Bering Sea and Aleutian Islands Yellowfin Sole Stock Assessment Development for September 2025

Authors: Meaghan Bryan and Ingrid Spies

Model names

Model 23.0: This is the accepted model from 2024, a bespoke model ('flatfish model') written in ADMB. This model has been used to assess yellowfin sole and Northern rock sole in the Bering Sea and Aleutian Islands since approximately 2016 (https://www.fisheries.noaa.gov/alaska/population-assessments/north-pacific-groundfish-stock-assessments-and-fishery-evaluation). Data weighting for survey age composition is based on the number of hauls from which otoliths were taken.

mSS3deter: This is the deterministic version of Model 23.0 in SS3 v3.30.23.2, with parameters fixed at values from Model 23.0.

Model 25.0: This is the estimation version of mSS3deter, with the following features:

- EBS survey catchability *q* incorporates the survey start date (but not EBS bottom temperature) as an environmental covariate.
- Data weighting for survey age composition is based on the number of hauls from which otoliths were collected and used for constructing age compositions, as in Model 23.0.

Model 25.0a: This is the proposed model for 2025 which builds on Model 25.0, except:

- Data weighting for survey age composition is based on the input sample size (ISS) methodology R package *sampleISS* (Hulson et al. 2024).

Introduction

In this document we present several modifications to the 2024 yellowfin sole assessment Model 23.0 (Spies et al. 2024) to make improvements and meet SSC requests. We explored the following modifications:

1. Consideration of Tier level: Investigating Tier 1 vs. Tier 3 for estimating recruitment.
2. Time-varying *q* should use survey start date, but not temperature.
3. Bridging the bespoke model to SS3 v3.30.23.2 (Methot et al. 2020).

Models will be presented in the following order. First, the deterministic model mSS3deter will be compared with Model 23.0 for the bridging analysis. Then we introduce estimation models Model 25.0 and Model 25.0a. For the purpose of this document, data through 2024 are used in all models. All models presented in this document can be found here: https://github.com/mdbryan/yfs_ss3_2025.

Relevant SSC comments

SSC December 2024
The SSC supports the authors' plan to bridge to a Stock Synthesis model for the next assessment.

Authors' response: We have completed a bridge from the bespoke model 23.0 to SS3 v3.30.23.2 that provides a nearly exact match, and discrepancies are explained below. We are presenting here the bridged model, in anticipation of a full model in November.

SSC December 2024
The SSC encourages evaluation of the information content of the stock-recruitment relationships for BSAI yellowfin sole … around the use of Tier 1 versus Tier 3 harvest specifications …[similar to] EBS pollock.

Authors' response:
We have evaluated the benefits of retaining a Tier 1 vs. Tier 3 level and consider Tier 3 to be preferable based on the information content in the data that informs recruitment subsequent to the 1977 regime change.

1. Consideration of Tier level

The North Pacific Fishery Management Council (NPFMC) uses a tiered system to categorize and manage fish stocks. Tier 1 management makes assumptions about stock recruitment relationship (SRR) providing a reliable pdf of F_{MSY}, while Tier 3 management models are not reliant on theoretical stock recruitment relationship, and instead estimate recruitment using a normally distribution with estimated mean and variance. For Tier 3 where B_{MSY} and a pdf of F_{MSY} cannot be reliably estimated, proxies of $B_{35\%}$ and $F_{35\%}$ are used.

Effectively estimating a stock recruitment relationship requires data from periods of low and high spawning stock biomass and their associated recruitment values. Without sufficient contrast in the data, accurate estimation of the stock-recruitment curve is not possible, leading to uncertainty in the parameter estimates and the management advice derived from them (Zhou 2007, He and Field 2019). In particular, years with low stock sizes are important to detect whether compensatory density dependence is present, in which reproductive rates increase at low population densities (Rose et al. 2001). The opposite may also be the case, known as the allee effect, or depensatory density dependence, in which reproduction decreases at low densities (Rose and Cowan 2000).

In 2024 the EBS pollock assessment moved from a Tier 1 to a Tier 3 approach, primarily driven by uncertainty in recruitment curve parameters and lack of information in the data. We have evaluated whether the yellowfin sole assessment presents a similar case. It is common practice to evaluate recruitment within but not among thermal regimes (Wooster and Zhang 2004). Therefore, the first year of data used to estimate recruitment is 1978, which includes sex-specific fishery ages (starting in 1978) and survey ages (starting in 1979). The maximum age observed for yellowfin sole is 40; however, the size distribution was truncated during the 1970s and 1980s following the extreme decline in stock size during 1960-1978, and few fish attained age 20 in the late 1970s. During the years subsequent to 1977, the yellowfin sole stock had recovered significantly from its lowest spawning stock size of 63,000 t, and was approaching half a million

tons (459,000 t) (Figure 1). The post-1978 data used to fit the spawner-recruit Ricker curve do not include years in which stock sizes were <500,000 t, and the years with the lowest stock sizes are 1978 and 1979, which are both close to 500,000 t (Figure 2). Some information may be available to estimate recruitment during the early 1970s when the stock was smaller, because yellowfin sole are somewhat long-lived. However, these years cannot be incorporated because recruitment patterns are not expected to be consistent across regimes (Wooster and Zhang 2004). Therefore, similar to EBS pollock, the data used to fit the stock recruitment curve for yellowfin sole do not include contrasting population size at low stock sizes (<500,000 t).

This issue is relevant for the model bridge exercise to SS3 presented here. In SS3, the Ricker model directly estimates the expected number of recruits produced by a given stock size, with lognormal error. In the bespoke model 23.0, recruitment is estimated using a normally distributed mean and variance, and the objective function works to minimize the difference between these estimates and the Ricker predictions of recruitment. Given the fundamental differences between the bespoke and SS3 use of a stock-recruitment curve, as well as the data constraints informing the Ricker stock recruitment curve mentioned above, we recommend moving to a Tier 3 for the yellowfin sole stock assessment.

2. Time-varying *q* should use survey start date but not temperature

Subsequent to the introduction of the VAST survey index and age compositions, we were double counting temperature with *q* (catchability). We recommend removal of the temperature component of time-varying *q* and present sensitivity to this update. Correlated or redundant predictor variables in a statistical model, also referred to as collinearity, can lead to reductions in model accuracy and reference point estimation (Dormann et al. 2013, De Marco and Nóbrega 2018). In the yellowfin sole VAST methodology, the mean bottom temperature within specific strata is used in biomass indices and survey age compositions (Appendix 1). When the VAST index was first incorporated in the BSAI yellowfin sole (YFS) assessment in 2021, the redundancy and collinearity with the environmental (temperature) covariate on catchability went unnoticed. Since 2018, the BSAI YFS assessment annually varying catchability has used both EBS bottom temperature and survey start date, as shown in Eq. 1.

The rationale for the use of survey start date and temperature is as follows: earlier survey start dates usually encounter colder water, and since the timing of the survey start date is positively correlated with bottom water temperature, improvement in fitting the survey biomass estimates can be gained by estimating two new parameters (μ and γ). Akaike information criterion (AIC) was used to determine if the additional variables (S and T:S) improved the regression fit, and the improvement in fit was more than offset by the additional two parameters (Nichol et al. 2019).

However, the use of temperature in the annually varying catchability as well as the survey index constitute double-counting of data, and results in model overfitting (e.g. Figure 15). Therefore, we recommend the use of only the survey start date in future BSAI YFS assessment models. To eliminate the double counting of temperature, we used an index of survey start date (1982 – 2023) as the environmental link on catchability for the bridging analysis.

3. Bridging the bespoke model to SS3

We propose bridging the bespoke 'flatfish' Model 23.0 to an SS3 platform for consistency, exchangeability, and reproducibility. Of the 12 NPFMC flatfish assessments that were at one time all bespoke models in admb, seven have bridged to SS3, two remain as bespoke ADMB legacy models, and the two (BSAI and GOA) arrowtooth flounder assessments are transitioning to *rceattle* (Table 1). Working groups benefit from the use of the same platform for consistency across parameters and assumptions in each model. In addition, it provides exchangeability, so that team members can easily assist on stock assessments if the lead author is not available or requires assistance. In addition, there is a single up-to-date version of SS3 with regular, well-documented updates that is maintained externally, avoiding issues of template drift among users. Finally, we will eventually be required to transition to the FIMS platform, which is still in development, and allow for an easier transition from SS3 to FIMS.

Here we present a bridging analysis using the SS3 modeling framework with deterministic models structured to match the 2024 yellowfin sole assessment. The models have two fleets, a commercial fishery and the combined eastern Bering Sea (EBS) and northern Bering Sea (NBS) bottom trawl survey. The main data inputs include empirical weight at age from the fishery and survey, fishery catch and age composition, and survey biomass and age composition. Natural mortality is sex-specific and catchability is time-varying. Age-based selectivity curves for the fishery and survey are assumed to be logistic, and do not differ by sex. Fishery selectivity is time-varying, while survey selectivity is constant.

Bridging exercise – deterministic models Model 23.0 and 'mSS3deter'

All parameter values were fixed in a deterministic model run based on the 2024 yellowfin sole assessment Model 23.0 (Figures 3, 4, and 5). Model 23.0 estimates of catchability (q) vary over time (t) as a function of annual temperature (T), annual survey start date (S) and their interaction (T:S):

Eq. 1: $q_t = e^{-\alpha + \beta T + \gamma S + \mu T:S}$,

where α is time-invariant q, β is a time-varying covariate based on temperature, γ is a time-varying covariate based on survey start date, and μ is time-varying based on the interaction between temperature and survey start date.

The fixed annual catchability deviations in the deterministic SS3 model run were calculated from the estimate of the time-invariant q and the time-varying estimates of q and provide an exact match to Model 23.0:

Eq. 2: $q_{dev} = e^{-\alpha} - q_t$.

Selectivity for the fishery and survey in mSS3deter provided an exact match to Model 23.0 (Figures 4 and 5).

Recruitment in the SS3 deterministic model was assumed to follow mean recruitment (fixed to the value from 2024 assessment) with steepness fixed to 1, recruitment variation fixed to 0.6, and fixed annual recruitment deviations for 1954 – 2024 equivalent to values from the 2024

assessment (Figure 6). We also fixed the deviations used to initialize the age composition in 1954 because SS3 and Model 23.0 employ different approaches to initializing the population. SS3 initializes the population assuming equilibrium and includes a plus group:

Eq. 3: $N_{0,a,s} = \begin{cases} cR_0 e^{-aM_s} & a = 0\ to\ 3A-1 \\ \sum_{a=A}^{3A-1} N_{0,a,s} + \frac{N_{0,3A-1,s}e^{-Ms}}{1-e^{-Ms}} & a > A \end{cases}$

where $N_{0,a,s}$ is the sex-specific initial numbers at age, c is the assumed sex-ratio at recruitment and fixed at 0.5, R_0 is mean recruitment, and M is sex-specific natural mortality. The numbers in 1954 are then derived as:

Eq. 4: $N_{0,a,s} = N_{0,a,s} e^{dev_a}$.

The flatfish model initializes the population in 1954 as:

Eq. 5: $n_{1954,a,s} = \begin{cases} 0.5e^{\bar{R}} & a = 1 \\ e^{\bar{I} + \varepsilon_{a,s}} & a > 1 \end{cases}$,

where $\bar{R}$ is mean recruitment, $\bar{I}$ represents the mean number of age 2+ in 1954, and $\varepsilon_{a,s}$ is the sex and age specific deviation.

SS3 requires a single vector of age specific deviations to initialize the age composition in 1954. To do this, we needed to derive what the numbers at age in 1954 would have been assuming exponential decay and a plus group from the 2024 assessment:

Eq. 6: $N_{1954,a,s} = \begin{cases} 0.5e^{\bar{R}} & a = 1 \\ N_{a-1,s}e^{-M_s} & a > 1 \\ N_{A,s} = \frac{N_A e^{-M_s}}{1-e^{-M_s}} & a = A \end{cases}$.

The age specific deviations were then derived as follows:

Eq. 7: $dev_a = \sum_s \ln(n_{1954,a,s)} - \sum_s \ln(N_{1954,a,s)}$.

This resulted in a single deviation to input in SS3 and the combined male and female numbers at age in 1954 matched in the two models (Figure 7).

Model 23.0 and mSS3deter provided identical estimates of survey biomass (Figure 8), fits to the survey age compositions (Figure 9), fit to fishery age compositions (Figure 10), and identical numbers at age (Figure 11), as shown for ages 1 and above (1+), ages 6 and above (6+), and ages 9 and above (9+).

There is some discrepancy in spawning biomass based on mSS3deter vs. Model 23.0 early in the time series (Figure 13). Spawning biomass is calculated in both models using the same equation:

Eq. 8: $SSB = \sum_a N_{fa} e^{-(M+F)t} W_a M_a$,

where N_{fa} is the female numbers at age, M = natural mortality, F = fishing mortality, t is the fraction of the year prior to spawning, W_a is weight-at-age, and M_a = maturity at age. Model 23.0 and mSS3deter incorporate equivalent values for all these parameters, except the initial number of females. The combined male and female numbers-at-age are equivalent in Model 23.0 and mSS3deter (Figure 11). However, there is a difference between the male and female initial numbers at age. In Model 23.0, the initial numbers of males and females are equal, but SS3 adjusts the initial female numbers-at-age based on natural mortality (see Eq. 3), which is lower in females than males, resulting in higher number of initial females and therefore higher initial spawning biomass in mSS3deter than Model 23.0 (Figures 12 and 13). Spawning biomass does not differ once the initial age composition moves through the population.

Estimation phase – M25.0 and M25.0a

This section presents versions of mSS3deter that allow the SS3 models to estimate parameters, Model 25.0 and Model 25.0a. Parameters that were estimated include male natural mortality, mean log recruitment, age-specific deviations to initialize the 1954 age composition, annual recruitment deviations from 1954 to 2018, time-varying fishery selectivity, time-invariant survey selectivity, and time-varying catchability.

In these estimation models, time varying survey catchability is modeled using a linear offset and an environmental link to estimate annual deviations for 1982-2024:

Eq. 9: $q_y = q_{base} + P_t env_y$,

where P_t is a link parameter and env_y is the environmental index value (survey start date). The P_t is an estimated parameter and is a measure of the environmental variable's influence on q. Additionally, a normally distributed prior on catchability is used in SS3, where $q \sim N(-0.07176, 0.9)$ to match the prior of the 2024 assessment model.

Catchability (q) did not follow high and low temperature extremes as closely after the temperature covariates were removed (Figures 14 and 15). The addition of the ISS to survey age composition increased q slightly, approximately 5% over the entire time series over Model 25.0. Models 25.0 and 25.0a result in similar fits to the survey biomass estimates, but they did not fit years of extreme temperatures as closely as Model 23.0 (Figure 15). Survey selectivity (Figure 16) and fishery selectivity (Figure 17) matched almost identically, with a small shift to older ages in models without the temperature covariate on catchability. Fits to the survey and fishery age compositions were indistinguishable (Figures 18 and 19). Numbers at age differed slightly, early in the time series due to differences in calculation of initial numbers at age described in the bridging analysis, and later in the time series due to differences in catchability (Figure 20). Finally, estimation of spawning biomass for Model 25.0 and 25.0a were within the confidence intervals of the 2024 model, Model 23.0 (Figure 21).

Residual Patterns

One step ahead (OSA) residuals are considered the best way to observe residual patterns because they are independent, identically distributed (iid), and normally distributed. The OSA residuals shown in Figures 22, 23, 24, and 25 were calculated by removing the first age (age-1). For the survey, males and females show similar patterns, and the patterns are typically small-scale and overall do not exceed one standard deviation (Figures 22 and 23). For the fishery, the male and female patterns are similar for age compositional data, and generally the scale was below 2 standard deviations (Figures 24 and 25). The fishery and survey age compositions for both models show a reverse S-shaped curve in the Q-Q plot, indicating a heavy-tailed but symmetric distribution (Figures 26 and 27). The standardized deviation of normalized residuals (SDNR) was <1 for the survey and fishery for Model 25.0 and Model 25.0a. In general, an SDNR much greater than 1 is not consistent with a good fit to the data. A value less than 1 indicates that the data were fit better than expected, and is not a cause for concern (Francis 2011).

Retrospective analyses

Mohn's rho of a 10-year retrospective peel of spawning biomass was 0.09 for Model 25.0 and 0.07 for Model 25.0a, indicating that the input sample sizes generated by method of Hulson et al. (2024) improved the model. The retrospective plots are consistent with little retrospective pattern, and are not a cause for concern (Figures 28 and 29).

Likelihood profiles

Likelihood profiles were performed on several key parameters: catchability (Figure 29), initial recruitment (R0, Figure 30), and male natural mortality (Figure 31). A likelihood profile on survey catchability (q) was performed (Figure 29); survey data indicates a higher value and fishery indicates somewhat lower, consistent with the survey and fishery influence on initial recruitment. The likelihood profile for initial recruitment (R0) suggests that the global MLE for R0 is approximately 14.8 (Figure 30). Fishery data indicates that initial recruitment is somewhat higher, while survey age composition data indicates R0 is somewhat lower. The profile of male natural mortality showed the MLE at 0.12, with the fishery data indicating somewhat higher (0.14) and the survey lower (0.10) (Figure 31).

Conclusion

Overall, the authors prefer adopting Model 25.0a for ABC/OFL and status determination advice under the FMP. This model makes use of SS3, a platform that is easily accessible to multiple users, providing more flexibility and a straightforward transition to FIMS in the future. The model ameliorates current collinearity of temperature covariates and updates data weighting to the best scientific methodology (surveyISS).

References

Francis, R.C., 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(6), pp.1124-1138.

He, X. and Field, J.C., 2019. Effects of recruitment variability and fishing history on estimation of stock-recruitment relationships: two case studies from US west coast fisheries. *Fisheries Research*, *217*, pp.21-34.

Methot, R. D., Jr., C. R. Wetzel, I. G. Taylor, and K. Doering. 2020. Stock Synthesis User Manual Version 3.30.15. U.S. Department of Commerce, NOAA Processed Report NMFS-NWFSC-PR-2020-05. https://doi.org/10.25923/5wpn-qt71.

Nichol, D.G., Kotwicki, S., Wilderbuer, T.K., Lauth, R.R. and Ianelli, J.N., 2019. Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey and its effect on estimates of survey biomass. Fisheries Research, 211, pp.319-330.

Rose, K.A. and Cowan Jr, J.H., 2000. Predicting fish population dynamics: compensation and the importance of site-specific considerations. *Environmental Science & Policy*, *3*, pp.433-443.

Rose, K.A., Cowan Jr, J.H., Winemiller, K.O., Myers, R.A. and Hilborn, R., 2001. Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish and Fisheries*, *2*(4), pp.293-327.

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

Zhou, S., 2007. Discriminating alternative stock–recruitment models and evaluating uncertainty in model structure. *Fisheries Research*, *86*(2-3), pp.268-279.

Tables

Table 1. Current NPFMC flatfish model platforms in SS3 (left column), and other platforms, with specifics in parentheses (right column).

SS3	Other platforms
BSAI Greenland Turbot	BSAI Kamchatka (ADMB Bespoke)
BSAI Flathead Sole	BSAI + GOA ATF (rceattle)
GOA Flathead Sole	BSAI Northern Rock Sole (ADMB Bespoke)
GOA Northern & Southern Rock Sole	
GOA Rex Sole	
BSAI Alaska Plaice	
GOA deepwater flatfish	

Figures

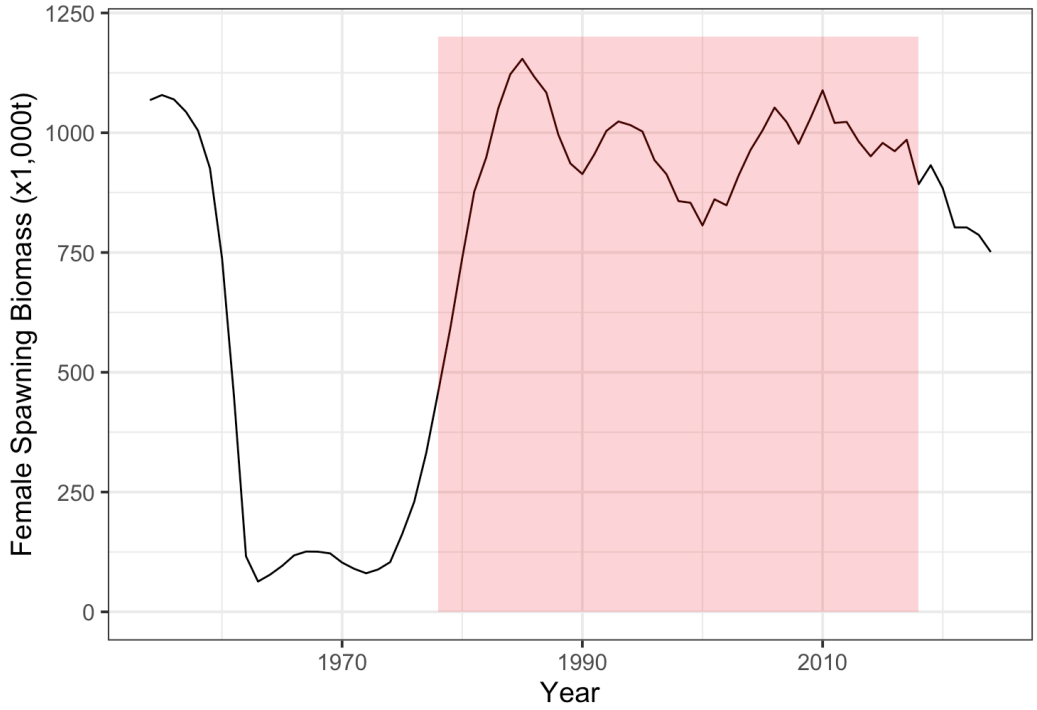


Figure 1. Model 23.0 estimate of spawning biomass from the 2024 BSAI yellowfin sole stock assessment. The shaded pink region indicates the years used to fit the Ricker stock recruitment curve in the model, which include the post-regime shift years 1978-2018.

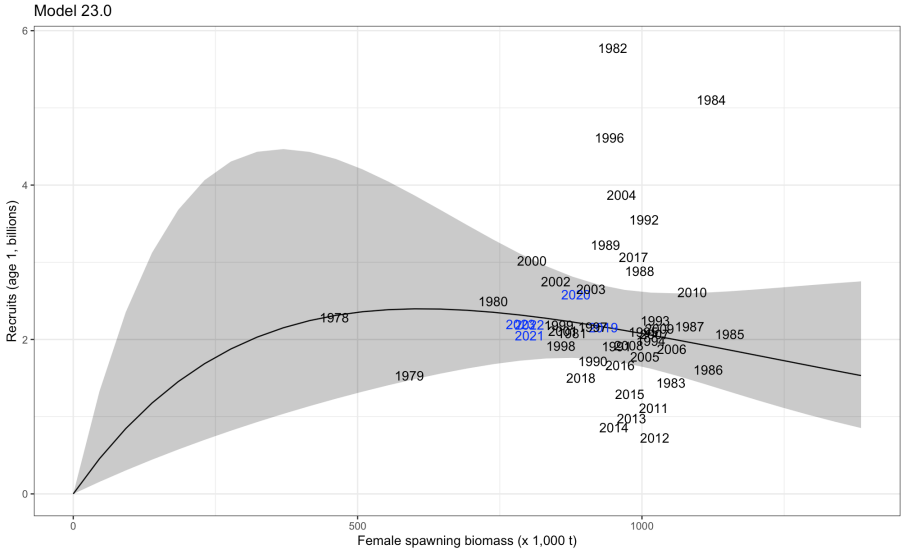


Figure 2. Ricker stock recruitment curve, with 95% confidence intervals fit to recruitment and spawning biomass estimates, 1978 – 2024. Years in blue (2019 – 2024) were recent and were not included as data fit to the model.

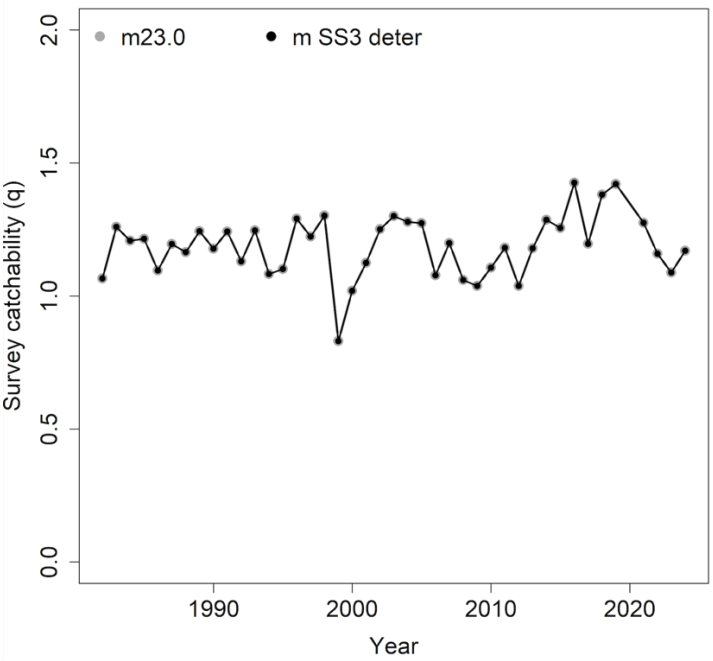


Figure 3. Deterministic run showing survey catchability for Model 23.0 and mSS3deter.

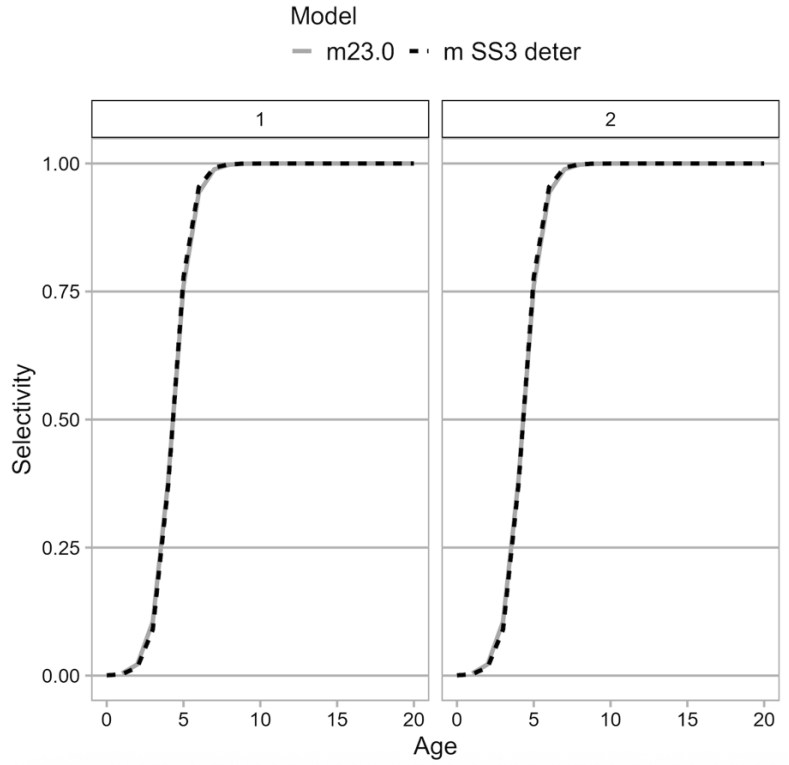


Figure 4. Deterministic run showing survey selectivity for males and females for Model 23.0 and mSS3deter.

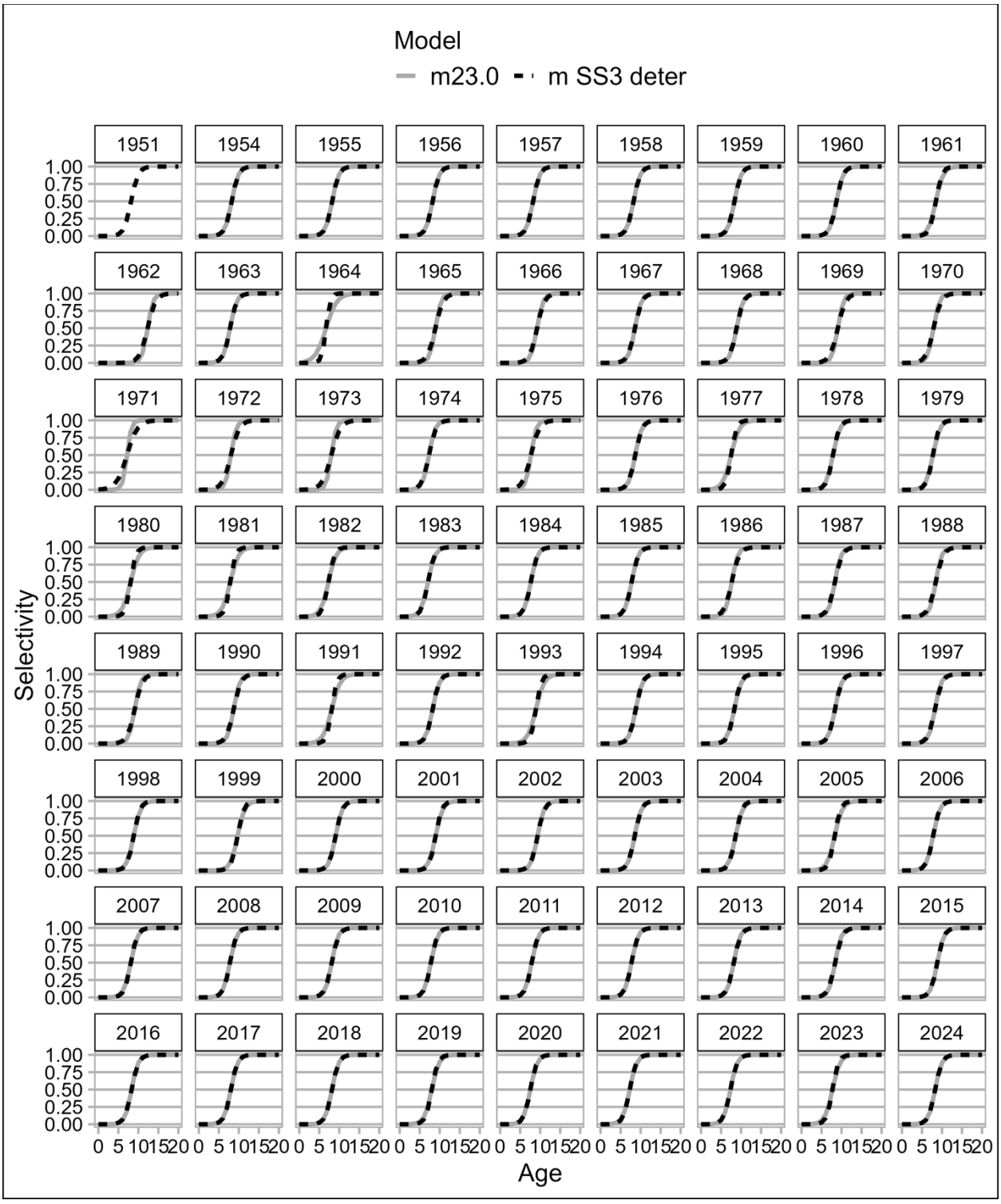


Figure 5. Deterministic run showing fishery selectivity for Model 23.0 and mSS3deter.

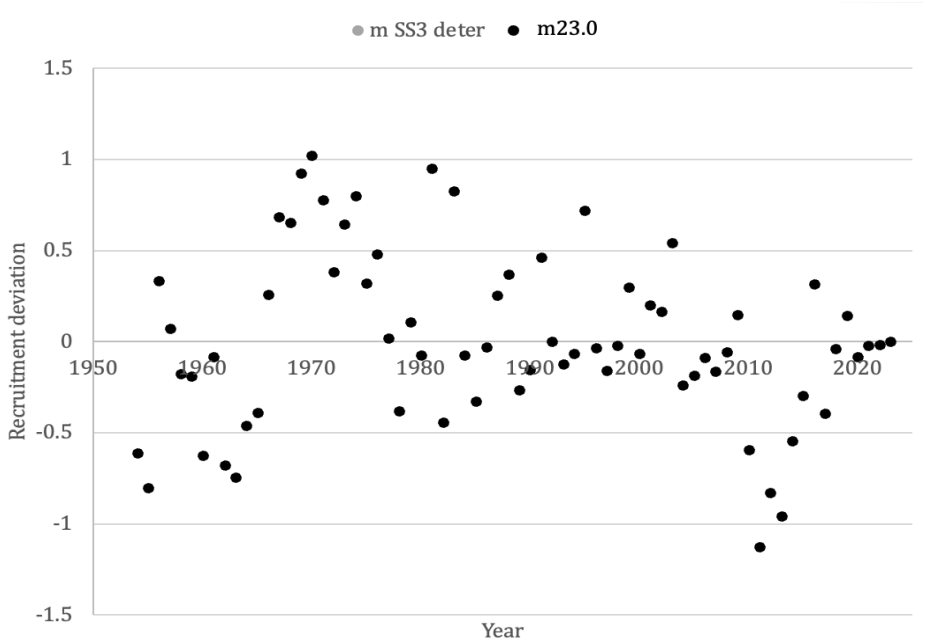


Figure 6. Deterministic run showing time series of recruitment deviations (1954 – 2023). Note that deviations for Model 23.0 and mSS3deter are overlapping.

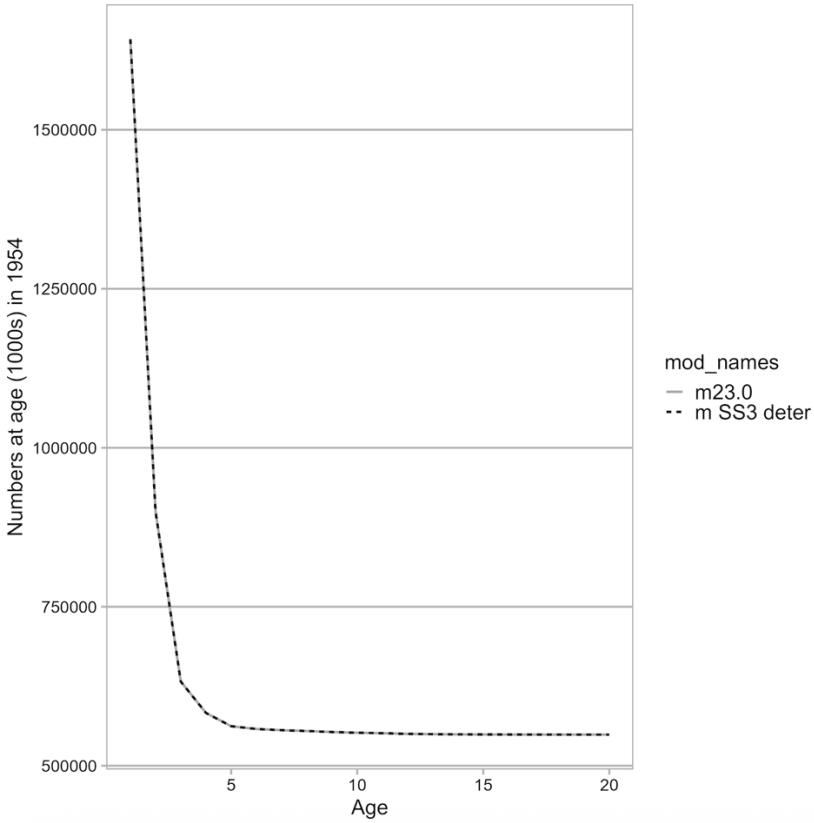


Figure 7. Initial numbers at age (males and females combined) for Model 23.0 and deterministic SS3 model (mSS3deter).

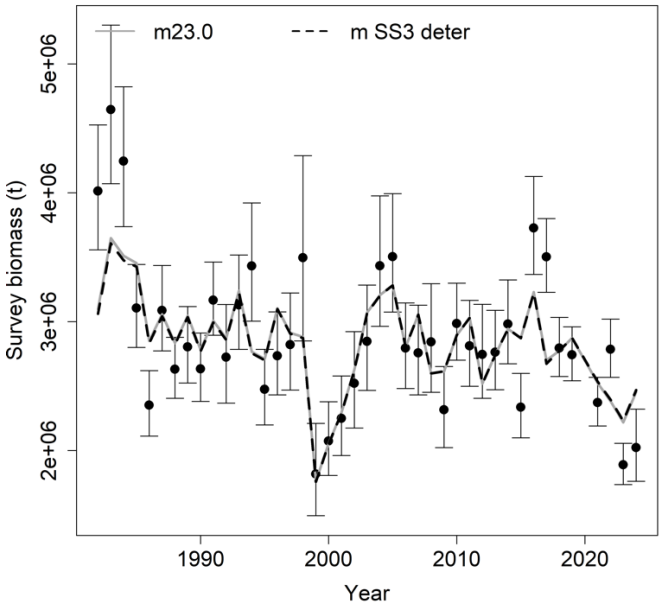


Figure 8. Deterministic run showing survey biomass estimate for Model 23.0 and mSS3deter.

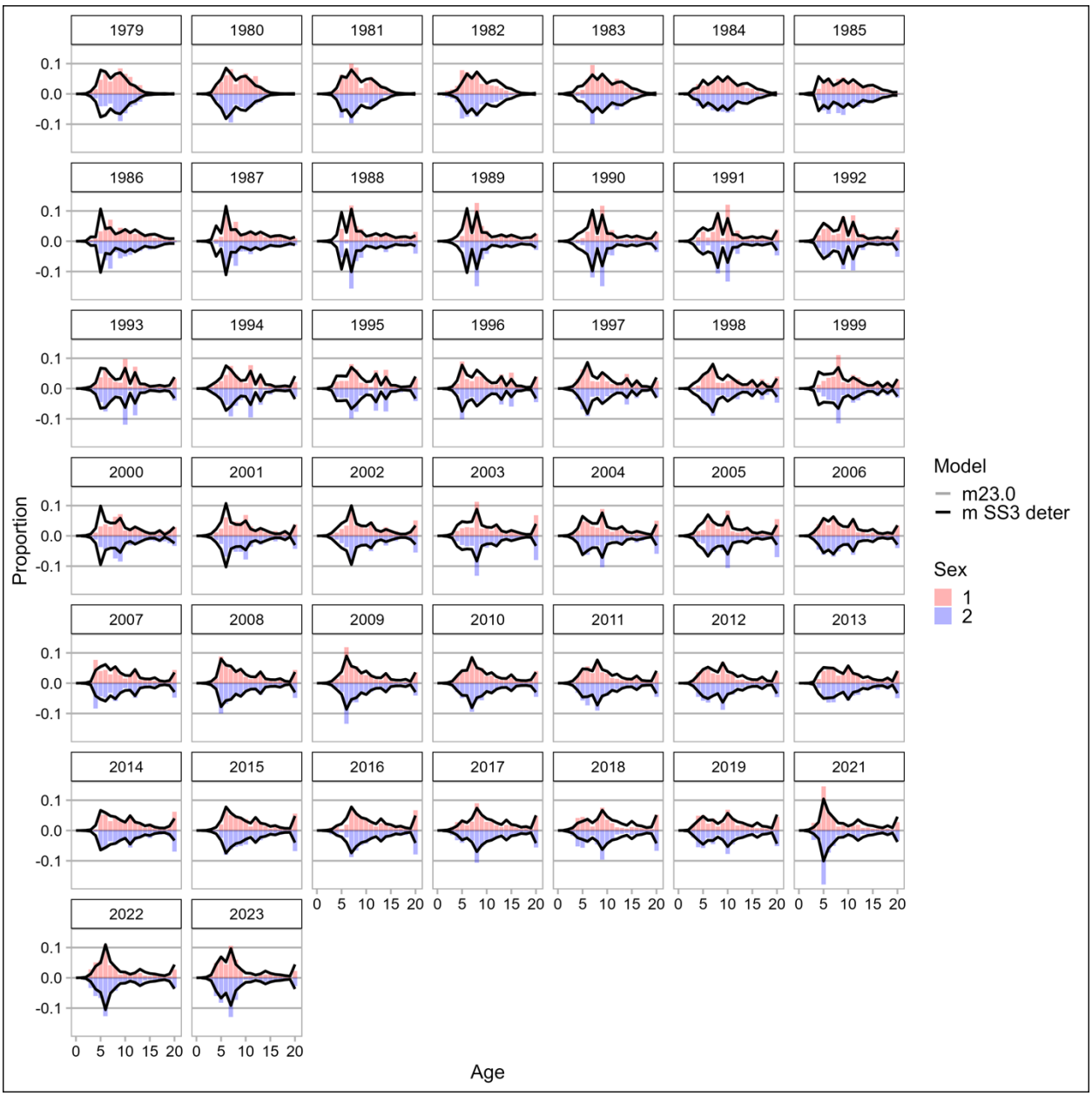


Figure 9. Deterministic run showing fits to survey age composition for males and females for Model 23.0 and mSS3deter.

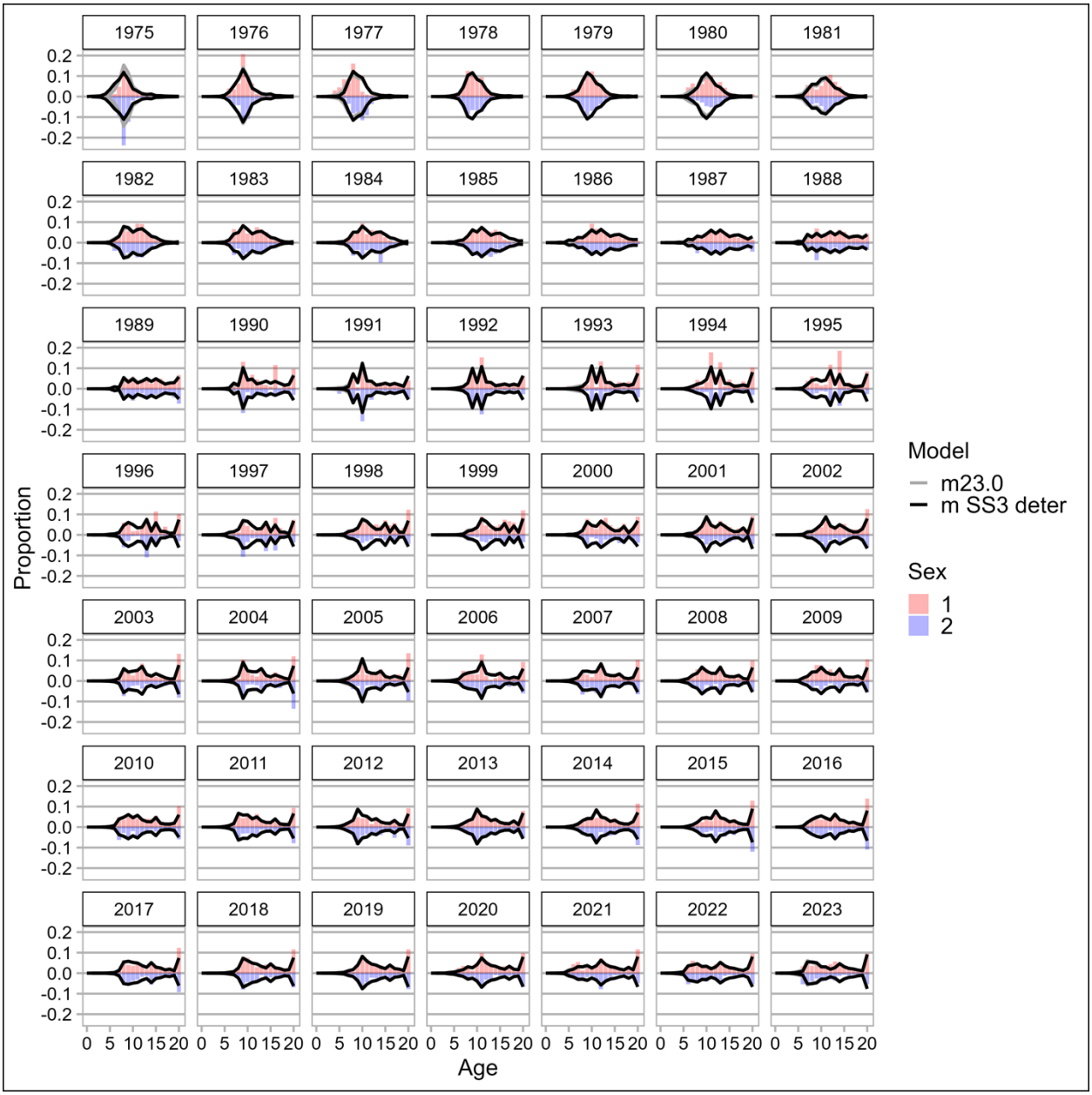


Figure 10. Deterministic run showing fits to fishery age composition for males and females for Model 23.0 and mSS3deter.

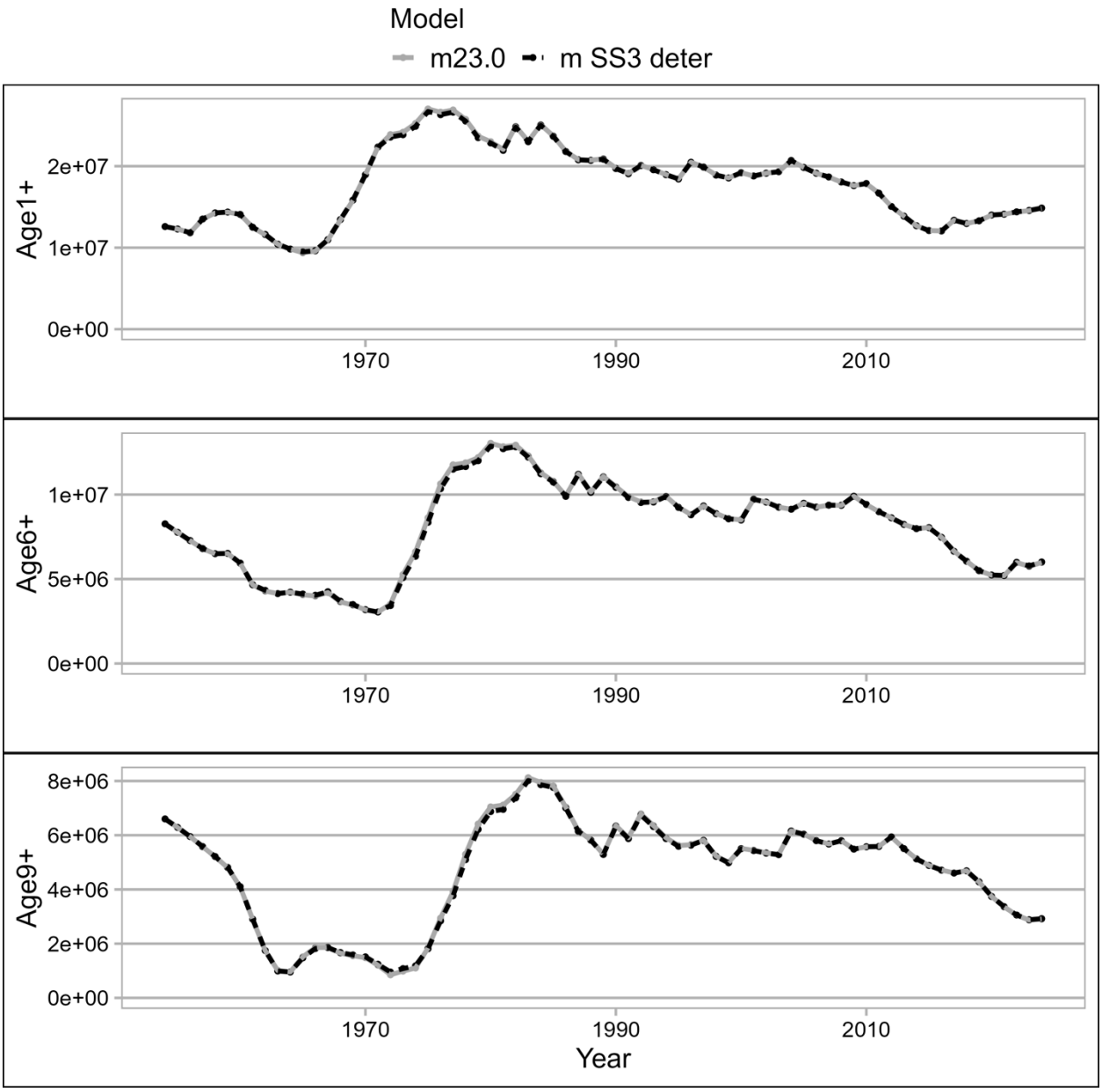


Figure 11. Deterministic runs of Model 23.0 and mSS3deter showing numbers at Age 1+, Age 6+ and Age 9+.

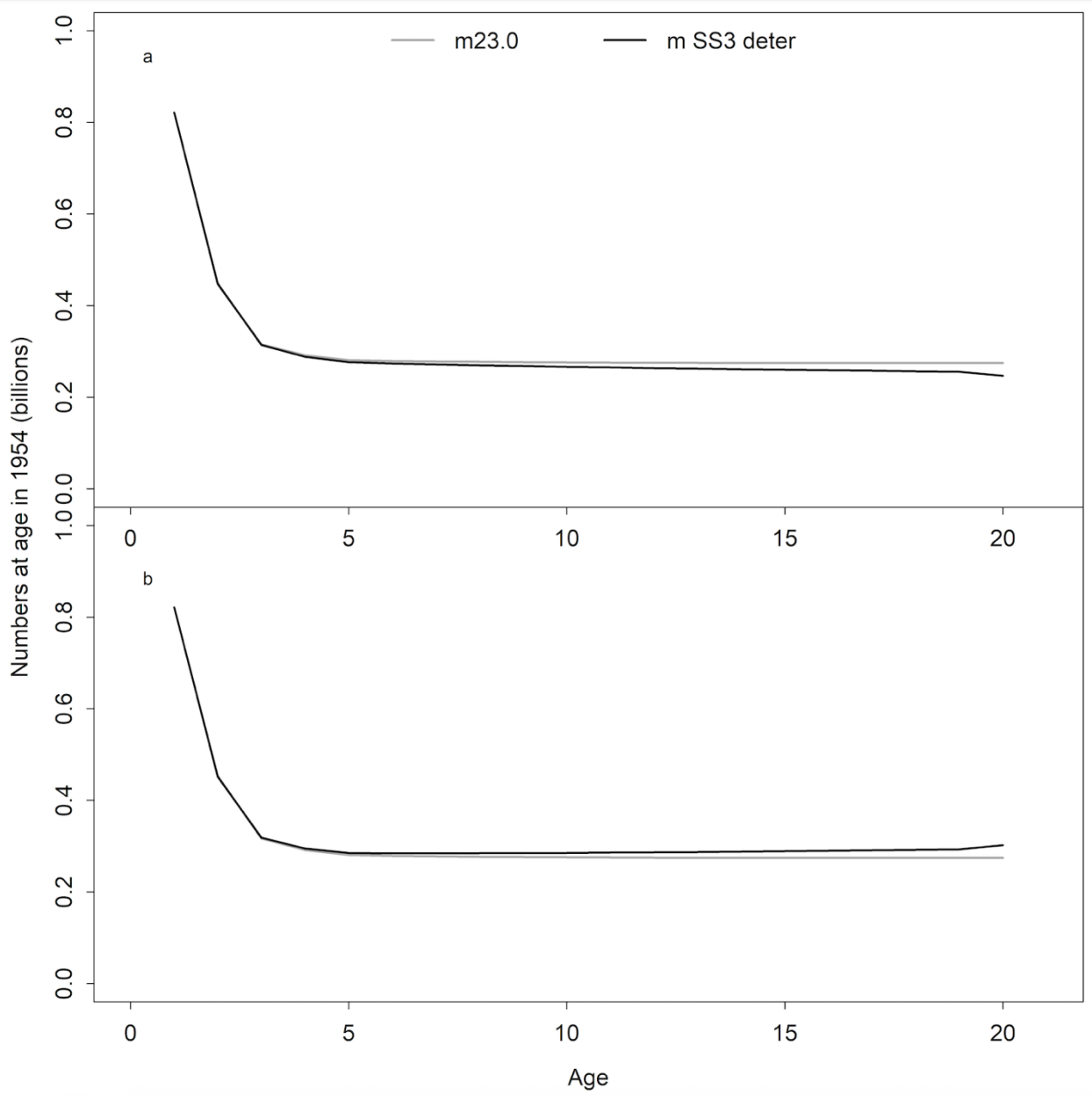


Figure 12. Initial numbers at age for males (panel a.), and females (panel b.), showing that SS3 (mSS3deter) incorporates different natural mortality for males and females (see Eq. 3). Model mSS3deter has fewer males (panel a.) than Model 23.0 but more females (panel b.), because males have higher natural mortality than females.

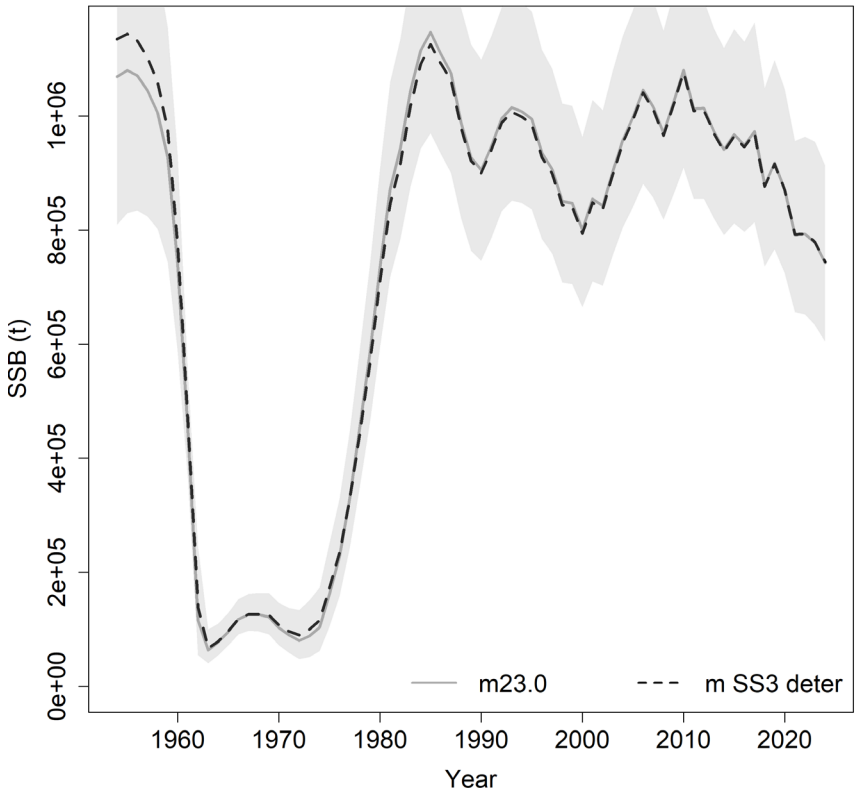


Figure 13. Deterministic runs of Model 23.0 and mSS3deter showing spawning stock biomass.

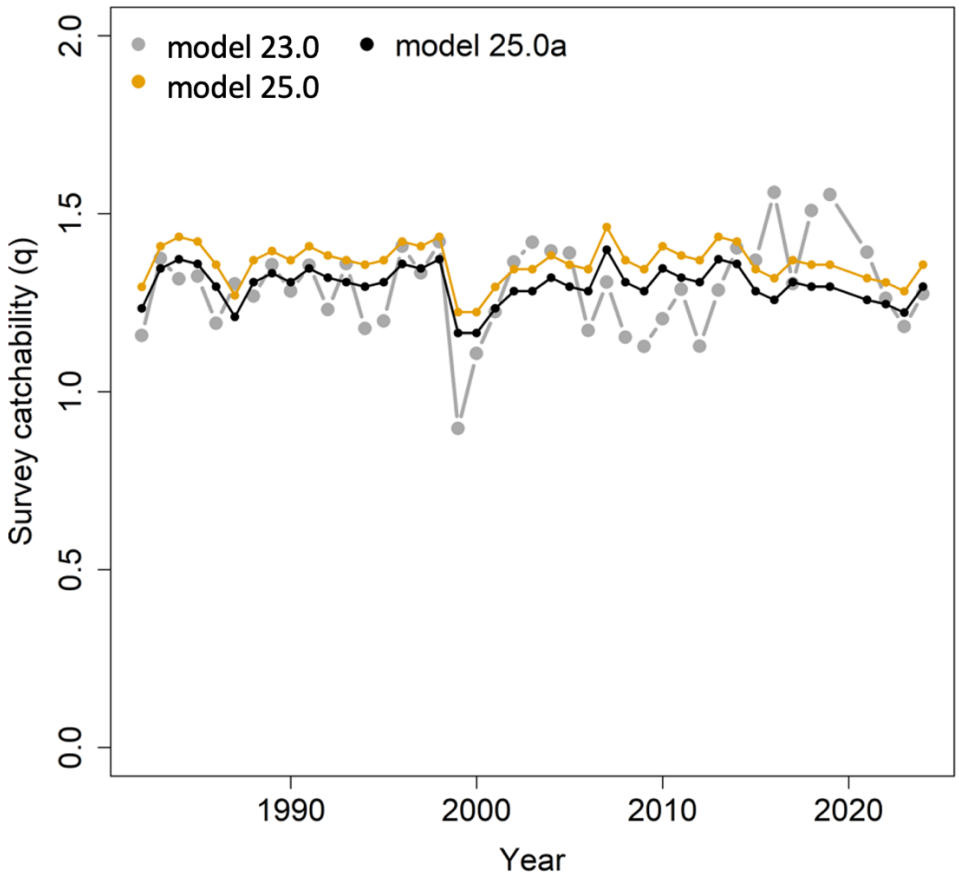


Figure 14. Estimation runs showing catchability for Model 23.0, Model 25.0, and Model 25.0a.

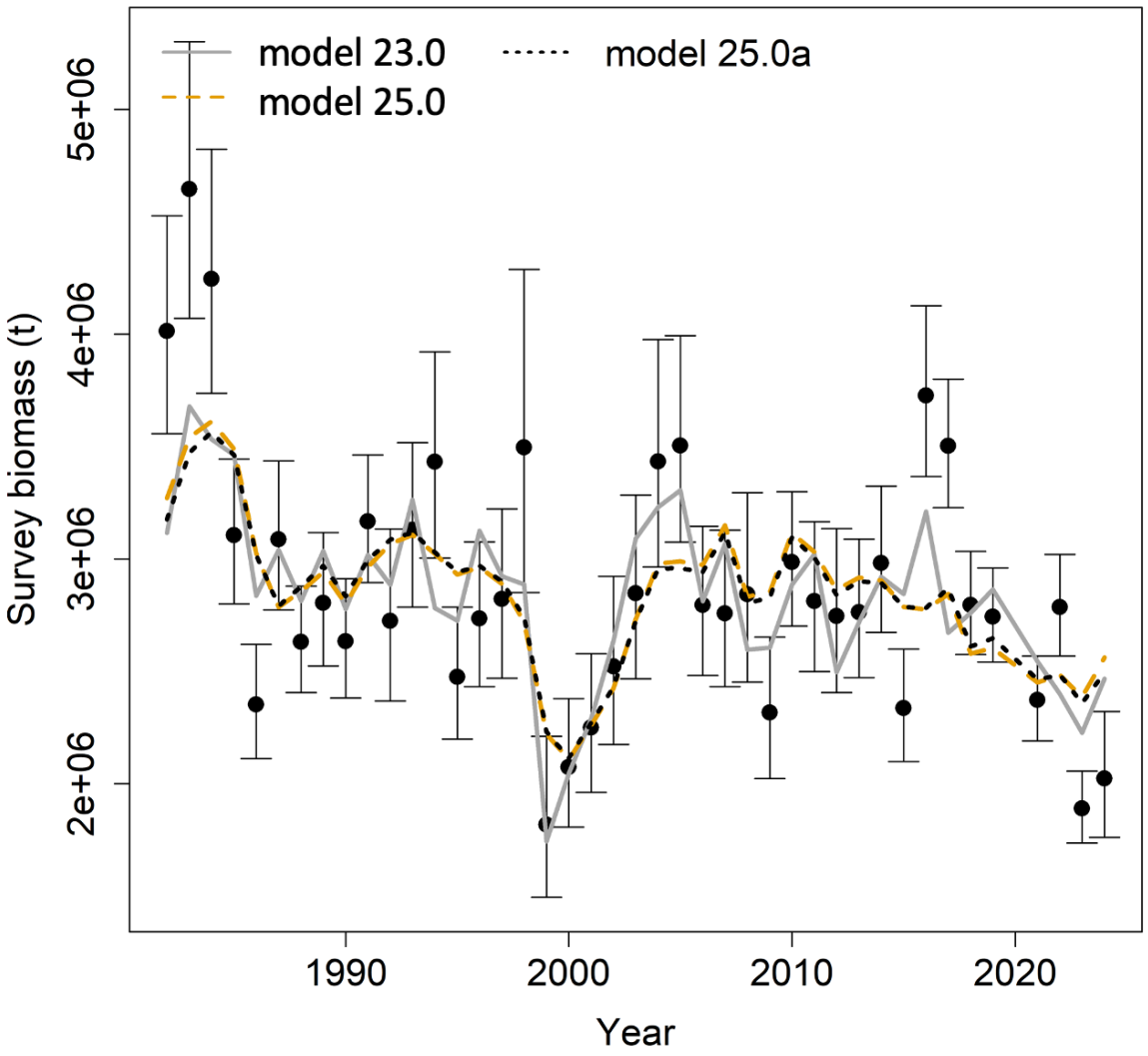


Figure 15. Estimation runs showing fit to survey biomass for Model 23.0, Model 25.0, and Model 25.0a.

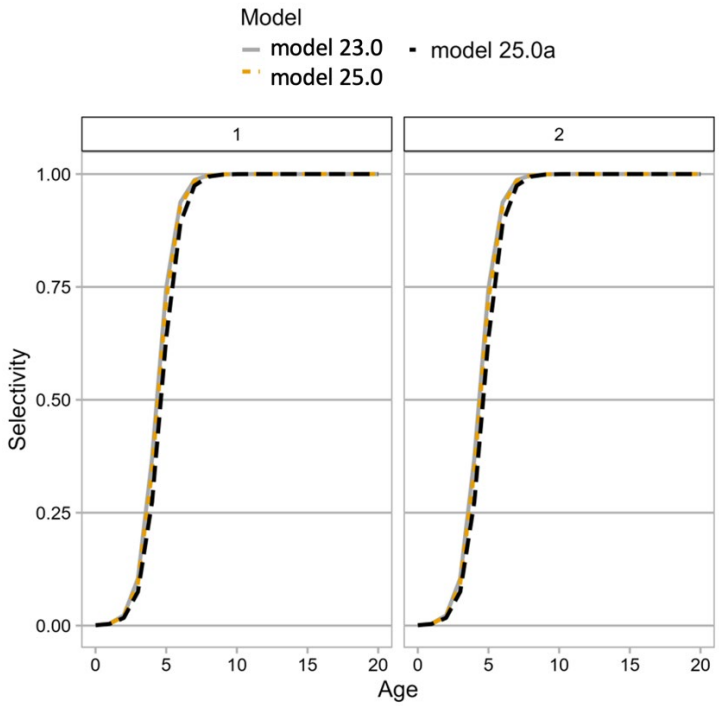


Figure 16. Estimation run showing survey selectivity for males and females for Model 23.0, Model 25.0, and Model 25.0a.

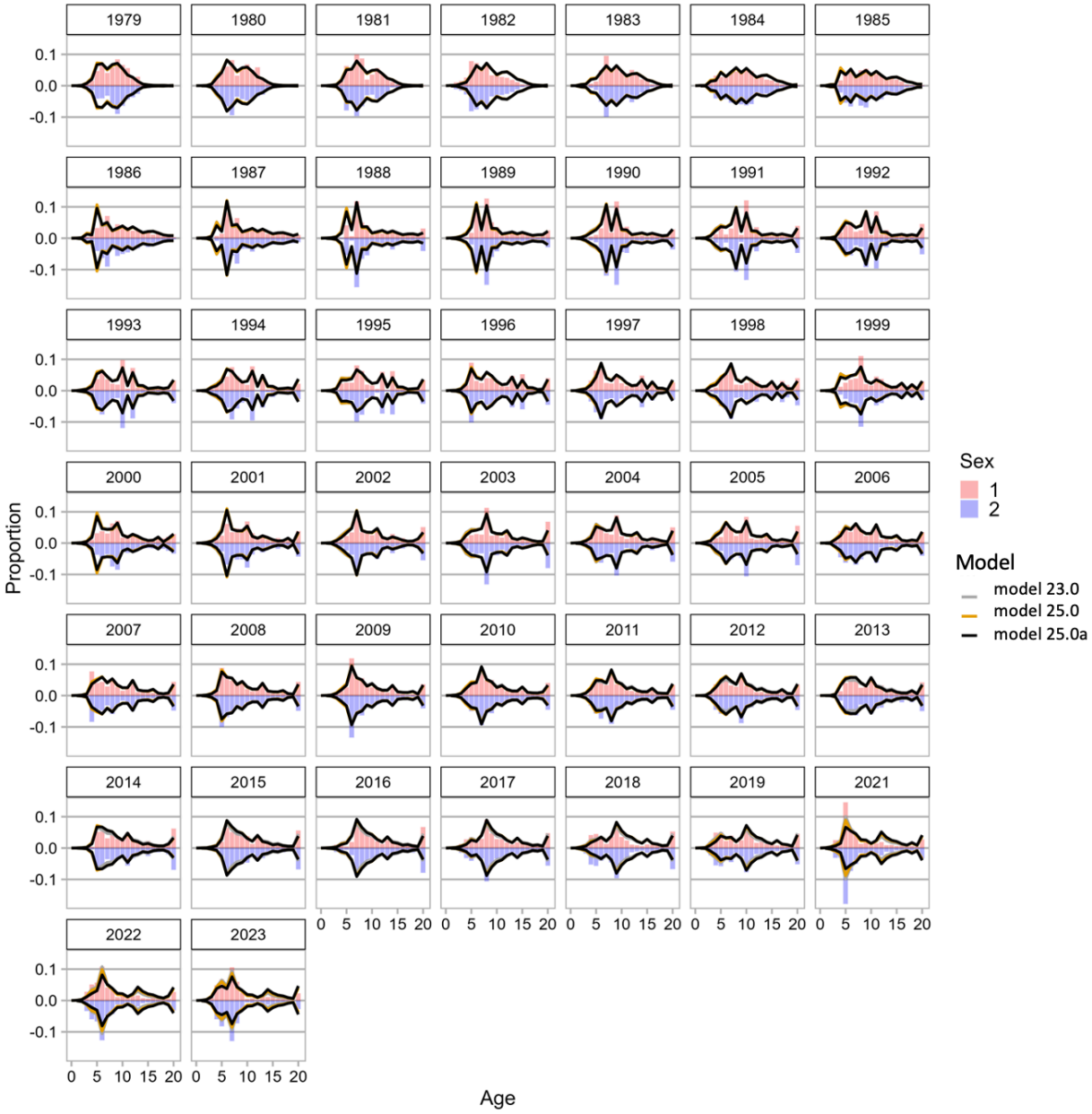


Figure 17. Estimation run showing fit to survey age composition for males and females for Model 23.0 , Model 25.0, and Model 25.0a.

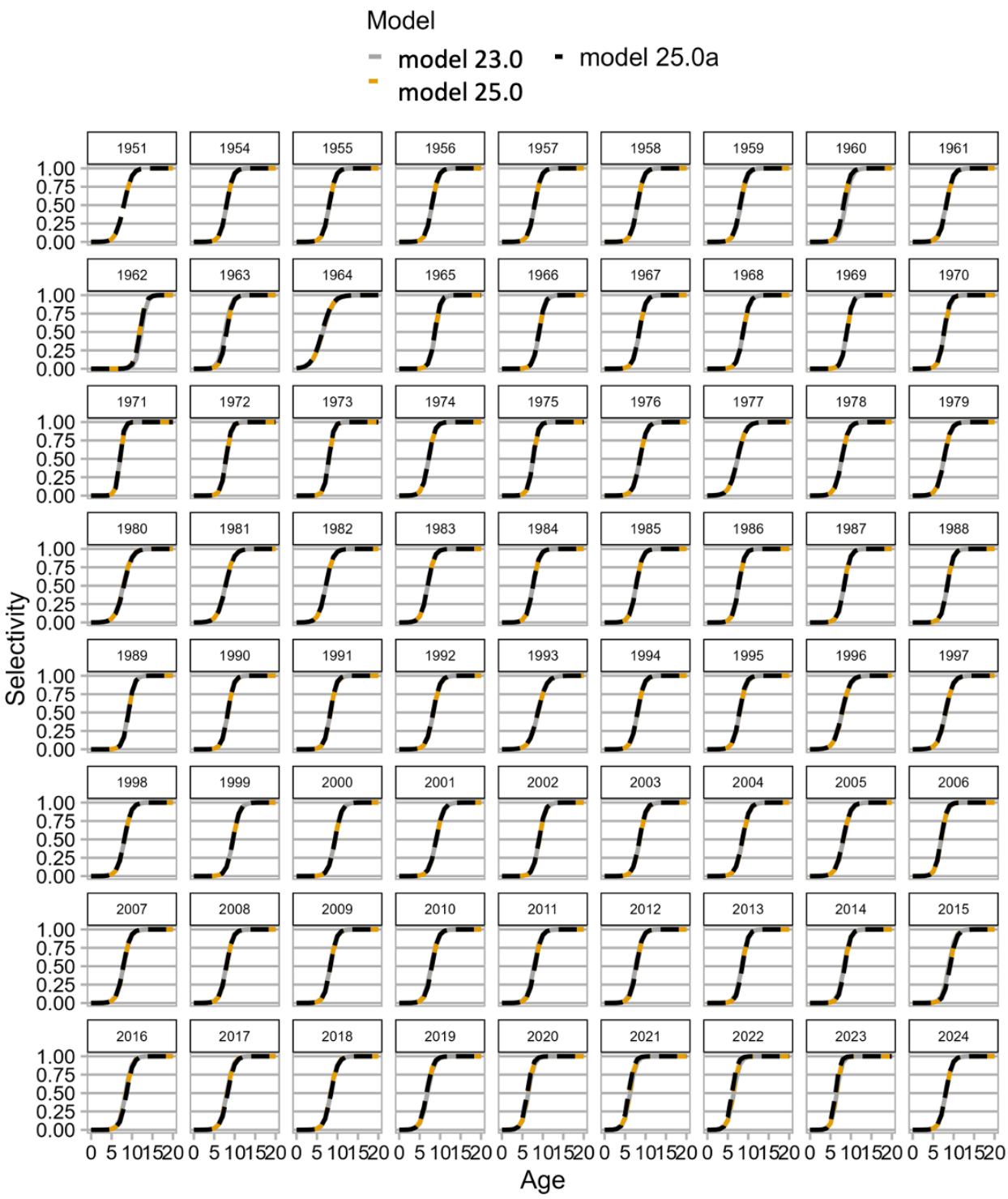


Figure 18. Estimation run showing fishery selectivity for Model 23.0, Model 25.0, and Model 25.0a.

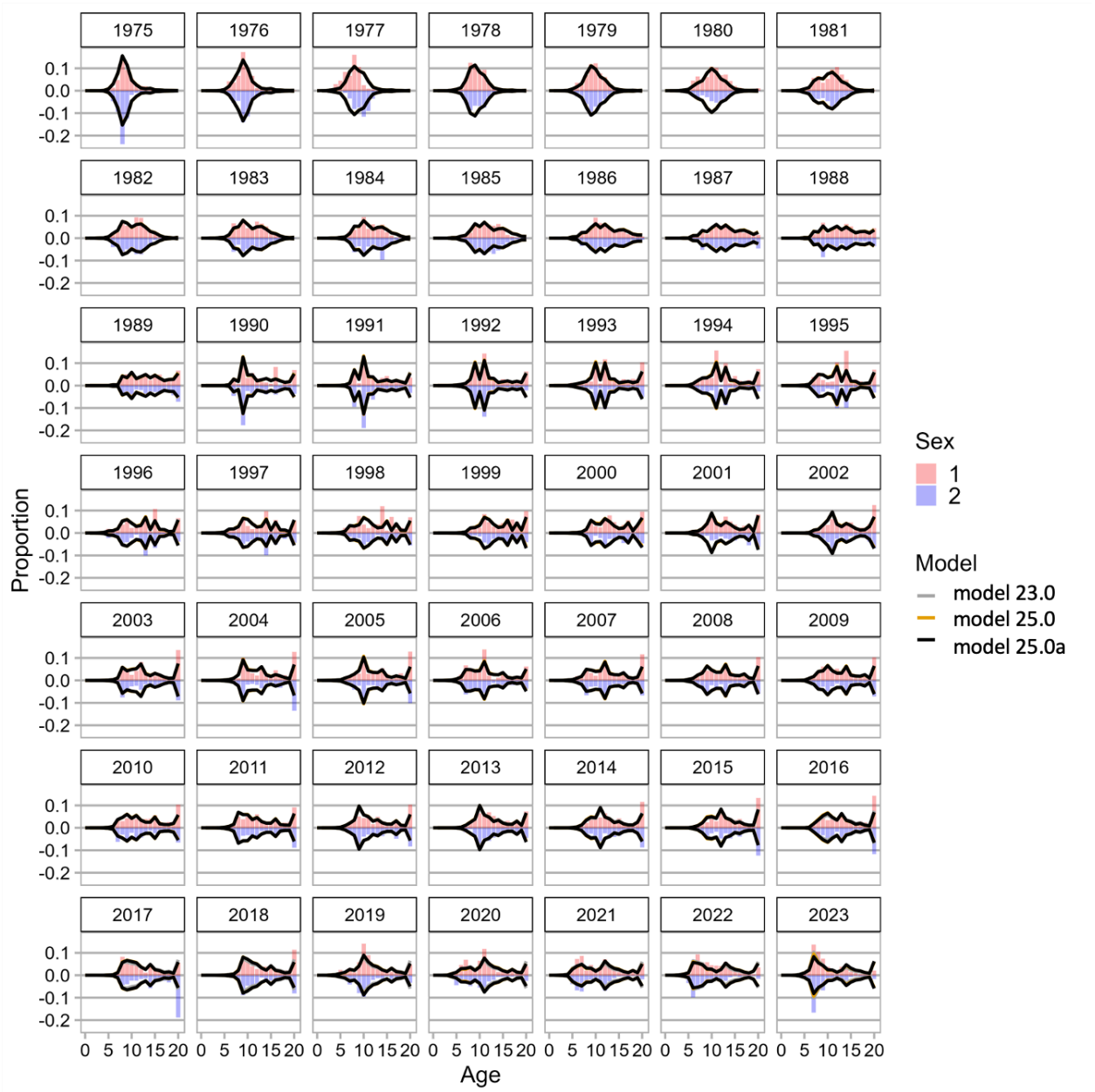


Figure 19. Estimation run showing fit to fishery age compositions for Model 23.0, Model 25.0, and Model 25.0a.

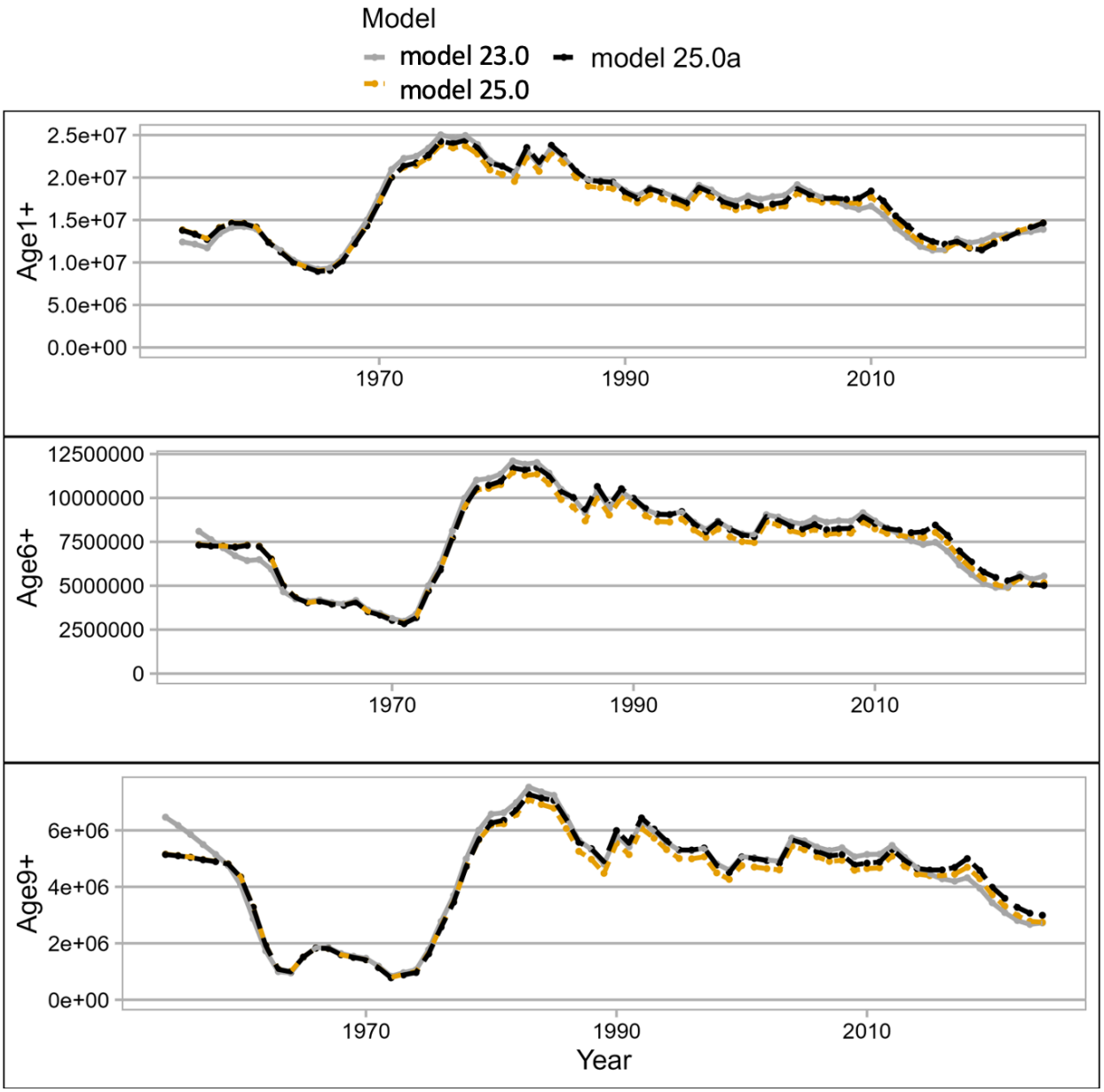


Figure 20. Estimation runs compositions for Model 23.0, Model 25.0, and Model 25.0a. showing numbers at all ages 1 and above (Age 1+), ages 6 and above (Age 6+), and ages 9 and above (Age 9+).

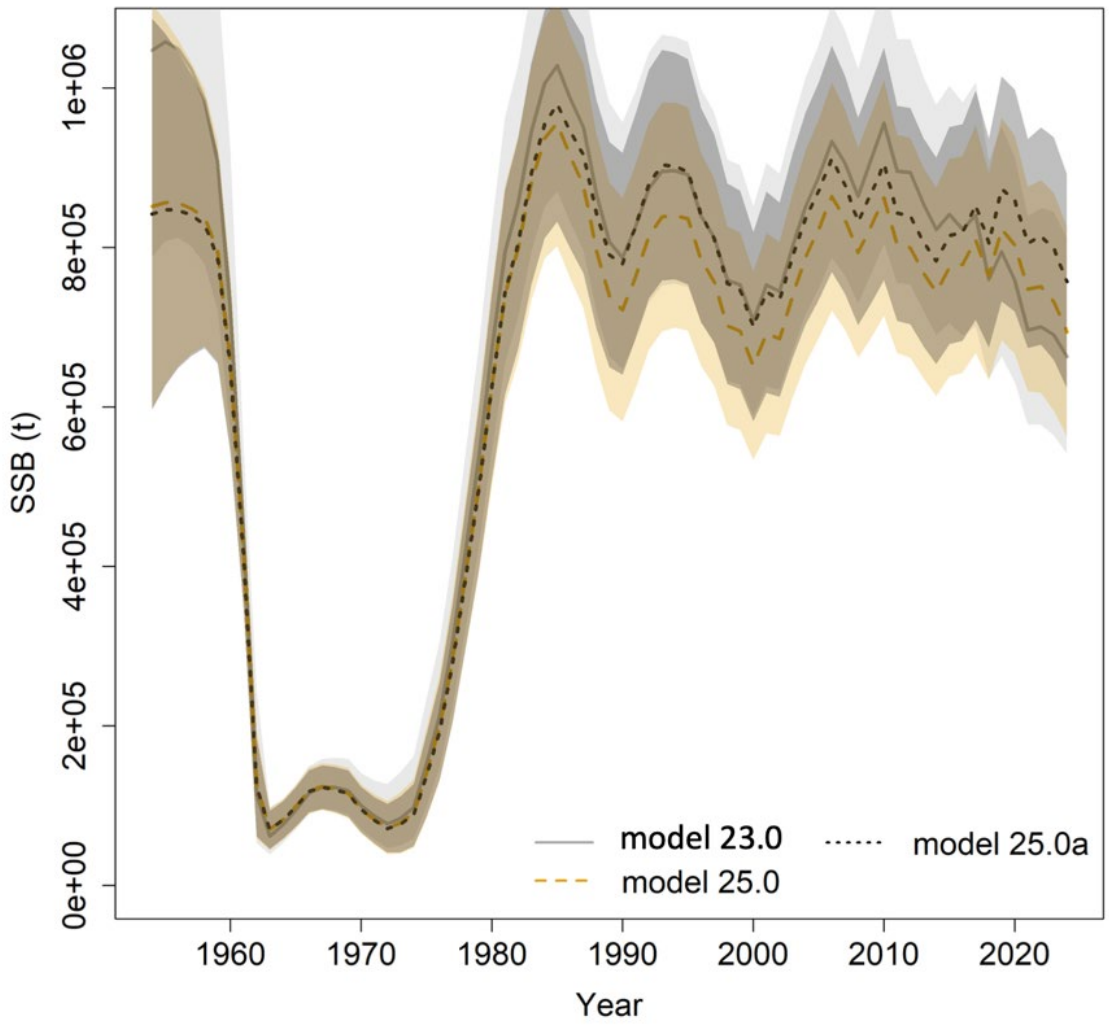
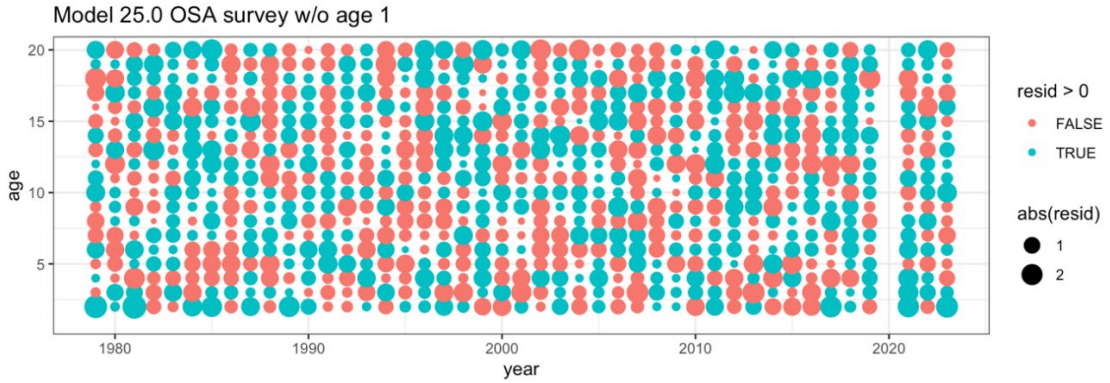


Figure 21. Estimation runs compositions for Model 23.0, Model 25.0, and Model 25.0a. showing estimates of spawning biomass.

a. Males



b. Females

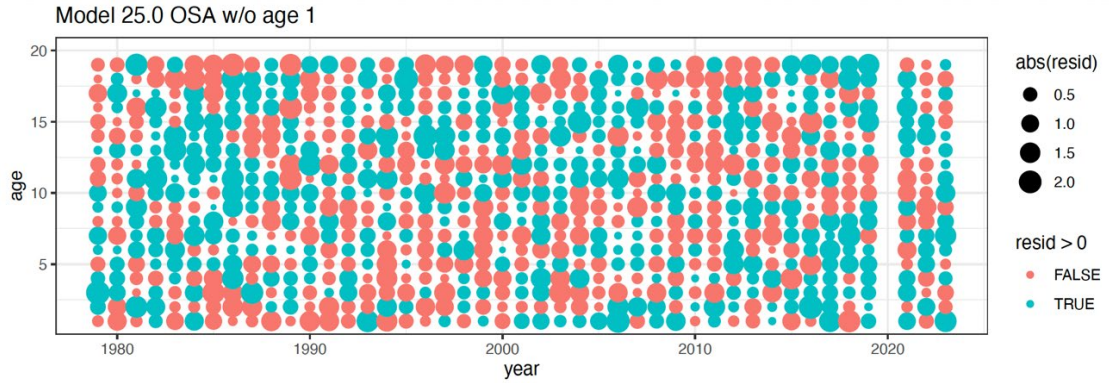


Figure 22. Model 25.0 survey age composition one step ahead males (upper panel), and females (lower panel).

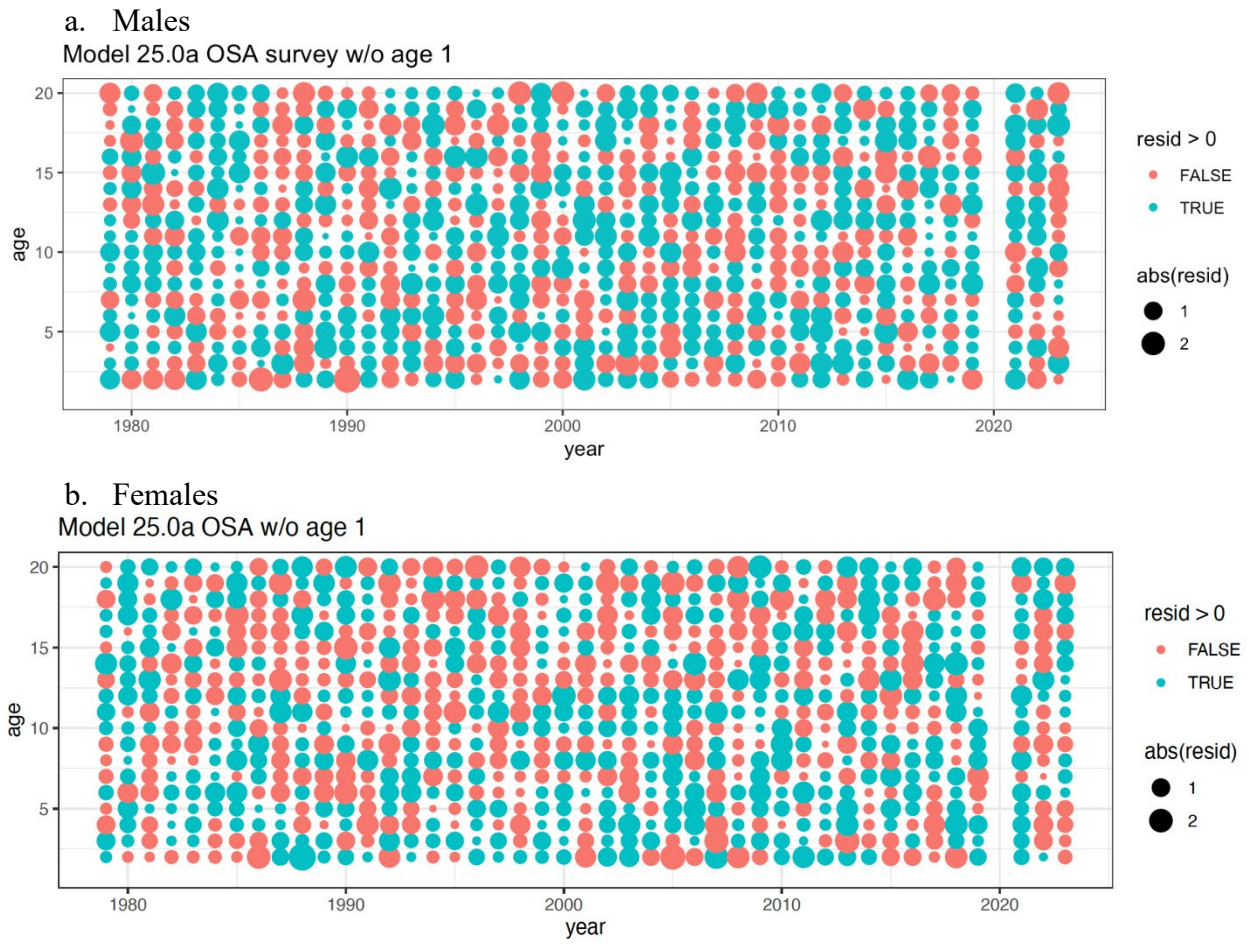
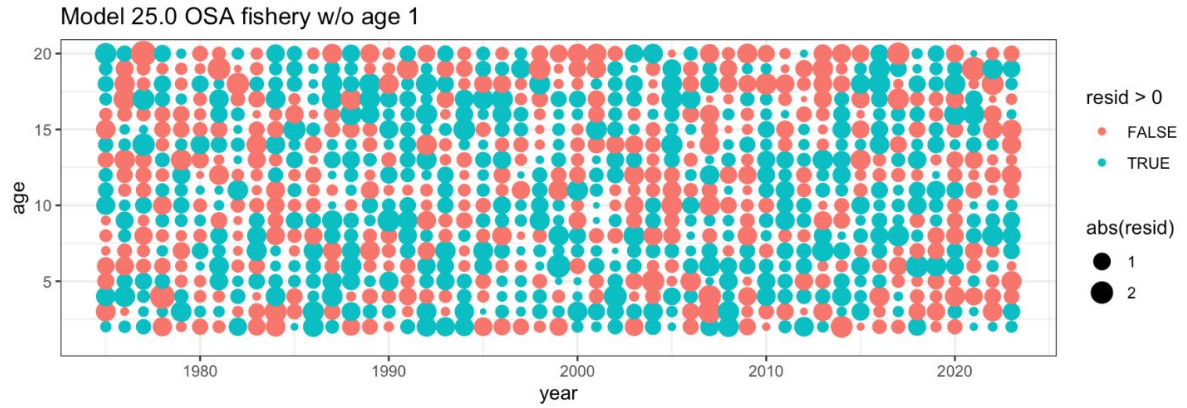


Figure 23. Model 25.0a survey age composition one step ahead males (upper panel), and females (lower panel).

a. Males



b. Females

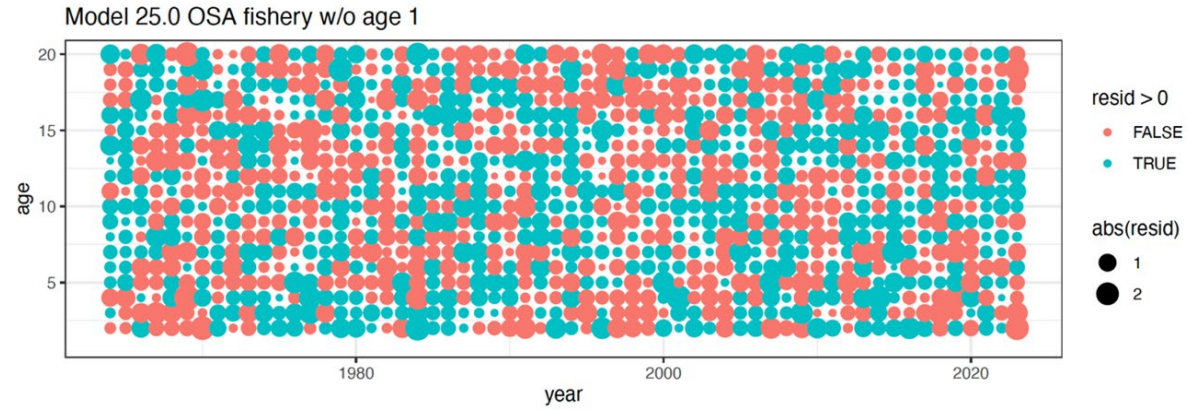
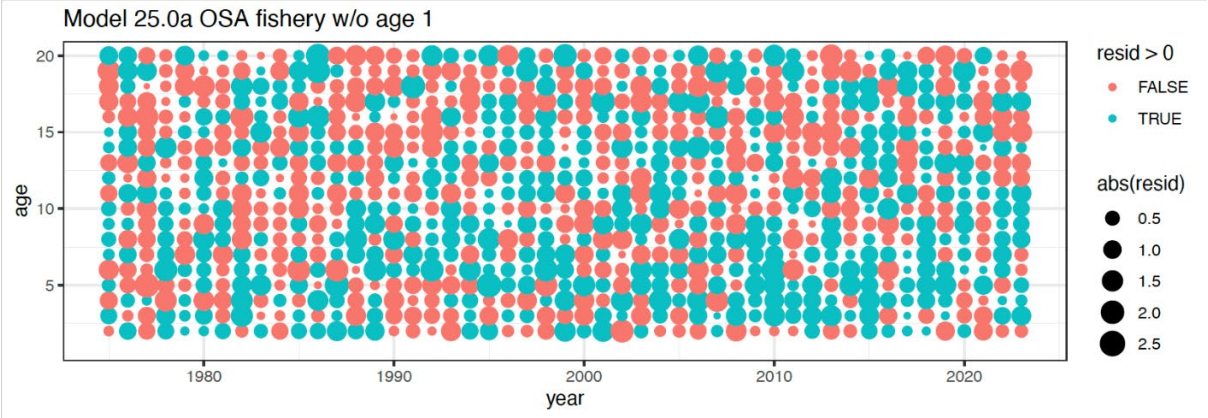


Figure 24. Model 25.0 fishery age composition one step ahead residuals males (upper panel) and females (lower panel).

a. Males



b. Females

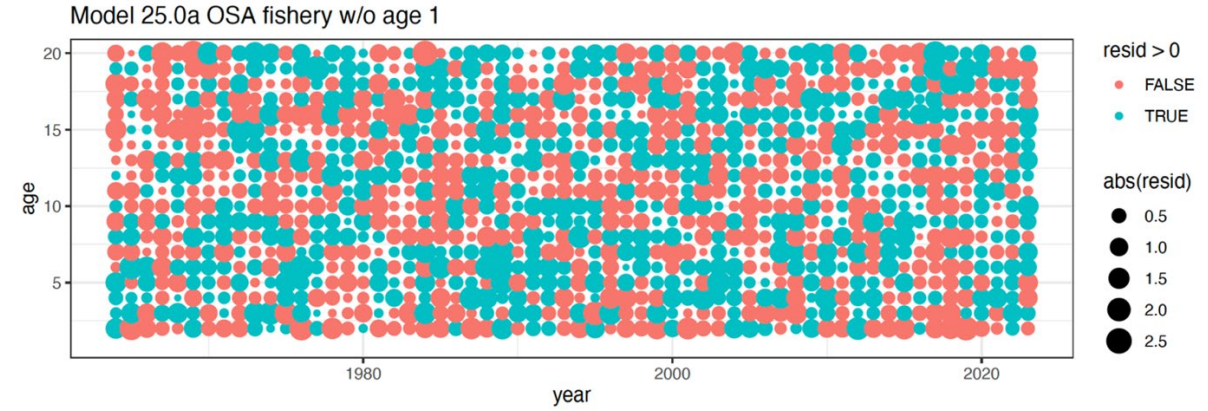


Figure 25. Model 25.0a fishery age composition one step ahead males (upper panel) and females (lower panel).

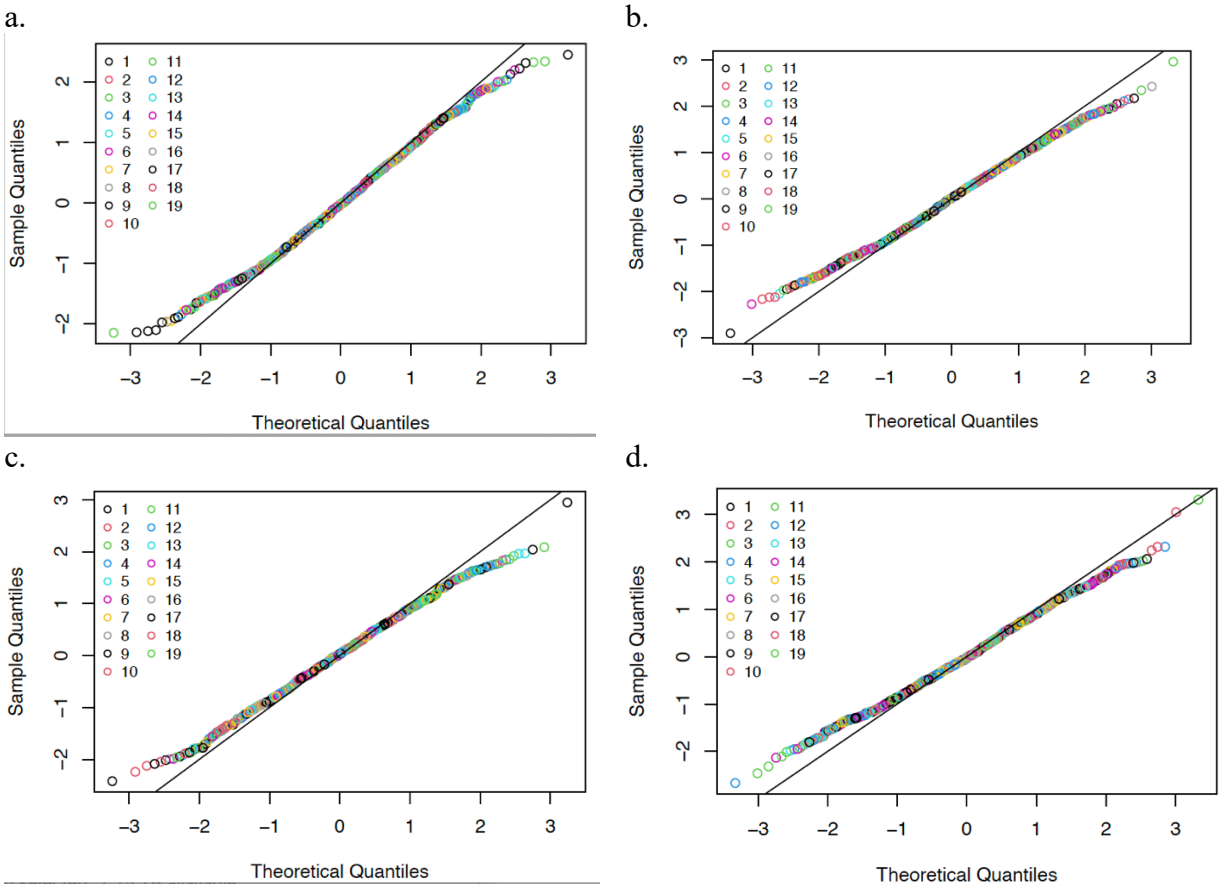


Figure 26. Females: Theoretical vs. sample quantiles for Model 25.0 survey (a.), Model 25.0 fishery (b.), Model 25.0a survey (c.), and Model 25.0a fishery (d.).

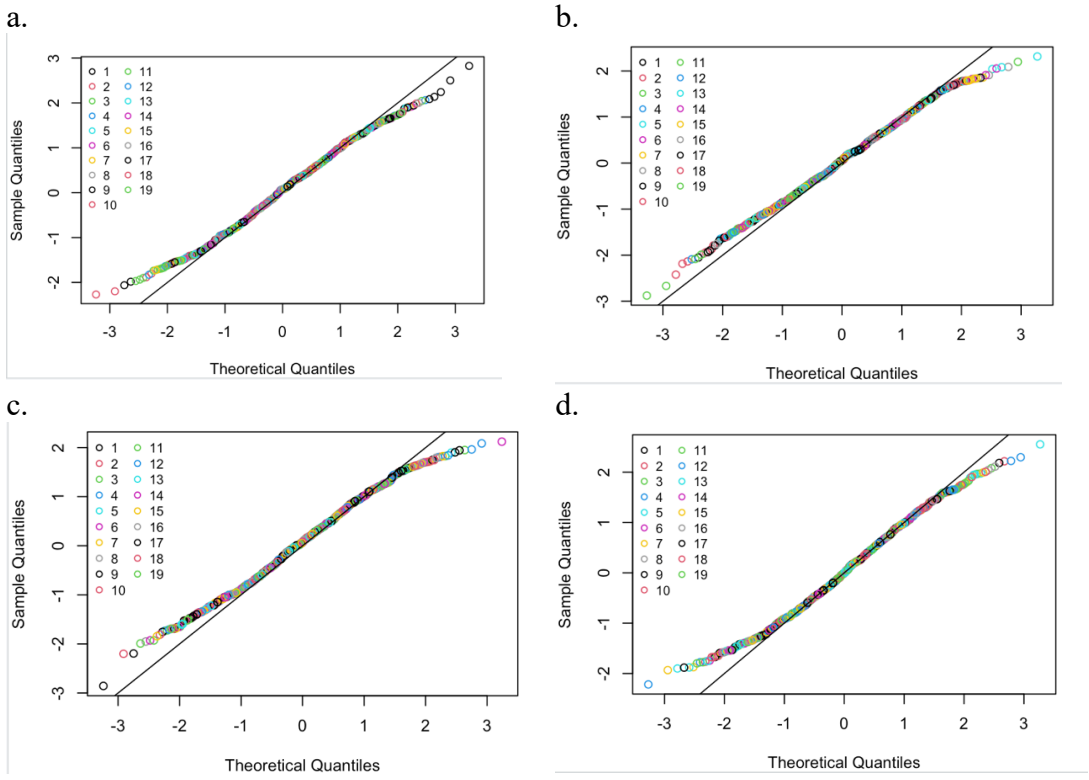
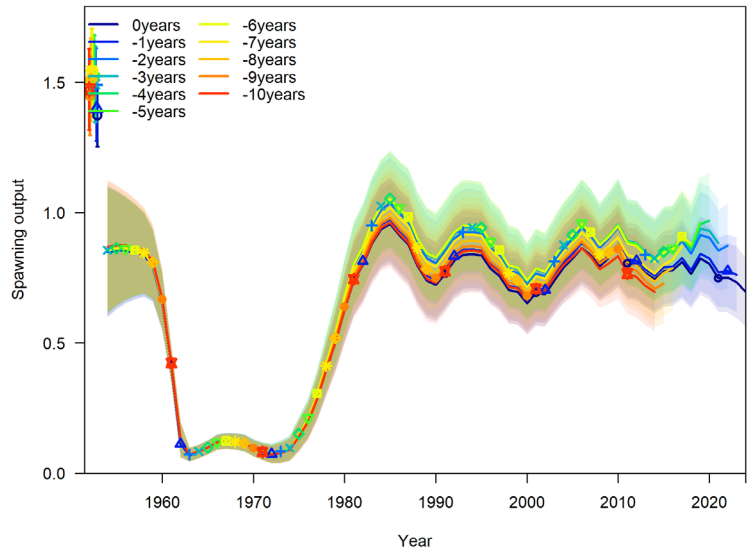


Figure 27. Males: Theoretical vs. sample quantiles for Model 25.0 survey (a.), Model 25.0 fishery (b.), Model 25.0a survey (c.), and Model 25.0a fishery (d.).

a.



b.

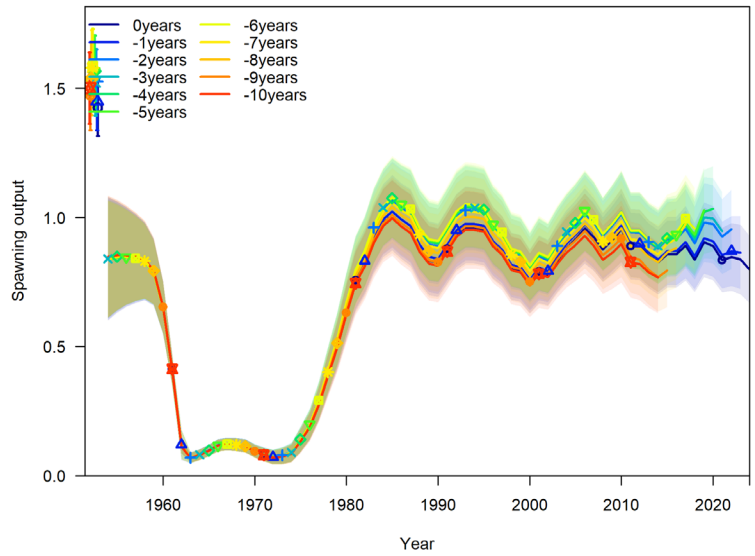


Figure 28. Retrospective plot of spawning biomass for Model 25.0 (panel a.) and Model 25.0a (panel b.).

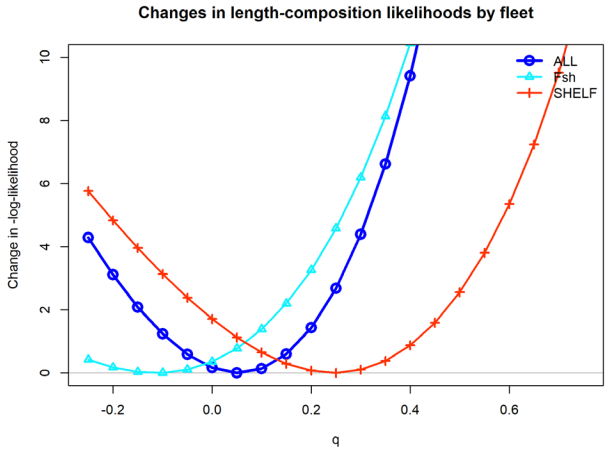


Figure 29. Model 25.0 likelihood profile over catchability, *q*.

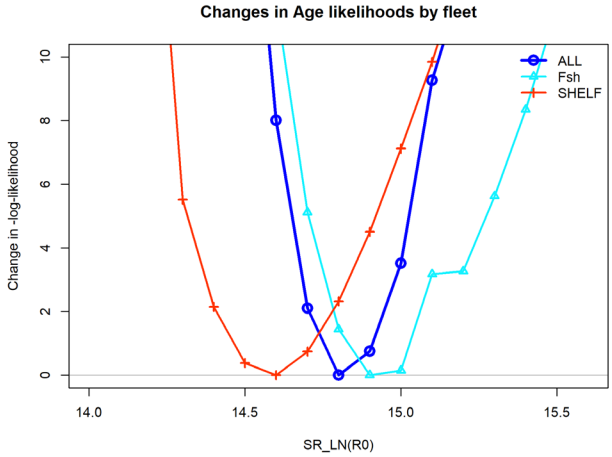


Figure 30. Model 25.0 likelihood profile over initial recruitment, R0.

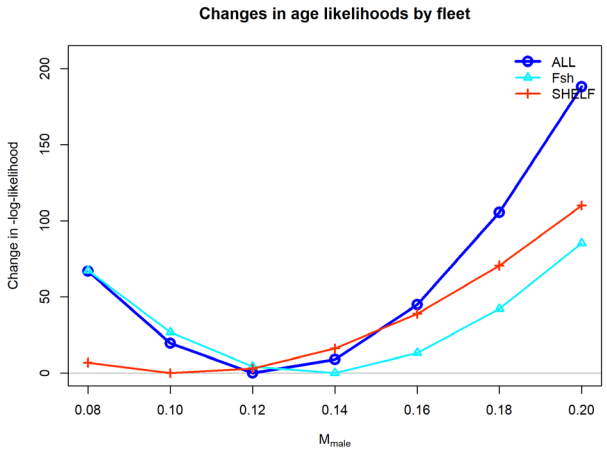


Figure 31. Model 25.0 likelihood profile over male M.

Appendix 1: VAST methodology applied to BSAI yellowfin sole.

Model-based abundance index methods (VAST) for the Bering Sea

For model-based indices in the Bering Sea, we fitted observations of numerical abundance or biomass per unit area (where the choice of abundance or biomass varied by stock at the request of assessors) from all sampling locations in the 83-112 bottom trawl survey of the EBS, 1982-2025, including exploratory northern extension samples in 2001, 2005, and 2006, as well as the 83-112 samples available in the NBS in 1982, 1985, 1988, 1991, 2010, 2017-2019, 2021-2023, and 2025. NBS samples collected prior to 2010 and in 2018 did not follow the 20 nautical mile sampling grid that was used in other survey years. Assimilating these unbalanced data requires extrapolating into unsampled areas, facilitated by including a spatially varying response to cold-pool extent (Thorson 2019). This spatially varying coefficient term was estimated for both linear predictors of a delta-model, and detailed comparison of results for EBS pollock has shown that it has a small but notable effect on these indices and resulting stock assessment outputs (O'Leary et al. 2020). Rather than using the cold-pool covariate for yellowfin sole, we instead used the mean bottom temperature within the inner and middle domain strata from an interpolated temperature product. All models were fitted in the *sdmTMB* R package (https://pbs-assess.github.io/sdmTMB; Anderson et al., 2022). All environmental data used as covariates were computed within the *coldpool* R package (https://github.com/afsc-gap-products/coldpool; Rohan et al., 2023).

We used a spatiotemporal Poisson-link delta-model (Thorson 2018) involving two linear predictors and a gamma distribution to model positive catch rates. We predicted population density across the entire EBS and NBS in each year, using AFSC GAP-approved prediction grids. Spatial and spatiotemporal fields were estimated using a spatial mesh including 250 "knots" whose locations were determined using a K-means algorithm to place the knots evenly over space, in proportion to the dimensions of the prediction grid. While the prior VAST model was fitted with a 750 knot mesh, we found that sdmTMB models fitted with mesh resolutions from 250-750 knots all provided extremely similar index estimates and that reducing the number of knots dramatically improved computation time. We estimated geometric anisotropy (how spatial autocorrelation declines with differing rates over distance in some cardinal directions than others) and included a spatial and spatiotemporal term for both linear predictors. To facilitate prediction of density in regions with unsampled years, we specified that the spatiotemporal fields were structured over time as an AR(1) process (where the magnitude of autocorrelation was estimated as a fixed effect for each linear predictor). However, in cases where the AR(1) correlation parameter ρ was estimated to be close to 1 for either model component, the model would be collapsed into a simpler structure by specifying $\rho = 1$, i.e., modeling spatiotemporal variation as a random walk. We did not include any temporal correlation for intercepts, which we treated as fixed effects for each linear predictor and year. Finally, we used epsilon bias-correction to correct for retransformation bias (Thorson and Kristensen 2016).

We checked model fits for convergence by confirming that the derivative of the marginal likelihood with respect to each fixed effect was sufficiently small (less than ~ 0.001) and that the Hessian matrix was positive definite, along with other model checks produced by the function *sdmTMB::sanity()*. We then checked goodness of model fit by computing Dunn-Smyth randomized quantile residuals (Dunn and Smyth 1996) and visualizing these using a quantile-quantile plot within the *DHARMa* R package (Hartig 2021). We also evaluated the distribution of

these residuals over space in each year, and inspected them for evidence of residual spatiotemporal patterns.

Model-based age composition methods

For model-based estimation of age compositions in the Bering Sea, we fitted observations of numerical abundance-at-age at each sampling location. This was made possible by applying a year-specific, region-specific (EBS and NBS) age-length key to records of numerical abundance and length composition. In subcategories (combinations of year, length, age, sex) that contained insufficient data, age composition was computed from length composition given a globally pooled age-length key. We computed these estimates in the *tinyVAST* R package (https://vast-lib.github.io/tinyVAST; Thorson et al., 2024), assuming a Poisson-link delta-model (Thorson 2018) involving two linear predictors, and a gamma distribution to model positive catch rates. We did not include any density covariates in estimation of age composition for consistency with models used in the previous assessments and due to computational limitations. We estimated the spatial range assuming geometric isotropy (i.e, spatial autocorrelation declines with equal rates over distance in all cardinal directions). We used the same extrapolation grid as implemented for abundance indices, but here we modeled spatial and spatiotemporal fields with a mesh with a coarser spatial resolution than the index model, here using 50 "knots". This reduction in the spatial resolution of the model, relative to that used for the abundance indices, was necessary due to the increased computational load of fitting multiple age categories. The estimates of abundance at age were bias-corrected with the same method as the abundance index. We implemented the same diagnostics to check convergence and model fit as those used for the abundance index.

Both the model-based bottom trawl index and age composition estimates were calculated using code in the AFSC GAP *model-based indices* GitHub repository (https://github.com/afsc-gap-products/model-based-indices). Design-based estimates of bottom trawl products were calculated using code in the *gapindex* R package (https://github.com/afsc-gap-products/gapindex).

References

Anderson, S.C., Ward, E.J., English, P.A. and Barnett, L.A.K., 2022. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. BioRxiv, pp.2022-03.

De Marco, P. and Nóbrega, C.C., 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PloS one*, *13*(9), p.e0202403.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J. and Münkemüller, T., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), pp.27-46.

Dunn, K.P., and Smyth, G.K., 1996. Randomized quantile residuals. Journal of Computational and Graphical Statistics 5, 1-10.

Hartig, F., 2021. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.0. http://florianhartig.github.io/DHARMa/.

Hulson, P.J.F. and Williams, B.C., 2024. Inclusion of ageing error and growth variability using a bootstrap estimation of age composition and conditional age-at-length input sample size for fisheries stock assessment models. *Fisheries Research*, *270*, p.106894.

O'Leary, C.A., Thorson, J.T., Ianelli, J.N. and Kotwicki, S., 2020. Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (*Gadus chalcogrammus*) stock assessment. Fisheries Oceanography, 29(6), pp.541-557.

Rohan, S.K., Barnett L.A.K., and Charriere, N., 2022. Evaluating approaches to estimating mean temperatures and cold pool area from AFSC bottom trawl surveys of the eastern Bering Sea. U.S. Dep. Commer., NOAA Tech. Mem. NMFS-AFSC-456, 42 p. https://doi.org/10.25923/1wwh-q418.

Thorson, J.T., 2018. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Canadian Journal of Fisheries and Aquatic Sciences. 75(9): 1369-1382. https://doi.org/10.1139/cjfas-2017-0266.

Thorson, J.T., 2019. Measuring the impact of oceanographic indices on species distribution shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnology and Oceanography, 64(6), pp.2632-2645.

Thorson, J.T., and Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research, 175, pp.66-74.

Thorson, J.T., Anderson, S.C., Goddard, P. and Rooper, C.N., 2024. tinyVAST: R package with an expressive interface to specify lagged and simultaneous effects in multivariate spatio-temporal models. arXiv preprint arXiv:2401.10193.