters are not estimated.

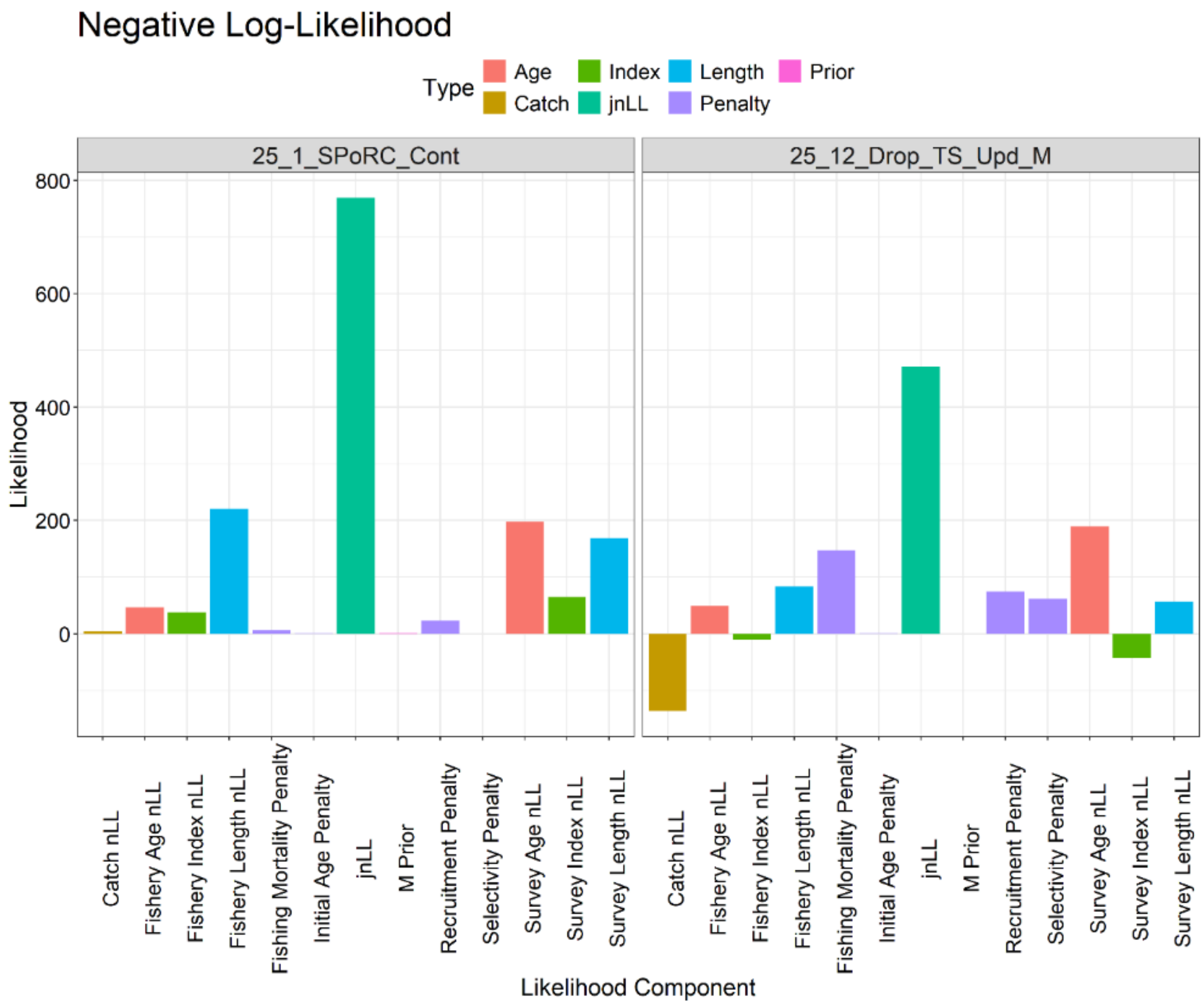


Figure A2.7. Joint negative log-likelihood (jnLL) components from the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*), where different colors represent the data sets. Note that due to differences in data and priors or penalties used in each model, these figures are not directly comparable across models.

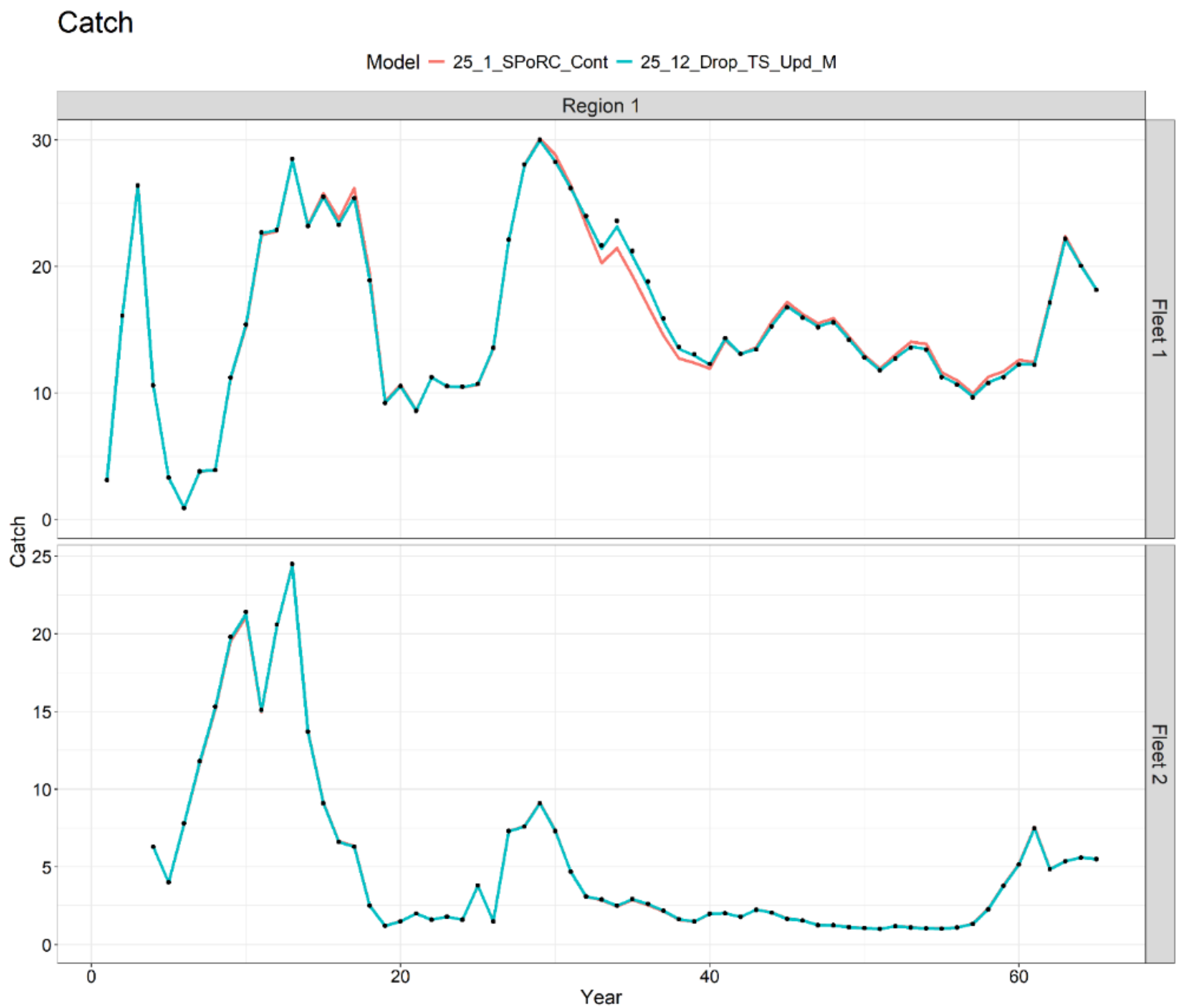


Figure A2.8. Fit to the catch data by fleet (fleet 1 is the fixed gear fishery and fleet 2 is the trawl fleet) for the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*), where different colors represent the different time blocks.

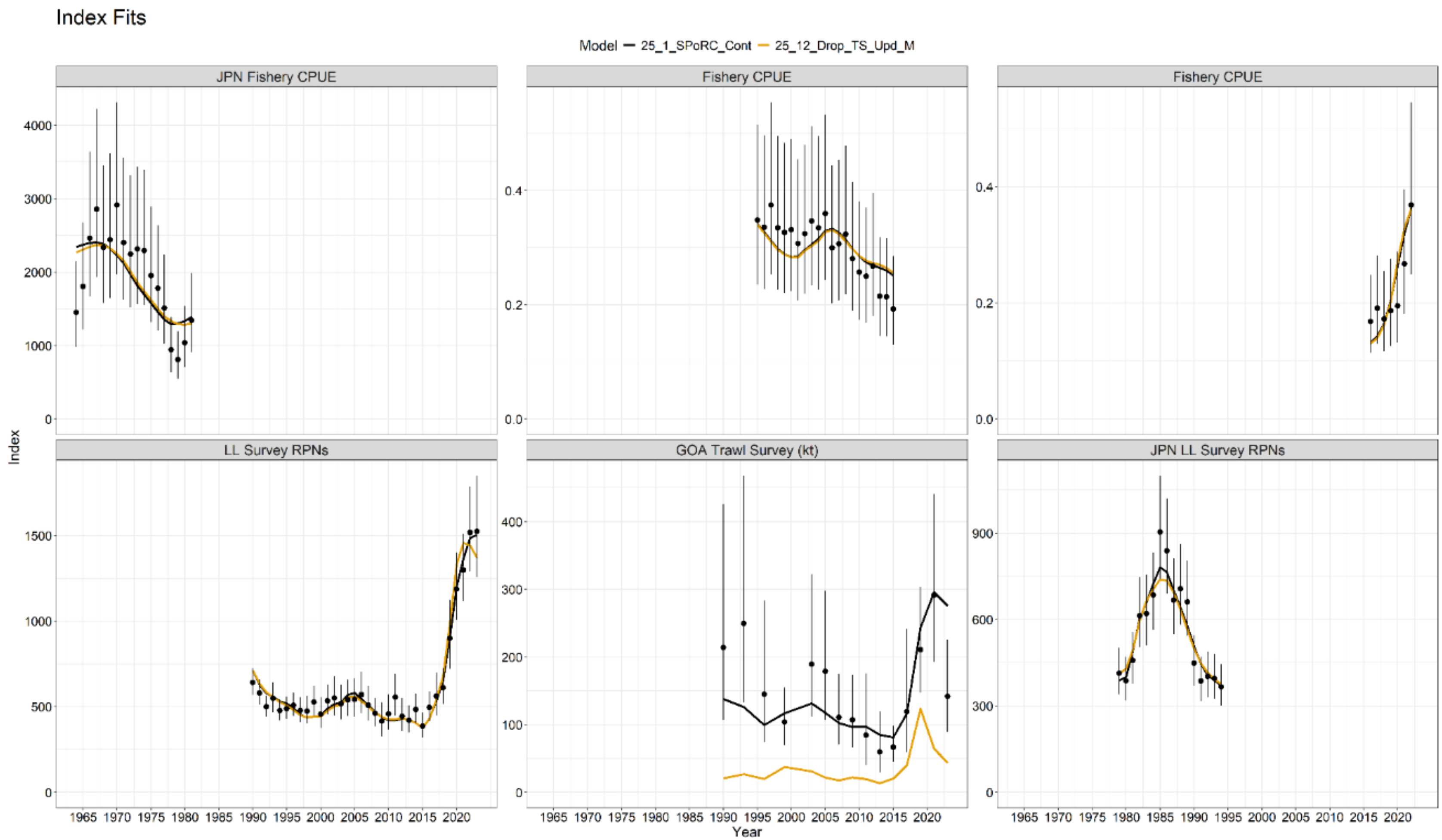


Figure A2.9. Fit to the fishery (top row) and survey (bottom panel) indices for the continuity model (*25.1_SPoRC_Cont*) and the author recommended model (*25.12_Drop_TS_Upd_M*). Note that the trawl survey is not fit in the author recommended model.

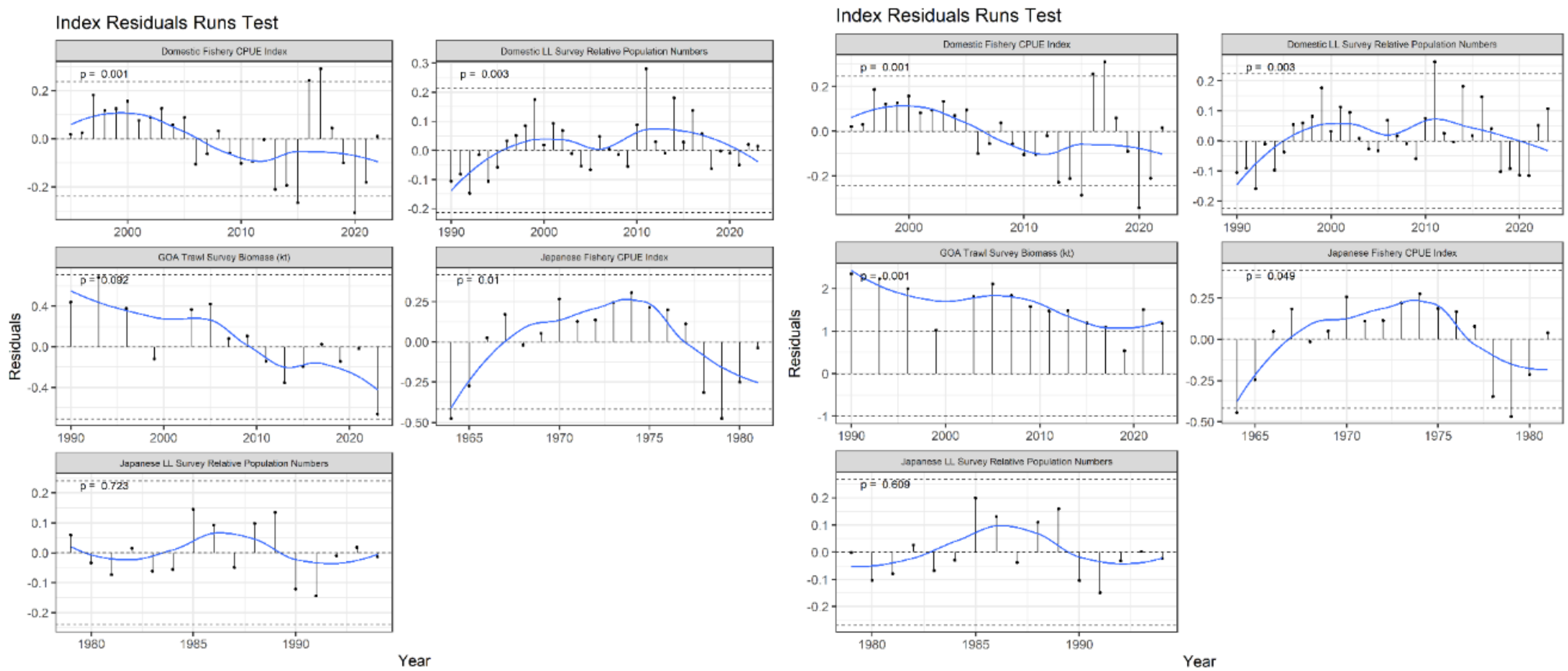


Figure A2.10. Results of index residuals runs test for the continuity model (*25.1_SPoRC_Cont*; left panel) and the author recommended model (*25.12_Drop_TS_Upd_M*; right panel).

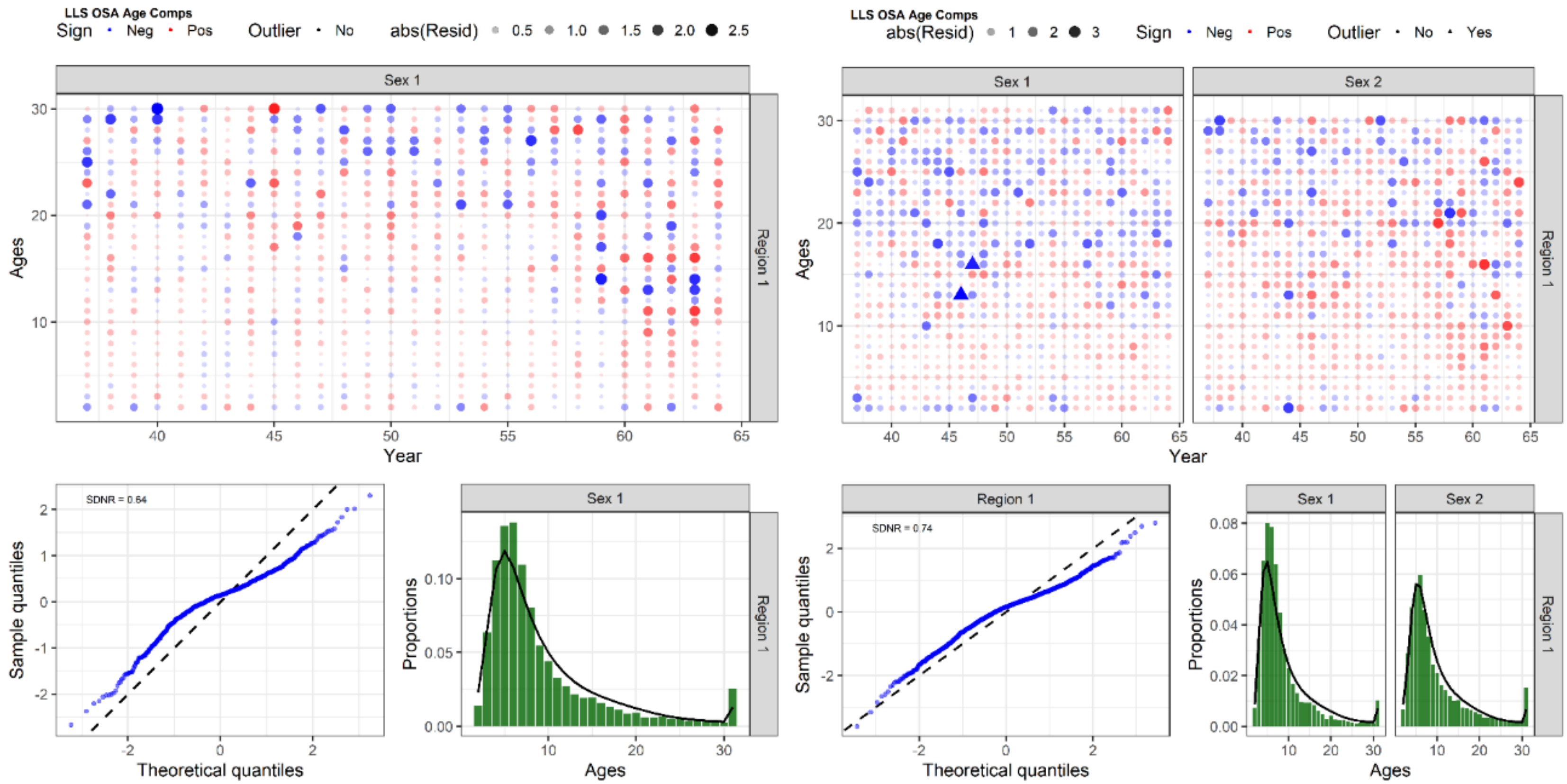


Figure A2.11. Fits to NOAA longline survey age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

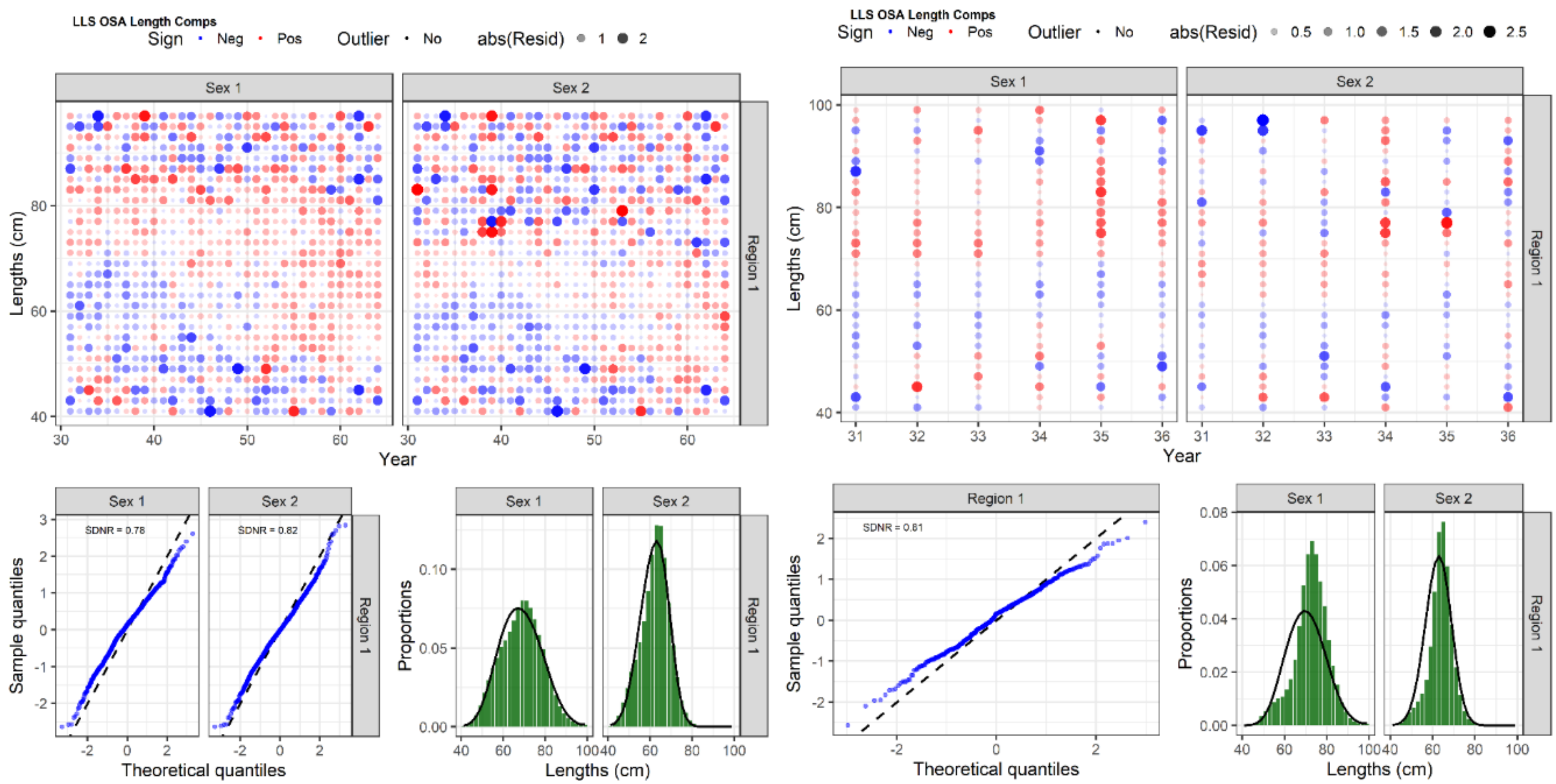


Figure A2.12. Fits to NOAA longline survey length composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

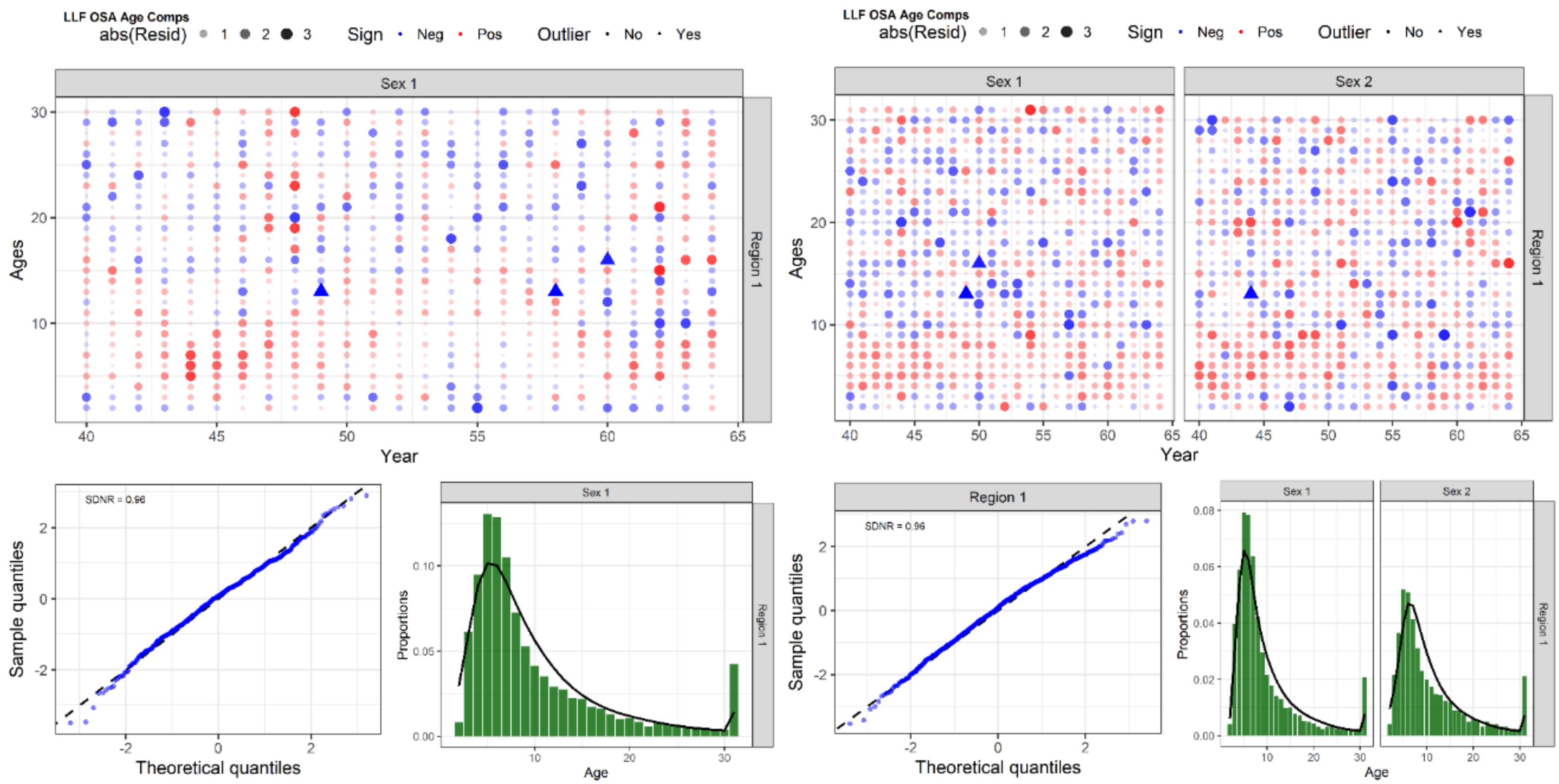


Figure A2.13. Fits to fixed gear fishery age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

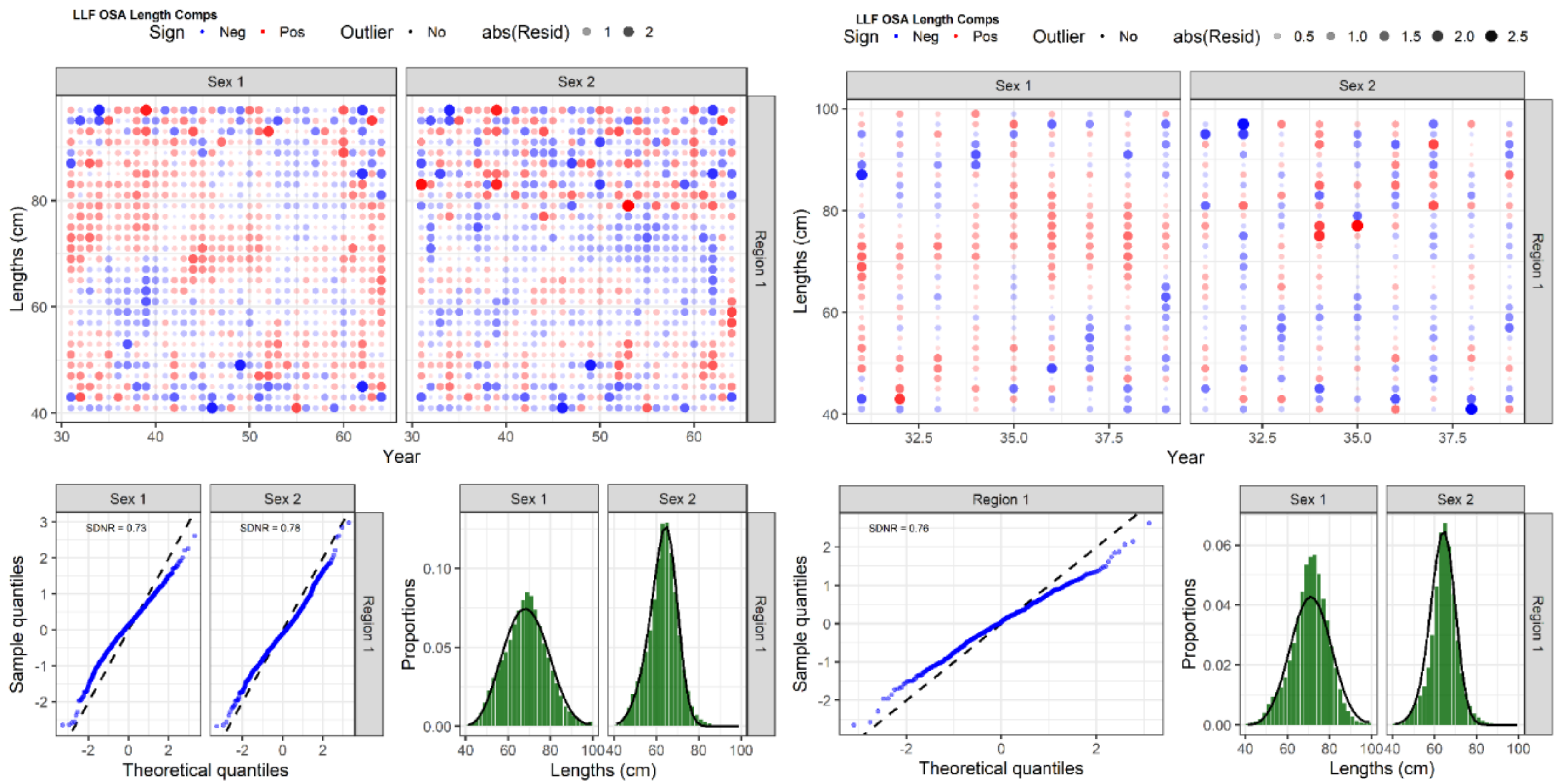


Figure A2.14. Fits to fixed gear fishery length composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

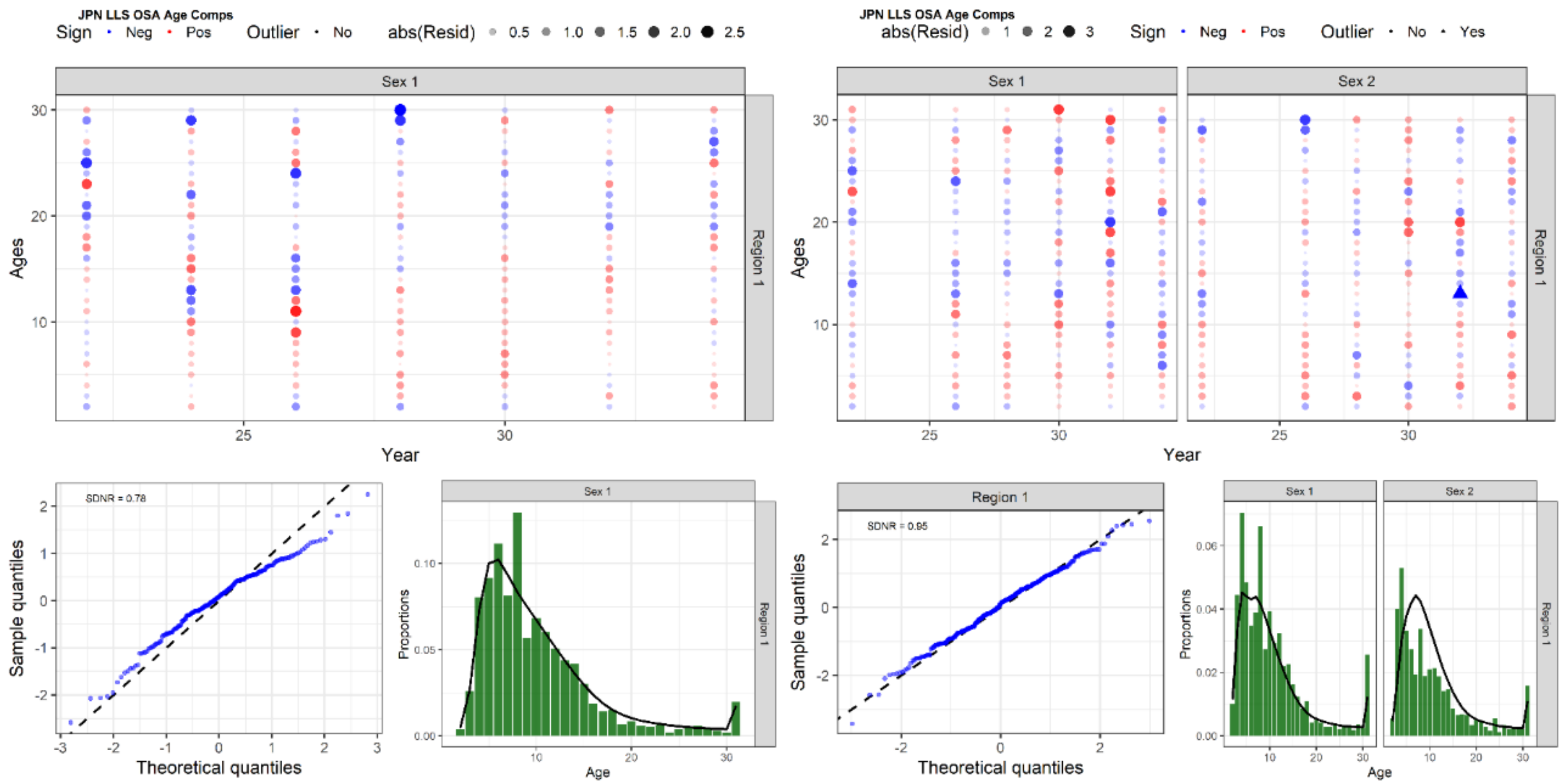


Figure A2.15. Fits to Japanese longline survey age composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

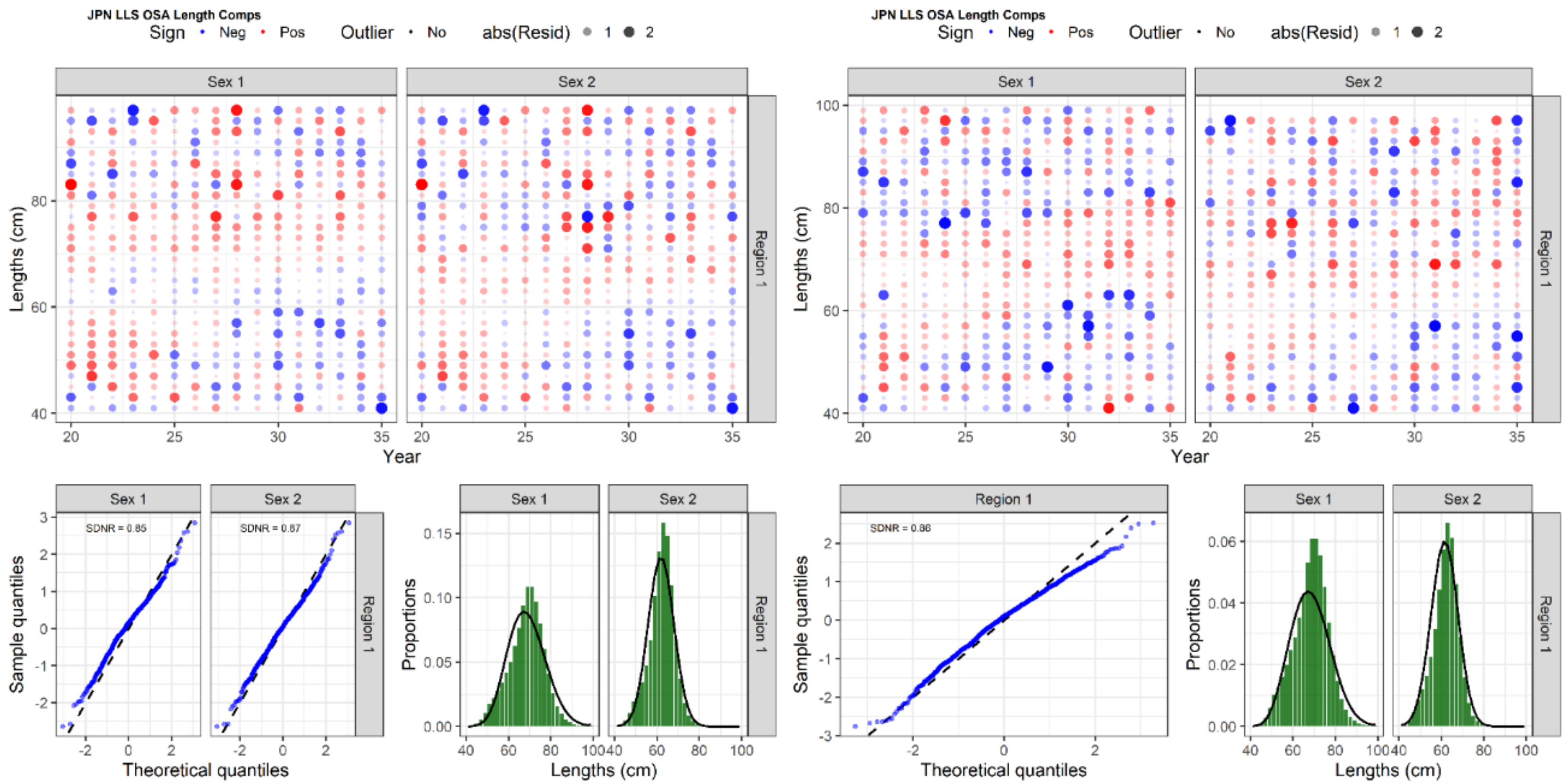


Figure A2.16. Fits to Japanese longline survey length composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

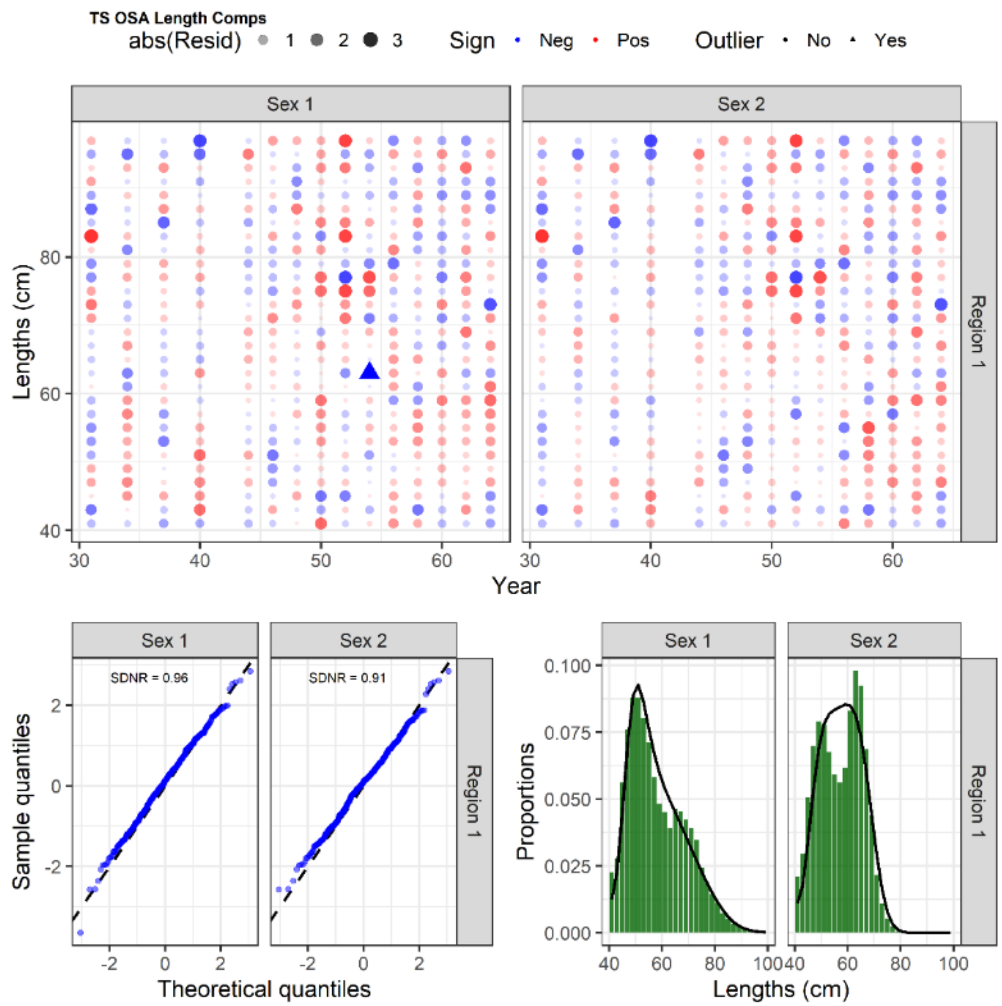


Figure A2.17. Fits to NOAA trawl survey length composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel). The trawl survey is not fit in model *25.12_Drop_TS_Upd_M*.

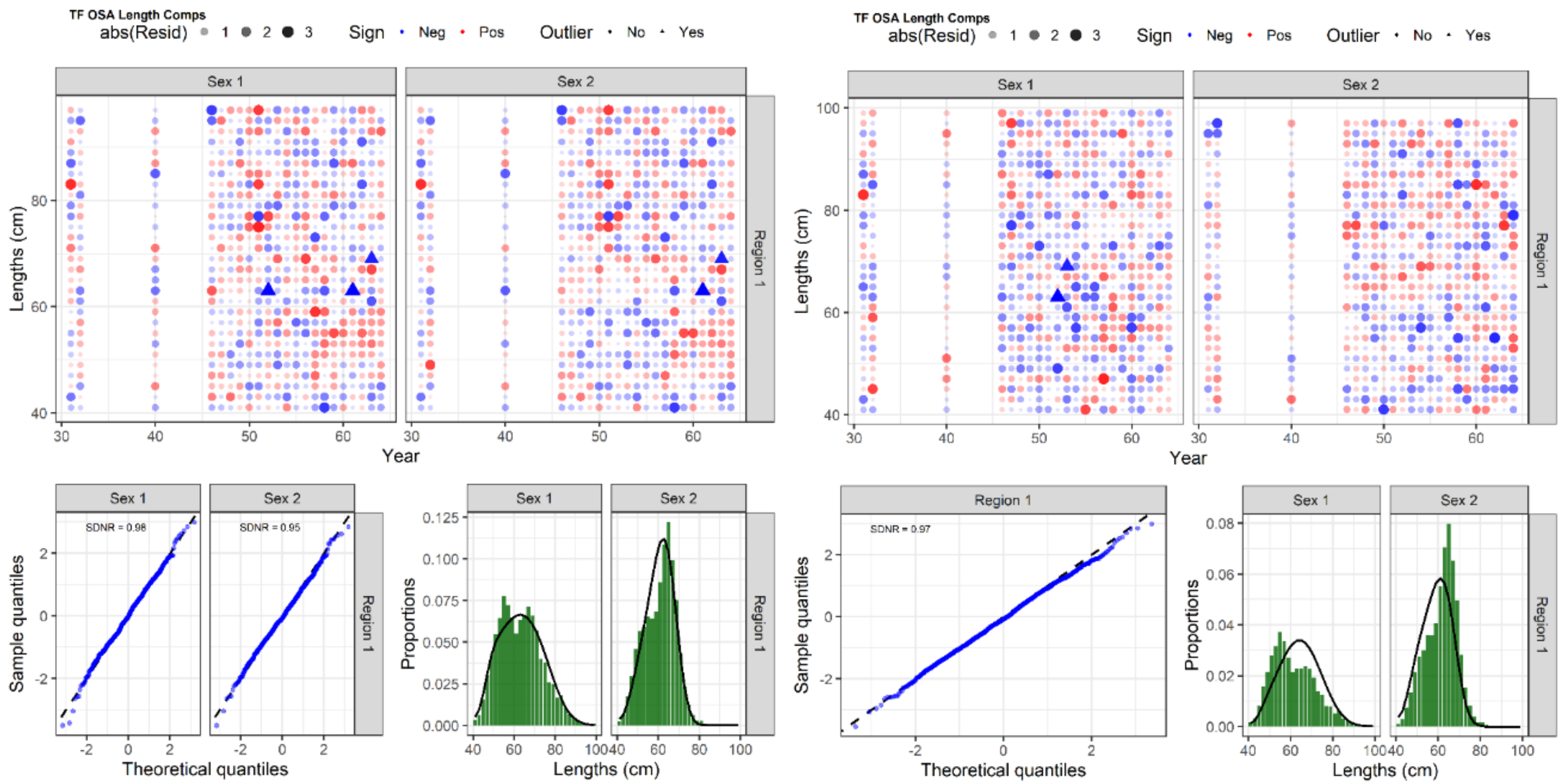


Figure A2.18. Fits to trawl fishery length composition data, including one step ahead (OSA) residuals (top panel), quantile residuals (bottom left panel), and aggregate (across all years) fit to the age compositions (bottom right panel; green bars are observed proportions and black lines are expected model estimated proportions), for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

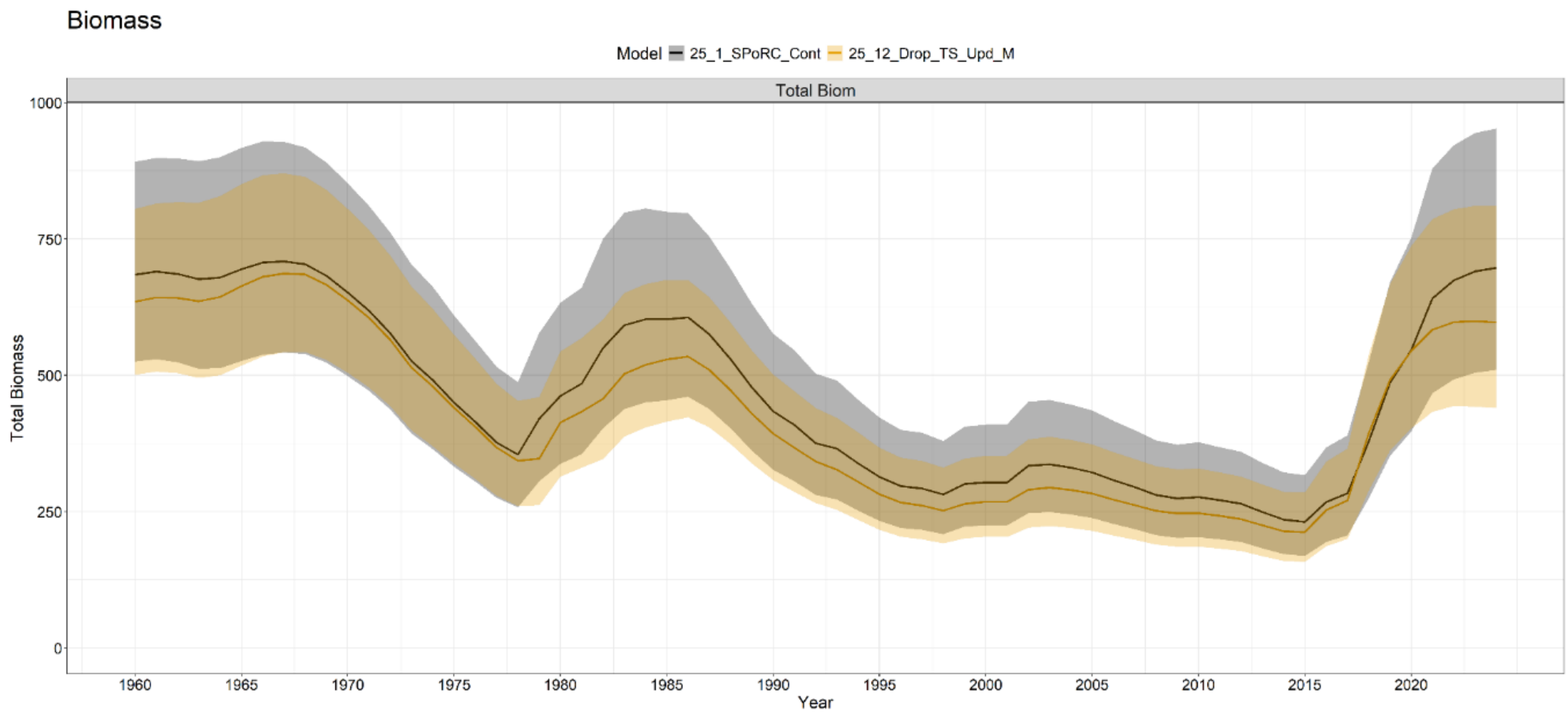


Figure A2.19. Time series estimates of total biomass (kt) for model *25.1_SPoRC_Cont* and model *25.12_Drop_TS_Upd_M*.

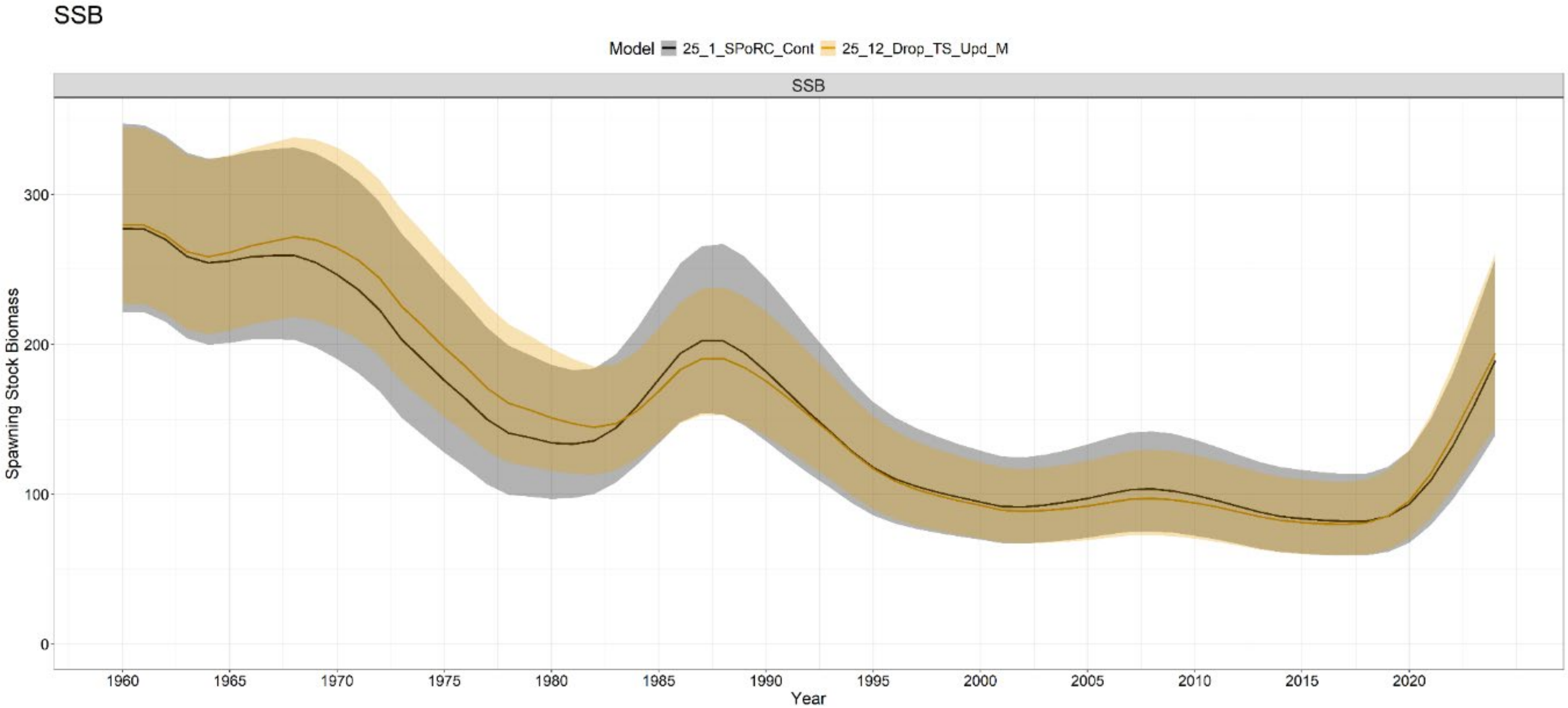


Figure A2.20. Time series estimates of spawning stock biomass (kt) for model *25.1_SPoRC_Cont* and model *25.12_Drop_TS_Upd_M*.

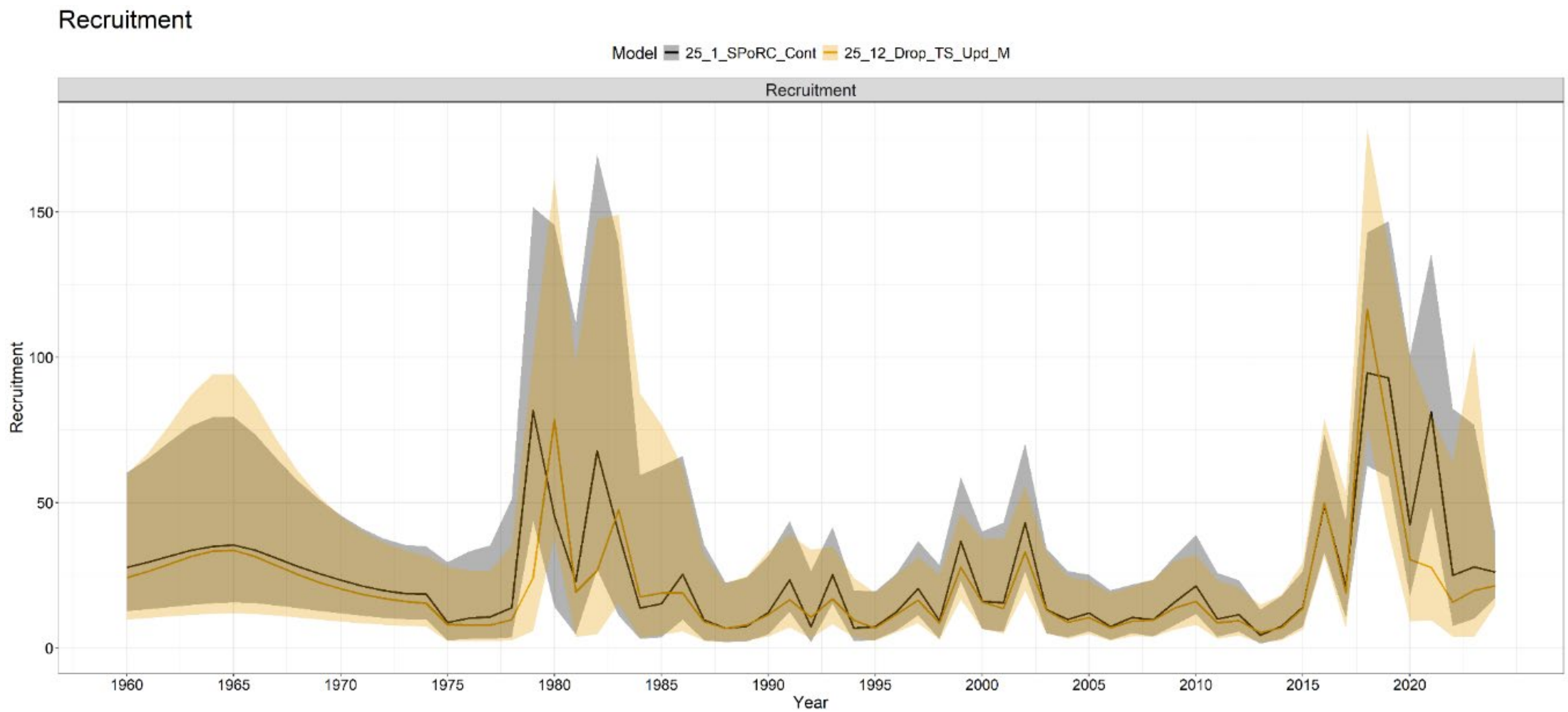


Figure A2.21. Time series estimates of recruitment (millions of fish) for model *25.1_SPoRC_Cont* and model *25.12_Drop_TS_Upd_M*.

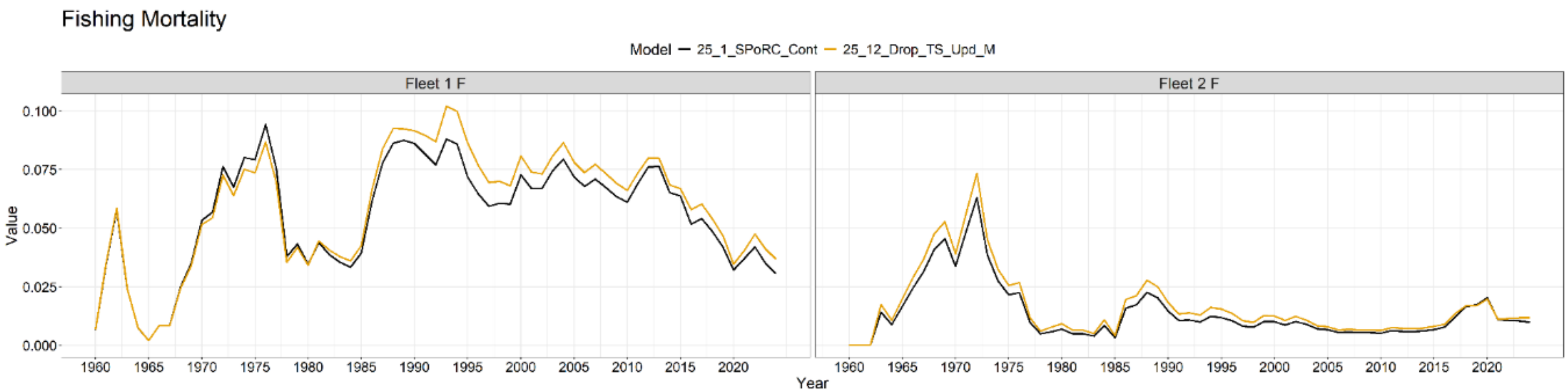


Figure A2.22. Time series estimates of fishing mortality (yr^{-1}) for model *25.1_SPoRC_Cont* and model *25.12_Drop_TS_Upd_M*. Fleet 1 is the fixed gear fishery and fleet 2 is the trawl fishery.

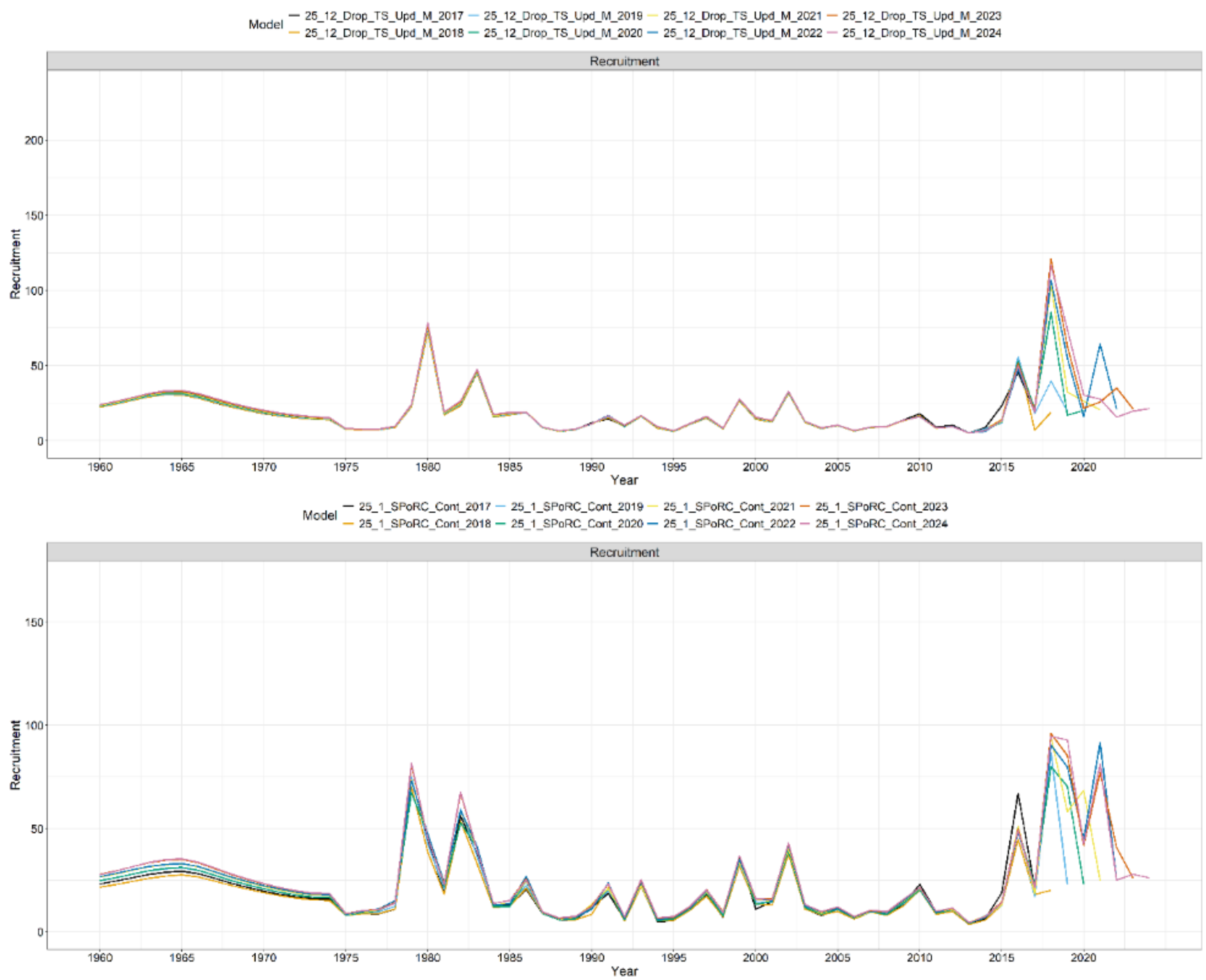


Figure A2.23. Comparison of the estimated recruitment (in millions of fish) time series from a seven year 'true' retrospective analysis for model *25.12_Drop_TS_Upd_M* (top panel) and *25.1_SPoRC_Cont* (bottom panel). The 'true' retrospective analysis removes a year of data for each peel, but also accounts for the lag in availability of compositional data (i.e., fishery ages and lengths as well as longline survey ages become available on a one year lag from associated catch or index data).

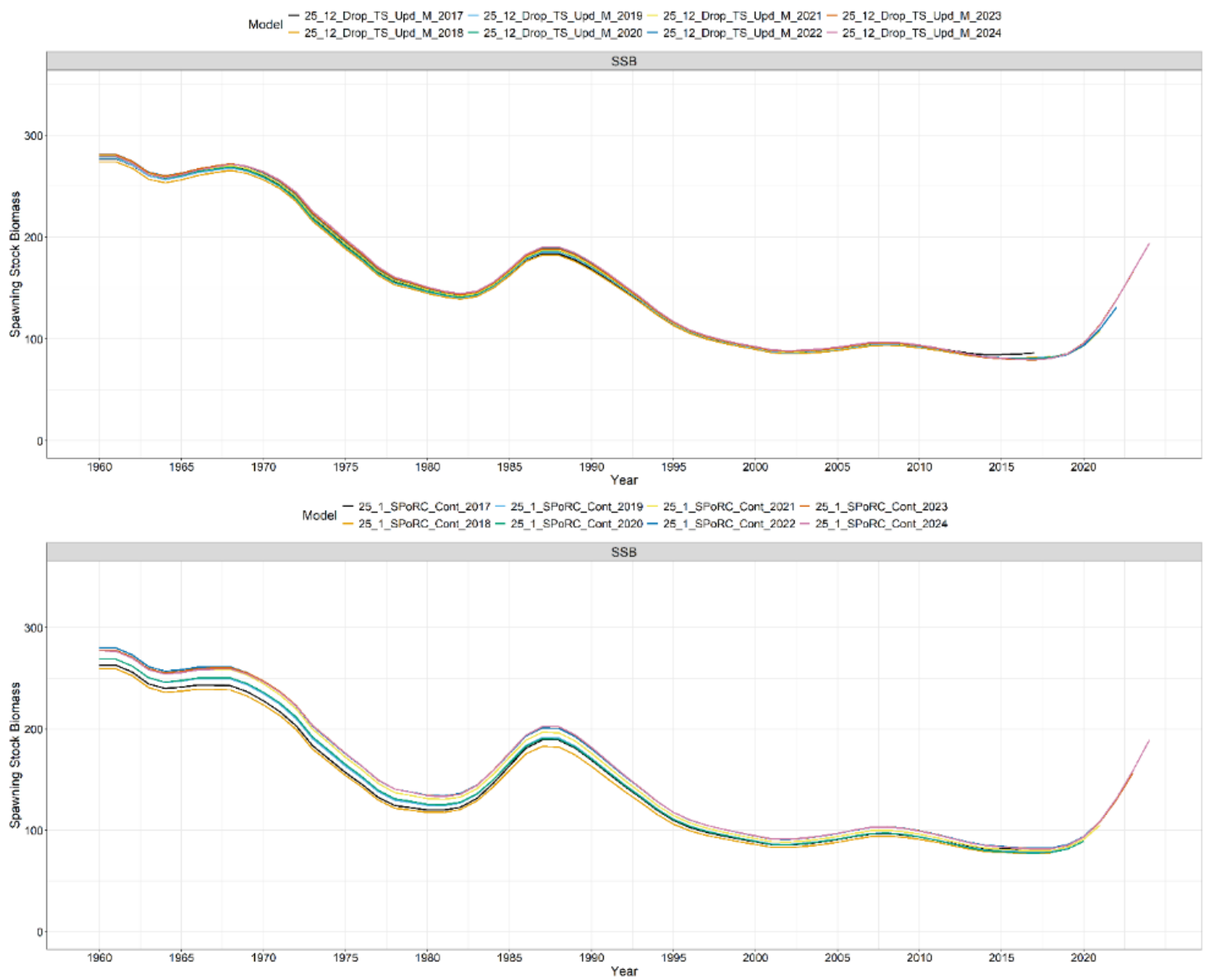


Figure A2.24. Comparison of the estimated spawning stock biomass (kt) time series from a seven year 'true' retrospective analysis for model *25.12_Drop_TS_Upd_M* (top panel) and *25.1_SPoRC_Cont* (bottom panel). The 'true' retrospective analysis removes a year of data for each peel, but also accounts for the lag in availability of compositional data (i.e., fishery ages and lengths as well as longline survey ages become available on a one year lag from associated catch or index data).

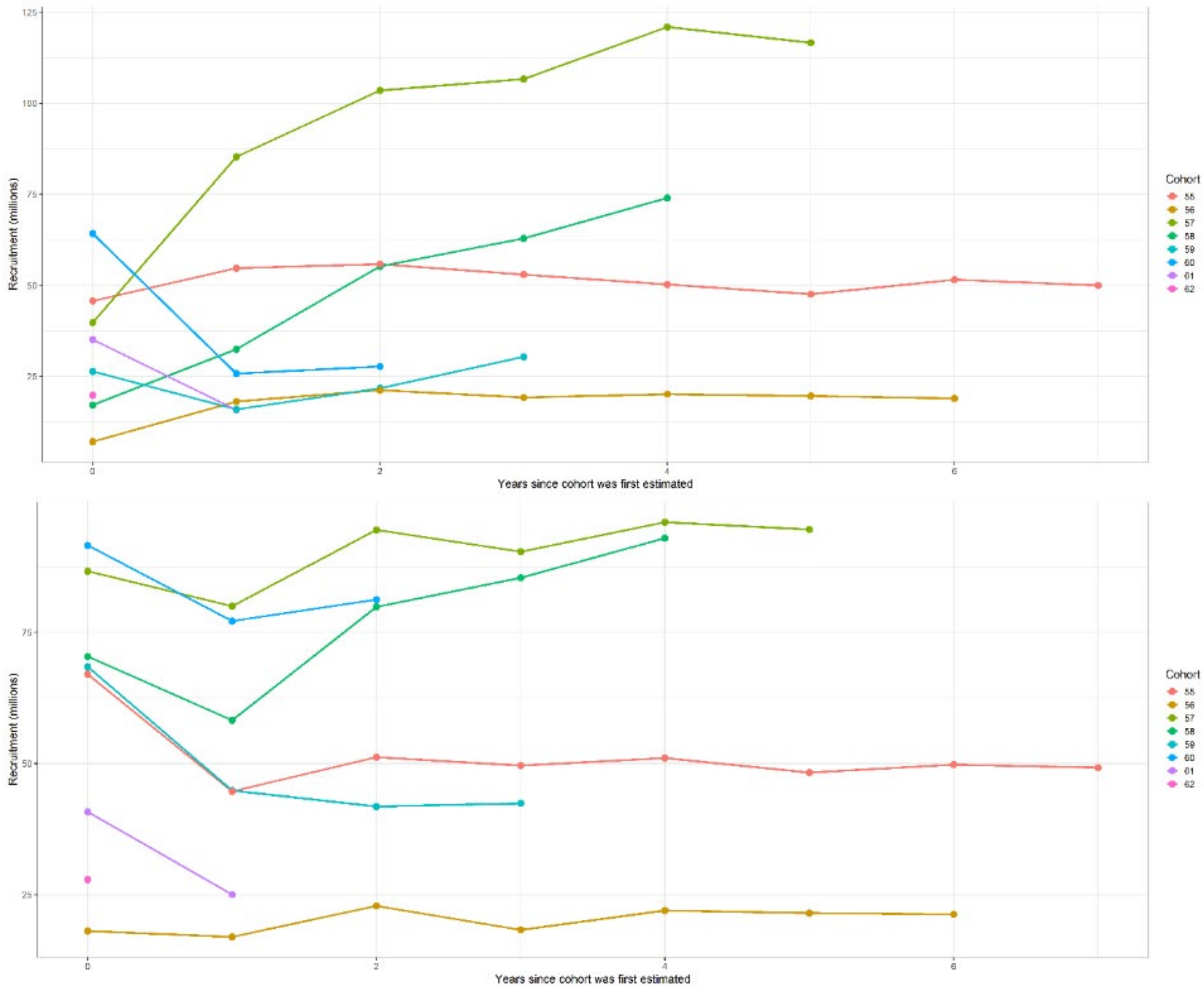


Figure A2.25. Comparison of a recruitment retrospective 'squid' plot representing how estimates of a given year class change as data are added in a seven year 'true' retrospective analysis for model *25.12_Drop_TS_Upd_M* (top panel) and *25.1_SPoRC_Cont* (bottom panel). The 'true' retrospective analysis removes a year of data for each peel, but also accounts for the lag in availability of compositional data (i.e., fishery ages and lengths as well as longline survey ages become available on a one year lag from associated catch or index data).

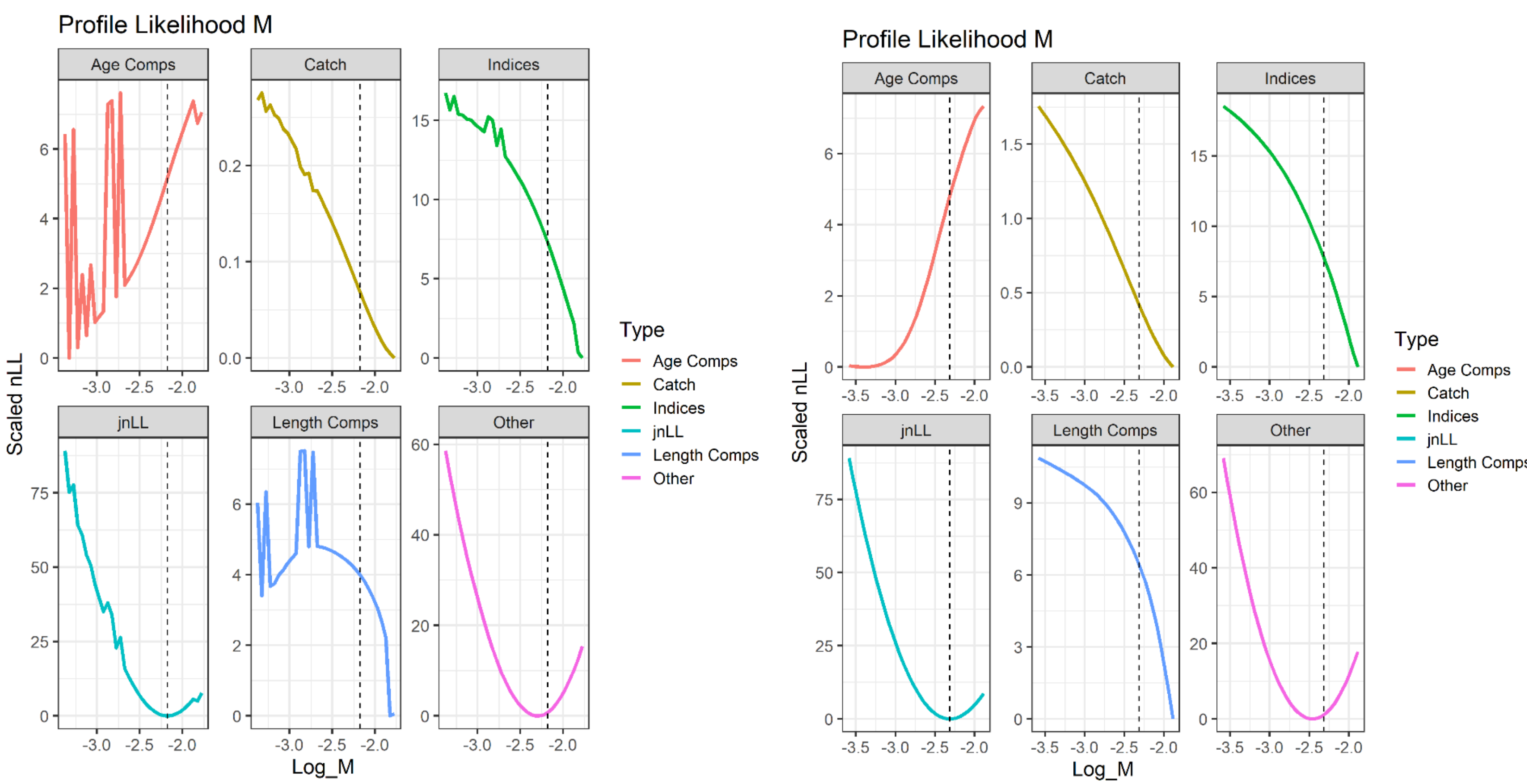


Figure A2.26. Likelihood profile for the natural mortality parameter (*M*), where panels represent the profile for each data type used in the model, for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

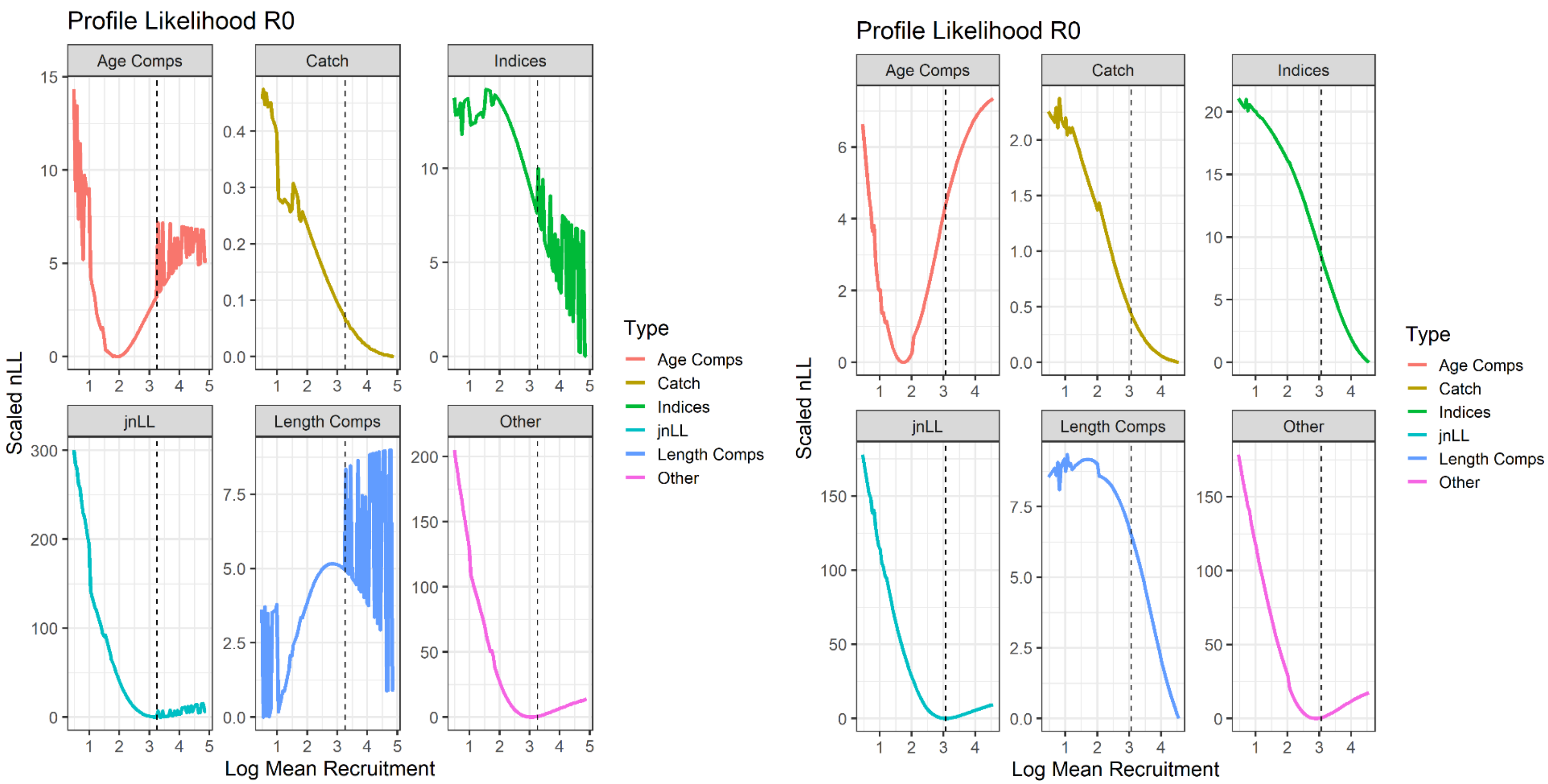


Figure A2.27. Likelihood profile for the mean recruitment (R_0), where panels represent the profile for each data type used in the model, for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel).

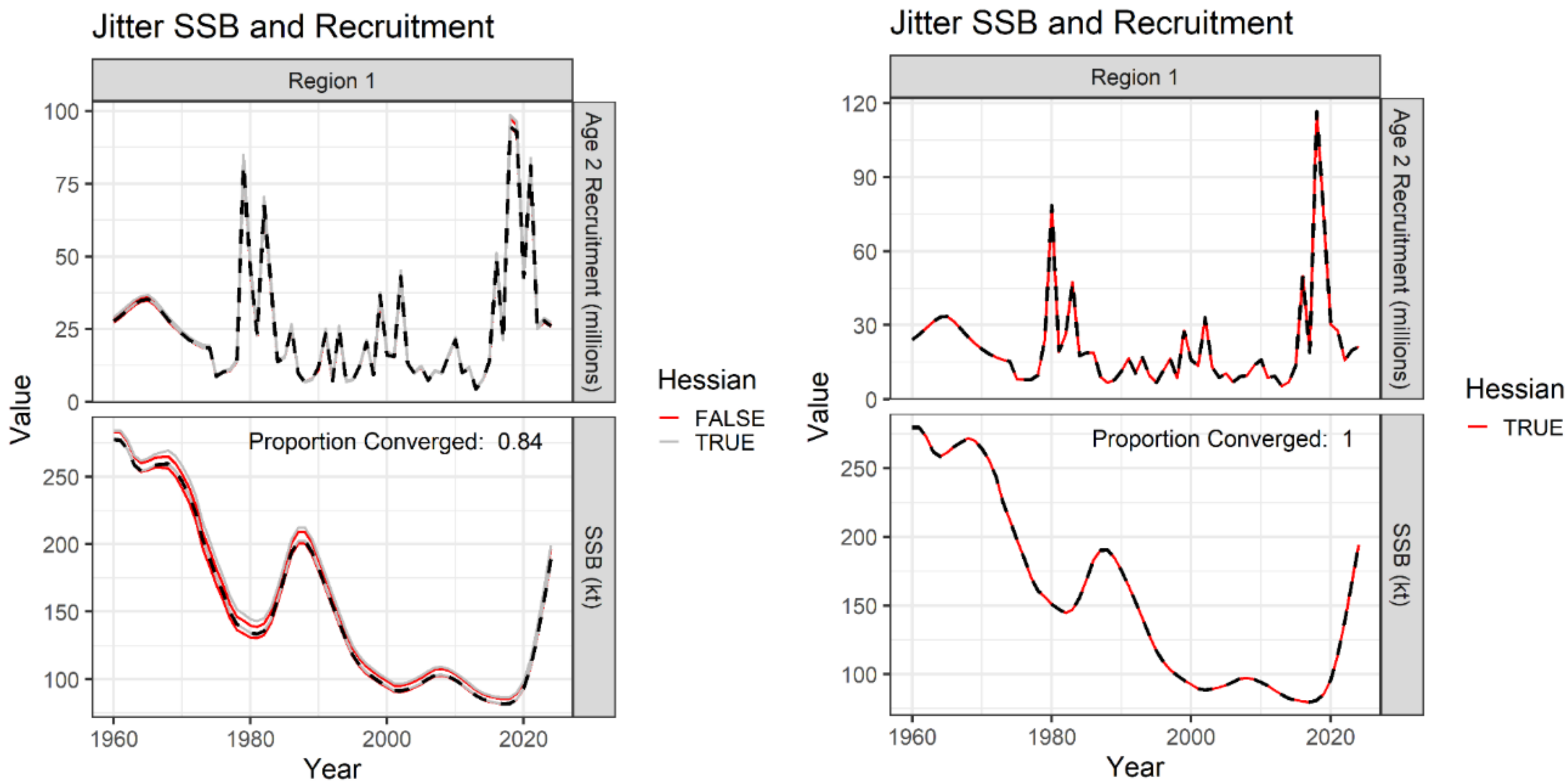
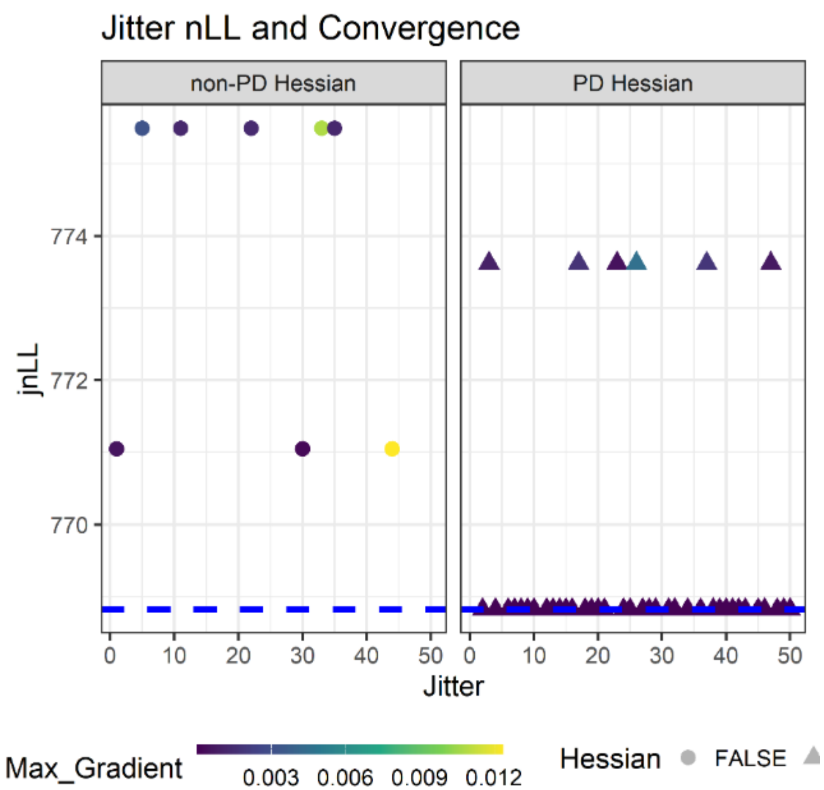


Figure A2.28. Results of a jitter analysis showing recruitment (top panel) and SSB (bottom panel) for each jitter iteration for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel). The starting parameter values were jittered using a ~$N(0,0.1^2)$.



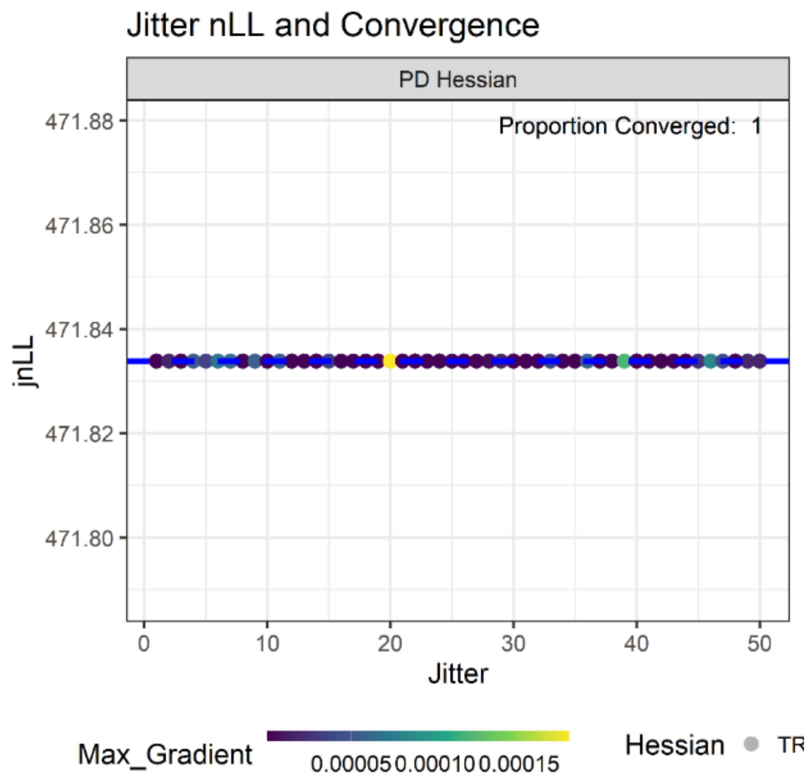


Figure A2.29. Results of a jitter analysis showing the joint negative log-likelihood values for converged (Positive Definite, PD, Hessian) and non-converged runs (non-Positive Definite Hessian) for model *25.1_SPoRC_Cont* (left panel) and model *25.12_Drop_TS_Upd_M* (right panel). The starting parameter values were jittered using a ~$N(0,0.1^2)$.