Bering Sea and Aleutian Islands Yellowfin Sole Stock Assessment Development for September 2026

Authors: Ingrid Spies and Meaghan Bryan

Summary:
A model transition from ADMB to SS3 was presented to Plan Team in September 2025 (NPFMC archive). The transition to SS3 was reviewed and accepted by CIE review team June 2026. Models presented in this document are implemented in SS3 (v.SS_V3_30_25).

Models presented for consideration:

All models presented in this document can be found here: https://github.com/afsc-assessments/BSAI_YFS, as well as r4ss output of all composition plots.

Model 25.0_2025:

- This is the base model, presented in September 2025. Documentation can be found on the NPFMC archive.
- 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 (Spies et al. 2024).
- Data partially updated for 2025.

The following models build off Model 25.0. All models have been updated with complete data for 2025. This includes the addition of two years of fishery weight-at-age data bootstrapped using catch-at-age methodology. An additional year of fishery ages are expected by October 1, 2026.

Model 25.0:

- This model is the same as Model 25.0_2025, but with updated data.

Model 25.0a: This model builds on Model 25.0, with:

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

Model 25.0b:

- This is the same as Model 25.0a but EBS survey catchability *q* incorporates EBS bottom temperature as an additive log-space environmental link (details below).

The data used in the models presented here include estimates of total catch, bottom trawl survey biomass estimates and their 95% confidence intervals, catch-at-age from the fishery, eastern Bering Sea survey bottom temperatures <100 m, and population age composition estimates from the bottom trawl survey. Weight-at-age and proportion mature-at-age from studies conducted during the bottom trawl surveys were also used. Estimates of fishery weight-at-age were based on catch-at-age methodology used in the walleye pollock assessment (Ianelli et al. 2024), following Kimura (1989) and modified by Dorn (1992).

	Model 25.0 (2025)	Model 25.0, 25.0a, 25.0b
Data source	Years	Years
Fishery catch	1954-2025*	1954-2025
Fishery age composition	1975-2023	1975-2024
Fishery weight at age	1975-2023	1975-2024
Survey biomass and standard error	1982-2024	1982-2025 (not 2020)
Bottom temperature	1982-2024	1982-2025 (not 2020)
Survey age composition	1979-2024	1979-2025 (not 2020)
Survey weight at age	1979-2024	1979-2025 (not 2020)
Age at maturity	combined 1992 and 2012 samples	

*Incomplete

CIE Review

In June 2026 the Center for Independent Experts conducted a CIE review for the yellowfin sole assessment for the Bering Sea and Aleutian Islands. The reviewers were: Patrick Cordue, Laurence Kell, and Joseph Powers. Reports can be found CIE Review Archive. We provide a detailed summary of each reviewer's comments below.

Patrick Cordue had several concerns, which are listed below.

1. Issue with growth in the model. He was critical of the empirical weight-at-age approach in the yellowfin sole assessment model; "…population model is scientifically invalid because growth is not really modelled (instead observations are used as biological parameters)." He states that growth modelling is required to ensure a valid population model.

2. Issue with EBS trawl series data. After examining methodology for trawl survey biomass estimation, he indicated that the EBS trawl series contains vessel effects which are structural rather than random. He notes that vessel is confounded with year; vessels used from 1982-1992 were different from vessels 1993-2024. He recommends estimating vessel effects, but this must be done within the two periods separately. He is skeptical this can be done in VAST, as can only be fitted as a random effect, and suggests applying year-specific corrections to the densities before input into VAST.

In addition, he supported the bridging analysis from the bespoke ADMB assessment to SS3, and stated that there should be no inadvertent differences between the two model implementations. His further suggestions included that the next assessment should continue to use bottom temperature and survey start date in the estimation of annual survey catchability (q). He ruled out 'double counting' temperature data as a covariate on catchability and in VAST biomass estimation. "The use of bottom temperature in VAST is not an issue because it is not used there to address annual variation in availability." Finally he recommended moving the assessment from Tier 1 to Tier 3; "The stock-recruitment relationship is not well estimated due to a lack of contrast. Therefore, this stock should reside in Tier 3 rather than Tier 1."

Authors' response:

1. The use of empirical weight-at-age is relatively commonplace in stock assessment models (e.g. Methot and Wetzel 2013, Methot et al. 2020), and simulations have shown that empirical weight at age can reduce estimation bias compared to modelling growth (e.g. Kuriyama et al. 2016). Future models will investigate incorporate growth as a function of bottom temperature.
2. We have shared the CIE review with survey personnel.
3. Two parameters on catchability in SS3 is problematic, as it is not allowed in the model structure. We have tested survey start date and bottom temperature as catchability covariates, and present those in Models 25.0a and 25.0b.
4. The assessment model will be conducted using a Tier 3 control rule for 2026. This was presented to the Plan Team in September 2025.

Laurence Kell indicated that the "SS3 assessment m25.0a for BSAI yellowfin sole is technically competent, draws on an appropriate and rich data set, and provides a coherent description of stock dynamics that is broadly consistent with survey and fishery evidence." He also noted that the transition from the ADMB model to SS3 had been carefully implemented and verified. In contrast to Patrick Cordue, he felt that the use of VAST/sdmTMB and tinyVAST for survey standardization and composition pre-processing was technically sound. He had several recommendations for the next few assessment cycles.

1. Standard SRR information-content diagnostics. Add likelihood profiles, residual plots, and simple regime or shift tests for alternative stock–recruitment forms and key parameters, and report how strongly the data inform B_{MSY} and F_{MSY} versus assumptions.
2. Basic M robustness sensitivities. Run and document a small set of age-structured and/or time-varying M scenarios (informed where possible by external life-history or ecosystem information) to show how B_{MSY}, F_{MSY}, stock status, and advice change relative to the scalar-M reference case.
3. Standardized convergence and identifiability reporting. For the operational SS3 run (m25.0a) and key sensitivities, routinely report maximum gradient, Hessian status, condition number, key parameter correlations, and a minimal likelihood-profile set for Tier-relevant parameters.
4. Clarify Tier 1 vs Tier 3 behavior. Present Tier 1 and Tier 3 reference points and advice outcomes side by side, with a qualitative narrative about differences in implied risk and precaution, explicitly flagging that risk-equivalence has not yet been demonstrated.
5. Robustness suite around productivity and process error. Build a limited set of structurally distinct, diagnostically acceptable models (e.g., alternative SRR forms, non-stationary productivity).

Authors' response:
Suggested diagnostics will be performed this assessment cycle where possible, and the remainder in future assessment cycles.

Joseph Powers supported transitioning to SS3 due to its improvement in consistency and flexibility, and utilizing a Tier 3 control rule. In addition, he supported analyses presented in the supporting assessment documents to justify these suggestions. He recommended that the candidate assessment model 25.0a is adequate for providing management advice.

Authors' response:
The 2026 assessment model will move forward as a Tier 3 assessment model in SS3 as suggested.

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 have presented this bridging exercise to the 2025 September Plan Team as well as the CIE review. The CIE review supported the SS3 model transition.

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. CIE reviewers support the shift to a Tier 3 control model.

Summary of Models

CIE review analysis provided confirmation that the use of bottom temperature in VAST calculations as well as a covariate for survey catchability does not represent 'double counting'. The use of bottom temperature in VAST is not an issue because it is not used in this context to address annual variation in availability. Additionally, we were faced with the challenge of modelling survey catchability q with only one environmental covariate, as it is constrained in SS3. Previously we had incorporated survey start date and bottom temperature, as well as their interaction. We incorporated survey start date in Model 25.0a (similar to Model 25.0 (2025)) and bottom temperature in Model 25.0b, and examined model fit to determine which covariate would be the best.

Model 23.0 presented in 2024 (Spies et al. 2024) incorporated time-varying catchability (q) that varied 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.

Model 25.0 catchability (q) estimates vary over time (t) as a function of the annual survey start date anomaly (S).

Eq. 2: $q_t = e^{-\alpha + \beta S}$,

where α is time-invariant and β is a based on the start date anomaly.

In SS3, the environmental link is coded as additive in log-space (201), with a flat, uninformative prior estimated in the phase 4.

Model 25.0b catchability (q) estimates also vary over time (t), but they vary as a function of annual bottom temperature anomaly (T) rather than survey start date/

Eq. 3: $q_t = e^{-\alpha + \beta T}$,

where α is time-invariant and β is a based on the bottom temperature anomaly.

Recruitment was assumed to follow mean recruitment, and recruits are defined as individuals of age 1. Recruitment was estimated within the model as the unfished recruitment level, or average number of age 1 fish entering the population annually in the absence of fishing pressure. Steepness was fixed to 1 and recruitment variation fixed to 0.6. Main recruitment deviations were estimated between 1954 and 2020. The value 2020 was selected because yellowfin sole do not exceed 50% selection until age 5. SS3 initializes the population assuming equilibrium and includes a plus group:

Eq. 4: $$N_{0,a,s} = \begin{cases} cR_0 e^{-aM_s} & a = 0 \; to \; 3A-1 \\ \sum_{a=A}^{3A-1} N_{0,a,s} + \frac{N_{0,3A-1,s} e^{-Ms}}{1-e^{-Ms}} & a > A \end{cases}$$

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

Model 25.0 initializes the population in 1954 as:

Eq. 6: $$n_{1954,a,s} = \begin{cases} 0.5e^{\bar{R}} & a = 1 \\ e^{\bar{I} + \varepsilon_{a,s}} & a > 1 \end{cases},$$

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.

Spawning biomass is calculated using:

Eq. 7: $SSB = \sum_a N_{fa} e^{-(M+F)t} W_a M_a$,

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.

Diagnostics

While a comparison of likelihoods is provided, composition data for each model was reweighted using Francis weighting methodology in SS3; therefore likelihoods are not directly comparable. Model 25.0 (2025) is also not directly comparable to the other configurations because it lacks complete 2025 data. Model 25.0 provides the lowest total likelihood, but the addition of *surveyISS* R package methodology to Models 25.0a and 25.0b provides a more robust and statistically sound weighting scheme for composition data because it incorporates observation variance through designed-based bootstrapping rather than simply relying on the number of hauls as a crude approximation (Hulson et al. 2024, Table 1). Model 25.0b improves the fit to the survey data by incorporating the survey bottom temperature to estimate the Eastern Bering Sea (EBS) survey catchability parameter, *q* (Table 1, Figure 4). Incorporating the complete 2025 data increased estimates for male natural mortality, spawning output and initial spawning biomass across all models (Table 1), but these values are comparable to parameter estimates in past assessments (Spies et al. 2024). Specifically, applying *sampleISS* for the survey sample size increased spawning output estimates during the final 10 years of the model time series (Model 25.0a, Figure 1). Including bottom temperature when estimating catchability lowered spawning biomass estimates, providing a better fit to survey indices (Model 25.0b, Figure 1). Model 25.0a yielded the highest unfished fraction in recent years, whereas Model 25.0b yielded the lowest (Figure 2). The standard deviation of normalized residuals (SDNR) for the fishery was between 0.83 – 0.90 for all models. Survey SDNR was lower for all models, between 0.74 – 0.82 (Table 1).

Recruitment patterns were generally similar across all models (Figure 3). Model 25.0b effectively links 100m bottom temperature to EBS survey catchability, correcting for high apparent CPUE. Without this environmental link, years with higher CPUE are incorrectly attributed to a larger population size. However, low SDNR for Model 25.0b (0.74/0.73, Table 1) may suggest overfitting the survey index. While the CIE review concluded that the use of mean annual 100 m EBS bottom temperature does not constitute double-counting in the model, data suggests the potential for overfitting. In sdmTMB, which was used to estimate the survey index, mean bottom temperature is specified as a density covariate rather than a catchability covariate. It is intended to be applied to differences in spatial density rather than factors confounded with catchability. Low temperature acts on yellowfin sole in multiple ways, including reducing mobility and therefore herding; delaying migration of fish returning from inshore spawning; and inhibiting growth (Nichol 1998, Somerton and Munro 2001, Nichol et al. 2019).

Kobe plots show highly similar patterns across all three configurations, strongly indicating that $F/F_{MSY} < 1.0$ and $SSB/SSB_{35\%} > 1.0$ (Figure 6). Additionally, all models passed the ss3diags runs test for CPUE (Figure 7). Mohn's rho was calculated for each model (Table 2) and ranged from 0.072 (Model 25.0a) to 0.0945 (Model 25.0b), with Model 25.0 in between at 0.087. For a medium- to long-lived species like yellowfin sole, these values fall well within the acceptable threshold range of -0.15 to 0.20.

Retrospective patterns from a 10-year peel of spawning stock biomass for all three models (Figure 8a, b, and c) and unfished spawning output (Figure 9a, b, and c) indicate that overall systemic bias is minimal. One-step-ahead (OSA) residuals for male and female age compositions were calculated for both the fishery and survey data using Model 25.0b (Figures 10 and 11). The OSA residuals revealed no obvious systematic patterns by sex. However, the fit to survey ages appears slightly misspecified, even though the selectivity patterns for survey and fishery follow a logical pattern (Figure 12), with survey selectivity higher at younger ages. Some adjustment to input sample sizes is needed to improve the fit to the survey age composition data, as there was a large discrepancy between effective and observed sample sizes (Figure 11).

Conclusion

No single model stands out as the definitive best option. Model 25.0 serves as the base model. Model 25.0a is improves upon it by incorporating the *surveyISS* data weighting methodology. Model 25.0b adds a temperature covariate on catchability, which may offer a further improvement, provided it does not cause overfitting. The three models presented here utilize Stock Synthesis (SS3), a widely accessible platform that offers operational flexibility and ensures a straightforward future transition to the Fisheries Integrated Modeling System (FIMS). Overall, these models successfully provide a Tier 3 harvest control rule, aligning with the recommendations of the 2026 Center for Independent Experts (CIE) reviewers.

References

Dorn, M.W., 1992. Detecting environmental covariates of Pacific whiting Merluccius productus growth using a growth-increment model. *Fish. Bull.*, *90*, pp.260-275.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R., and Punt, A. E. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. – ICES Journal of Marine Science, 72: 99–110.

Ianelli, J., T. Honkalehto, S. Wassermann, A. McCarthy, S. Steinessen, C. McGilliard, and E. Siddon. 2024. Assessment of walleye pollock in the eastern Bering Sea. North Pacific Fishery Management Council, Anchorage, AK. Available here.

Kimura, D.K. 1989. Variability in estimating catch-in-numbers-at-age and its impact on cohort analysis. In R.J. Beamish and G.A. McFarlane (eds.), Effects on ocean variability on recruitment and an evaluation of parameters used in stock assessment models. Can. Spec. Publ. Fish. Aq. Sci. 108:57-66.

Kuriyama, P.T., Ono, K., Hurtado-Ferro, F., Hicks, A.C., Taylor, I.G., Licandeo, R.R., Johnson, K.F., Anderson, S.C., Monnahan, C.C., Rudd, M.B. and Stawitz, C.C., 2016. An empirical weight-at-age approach reduces estimation bias compared to modeling parametric growth in integrated, statistical stock assessment models when growth is time varying. *Fisheries Research*, *180*, pp.119-127.

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

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., 1998. Annual and between-sex variability of yellowfin sole, *Pleuronectes aspera*. *Fishery Bulletin*, *96*, pp.547-561.

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.

Spies, I., Bryan, M., Barnett, L., Markowitz, E., Oyafuso, Z., and Siddon, E. 2024. Assessment of the Yellowfin Sole Stock in the Bering Sea and Aleutian Islands. North Pacific Fishery Management Council, Anchorage, AK. Available from https://www.npfmc.org/library/safe reports/.

Somerton, D.A. and Munro, P., 2001. Bridle efficiency of a survey trawl for flatfish. *Fishery Bulletin*, *99*(4), pp.641-653.

Table 1. Comparison of likelihoods and key parameters among Models 25.0, 25.0a, 25.0b, as well as Model 25.0 (2025), which was presented in 2025 and contains partial data for 2025. Note that Francis weighting was performed on each model separately in SS3; therefore, likelihoods are not directly comparable.

Label	M25.0 (2025)	M25.0	M25.0a	M25.0b
Data	Partial 2025	--------------Complete 2025--------------		
Catchability	Start date	Start date	Start date	Temperature
N. samples for agecomp	Nhauls	Nhauls	SurveyISS	SurveyISS
Number of parameters	235	235	235	235
Total likelihood	207.66	327.06	399.16	382.43
Survey likelihood	4.55	-5.36	-10.37	-20.97
Age comp likelihood	197.97	319.81	393.19	389.60
Parm priors likelihood	0.09	0.04	0.02	0.04
Recr_Virgin_billions	2.39	2.67	2.73	2.61
SR_LN(R0)	14.69	14.80	14.82	14.78
R0_for_SRR_bench	2,385,820	2,669,580	2,734,870	2,612,490
SR_BH_steep	1.00	1.00	1.00	1.00
NatM_uniform_Fem_GP_1	0.120	0.120	0.120	0.120
NatM_uniform_Mal_GP_1	0.123	0.132	0.132	0.132
L_at_Amax_Fem_GP_1	38.034	38.034	38.034	38.034
L_at_Amax_Mal_GP_1	34.030	34.030	34.030	34.030
VonBert_K_Fem_GP_1	0.137	0.137	0.137	0.137
VonBert_K_Mal_GP_1	0.161	0.161	0.161	0.161
SSB_Virgin	1,372,200	1,535,410	1,572,960	1,502,580
Bratio_2025	NA	0.77	0.78	0.71
$SDNR_{fishery}$		0.83/0.86	0.86/0.91	0.85/0.89
$SDNR_{survey}$		0.82/0.80	0.76/0.75	0.74/0.73

Table 2. Mohn's rho calculated using *r4ss::SSmohnsrho*, based on the afsc spawning stock biomass formulation of Mohn's rho (Hurtado-Ferro et al. 2015).

Model	Rho
25.0	0.087
25.0a	0.072
25.0b	0.095

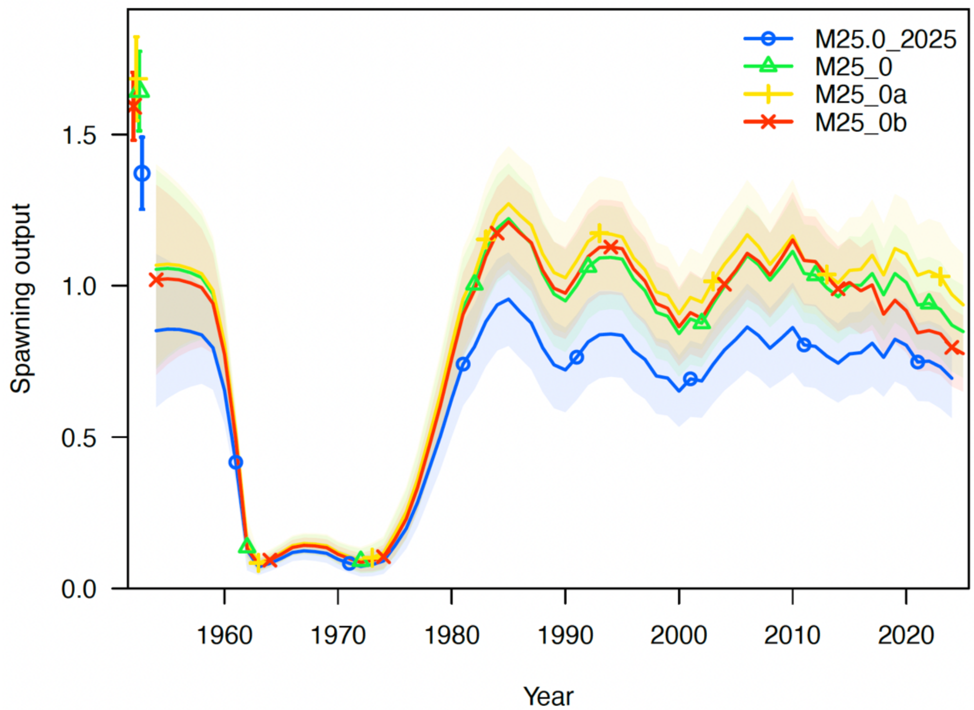


Figure 1. Time series of SSB_t/SSB_0 for Models 25.0 (2025) 25.0, 25.0a, and 25.0b.

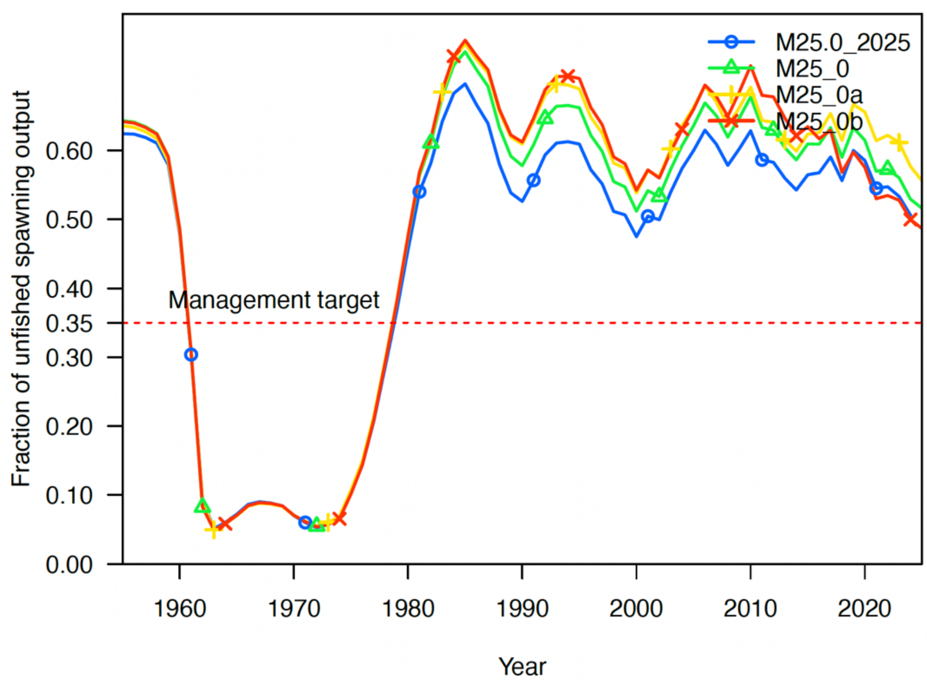


Figure 2. Timeseries of the fraction of unfished spawning biomass relative to management target (here SSB35%) for Models 25.0 (2025), 25.0, 25.0a, and 25.0b.

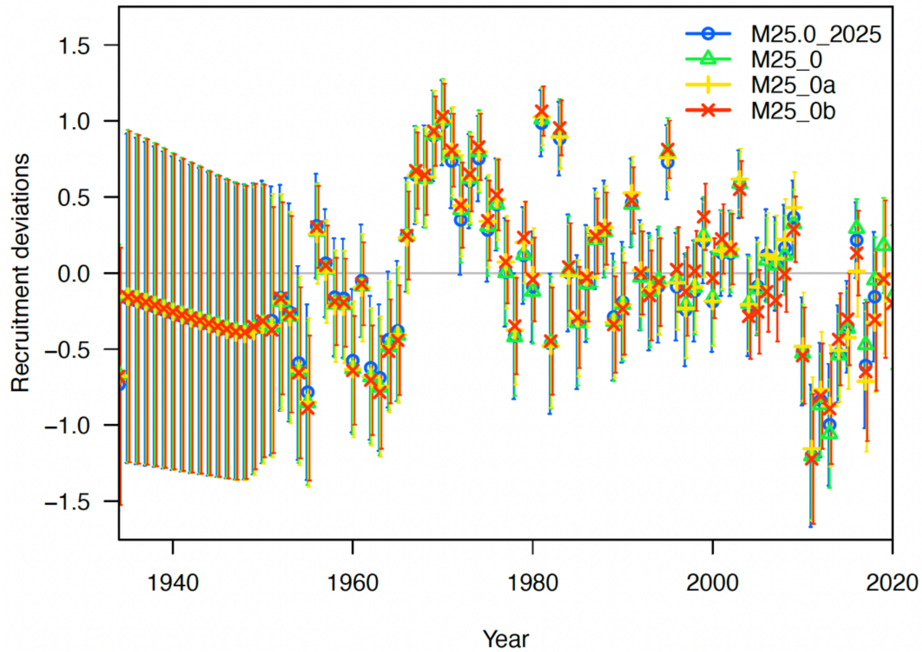


Figure 3. Time series of age-1 recruitment with 95% confidence intervals for Models 25.0 (2025), 25.0, 25.0a, and 25.0b.

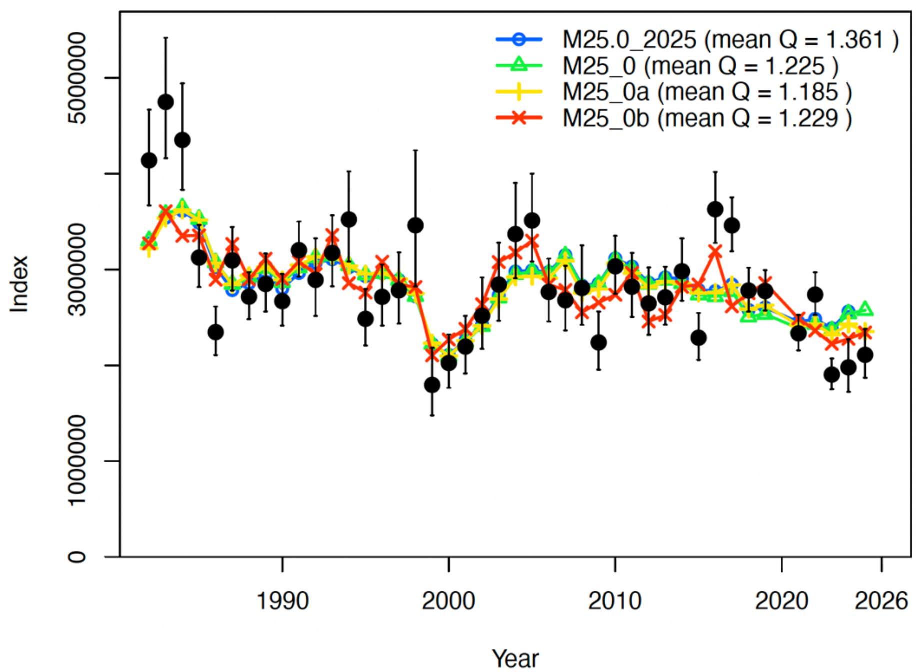


Figure 4. Time series of Models 25.0 (2025), 25.0, 25.0a, and 25.0b fit to the survey index, 1982 – 2025.

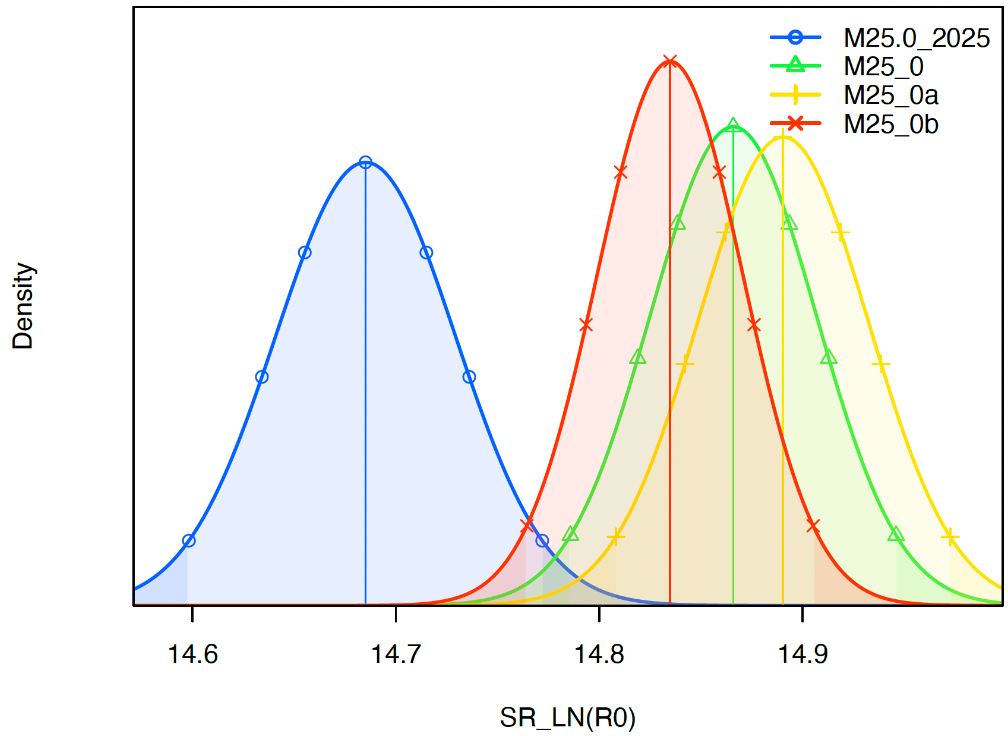


Figure 5. Model (25.0 (2025), 25.0, 25.0a, and 25.0b) density plots of estimates of log initial recruitment.

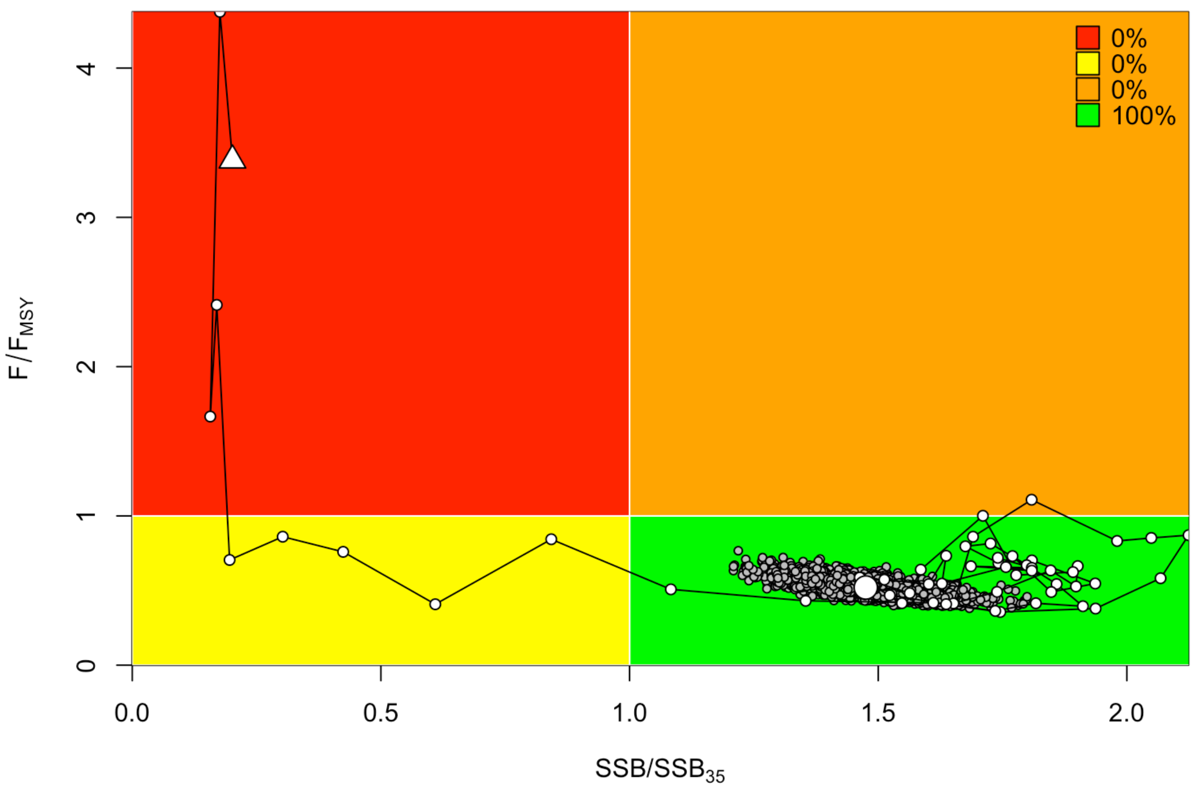


Figure 6a. Kobe plot for Model 25.0 with 5,000 Monte Carlo simulations. Years prior to 1970 were excluded because F/F_{MSY} exceeded 30 for those years. Triangle represents 1970, circle represents 2025.

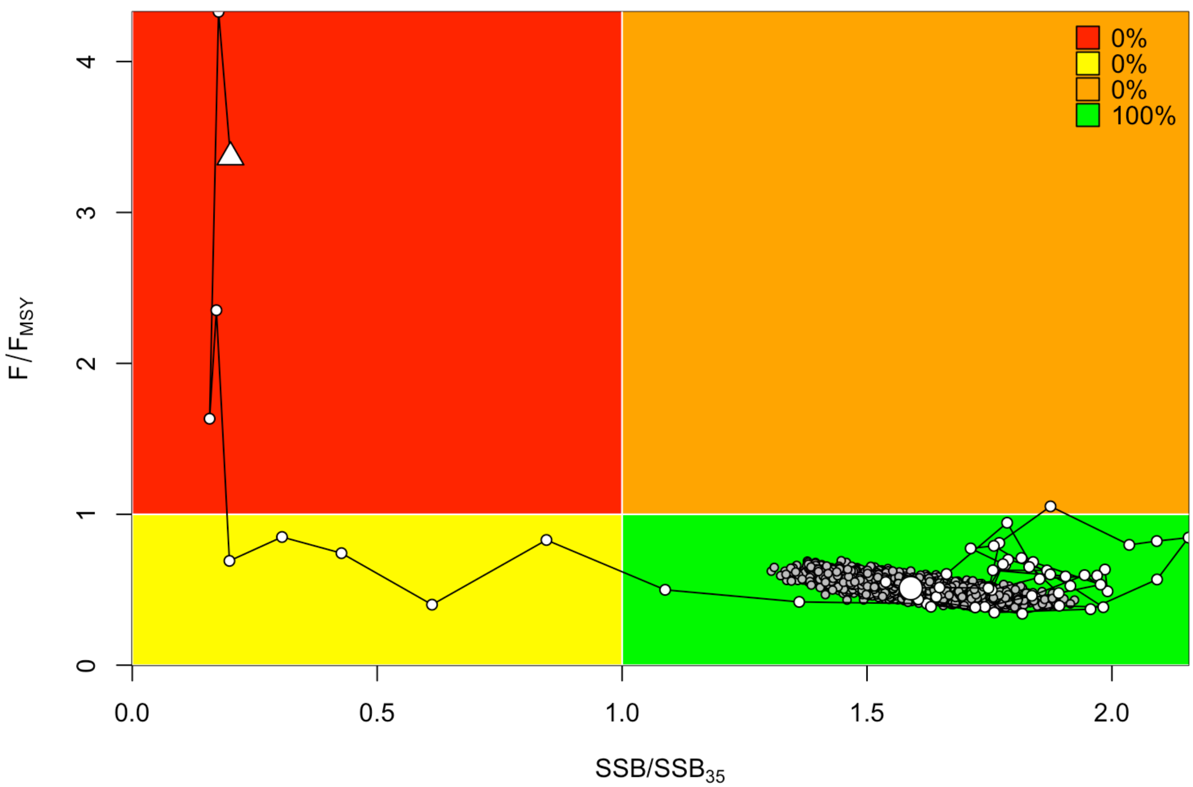


Figure 6b. Kobe plot for Model 25.0a with 5,000 Monte Carlo simulations. Years prior to 1970 were excluded because F/F_{MSY} exceeded 30 for those years. Triangle represents 1970, circle represents 2025.

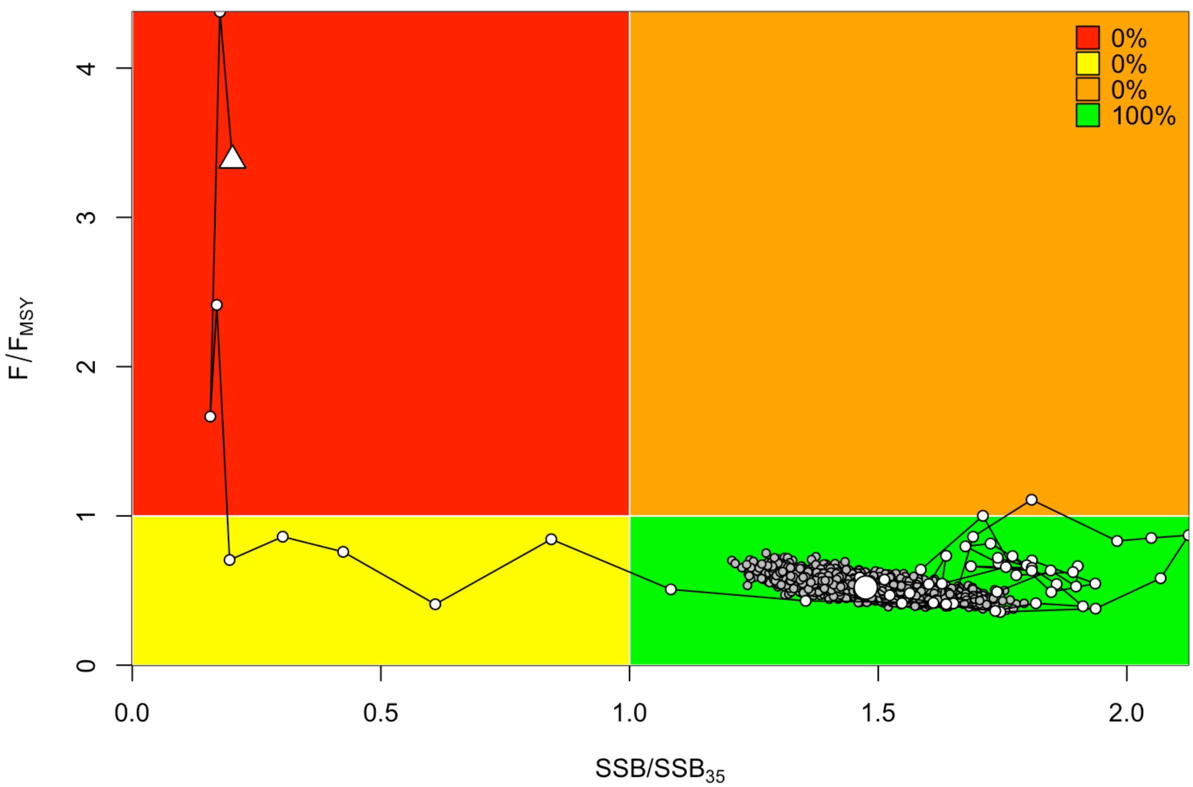
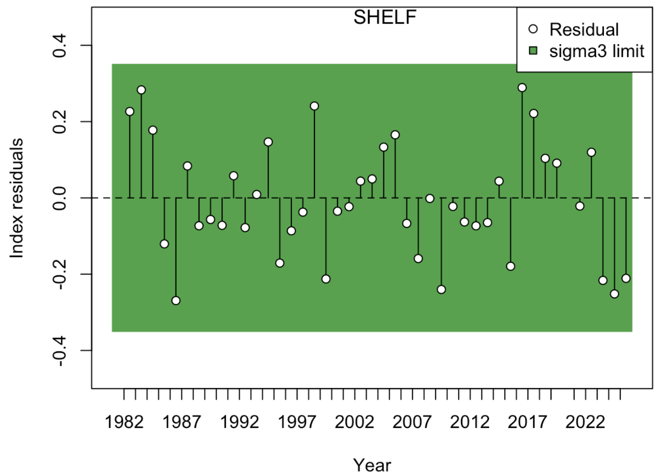
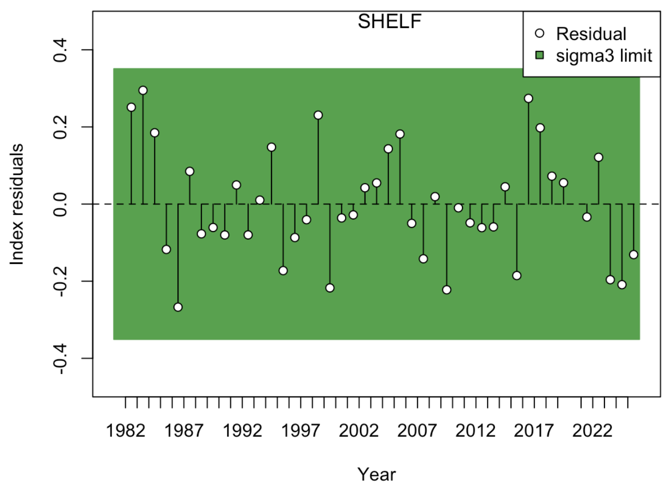


Figure 6c. Kobe plot for Model 25.0b with 5,000 Monte Carlo simulations. Years prior to 1970 were excluded because F/F_{MSY} exceeded 30 for those years. Triangle represents 1970, circle represents 2025.

M25.0



M25.0a



M25.0b

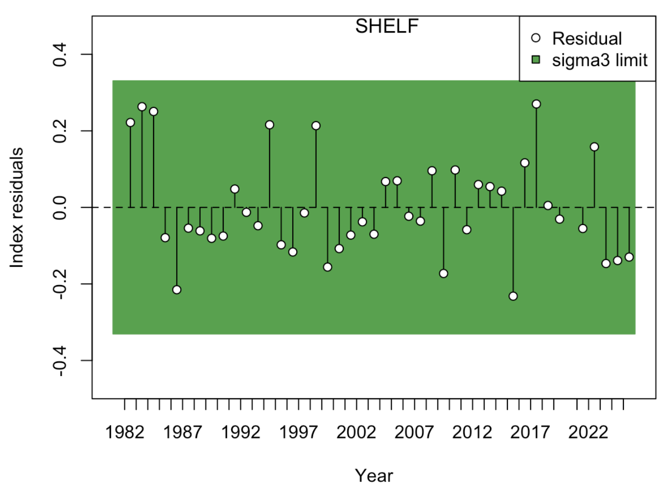


Figure 7. Residual plots for CPUE ss3diags runs test for Models 25.0, 25.0a, and 25.0b.

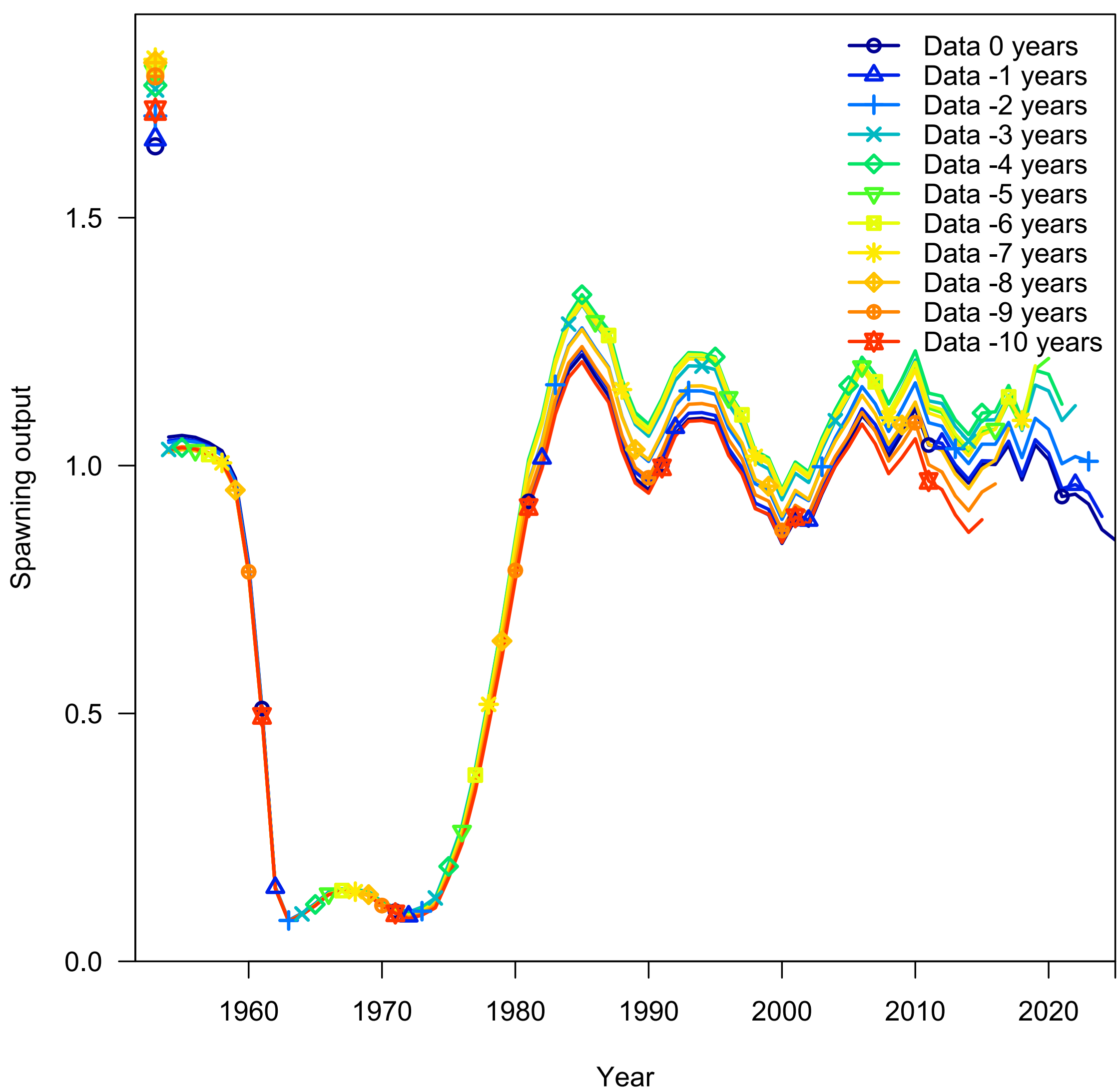


Figure 8a. Ten-year retrospective peel of spawning biomass for Model 25.0.

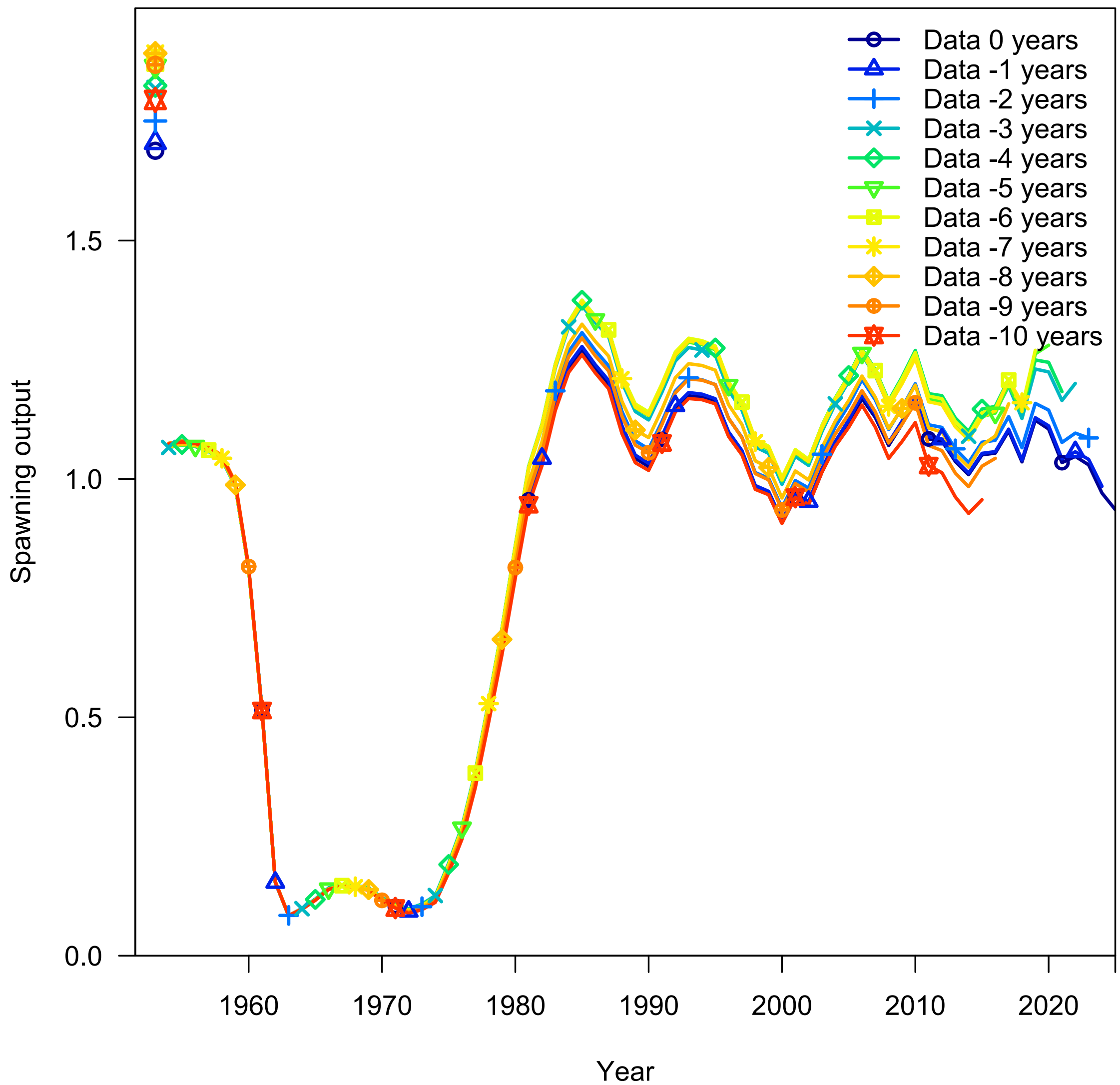


Figure 8b. Ten-year retrospective peel of spawning biomass for Model 25.0a.

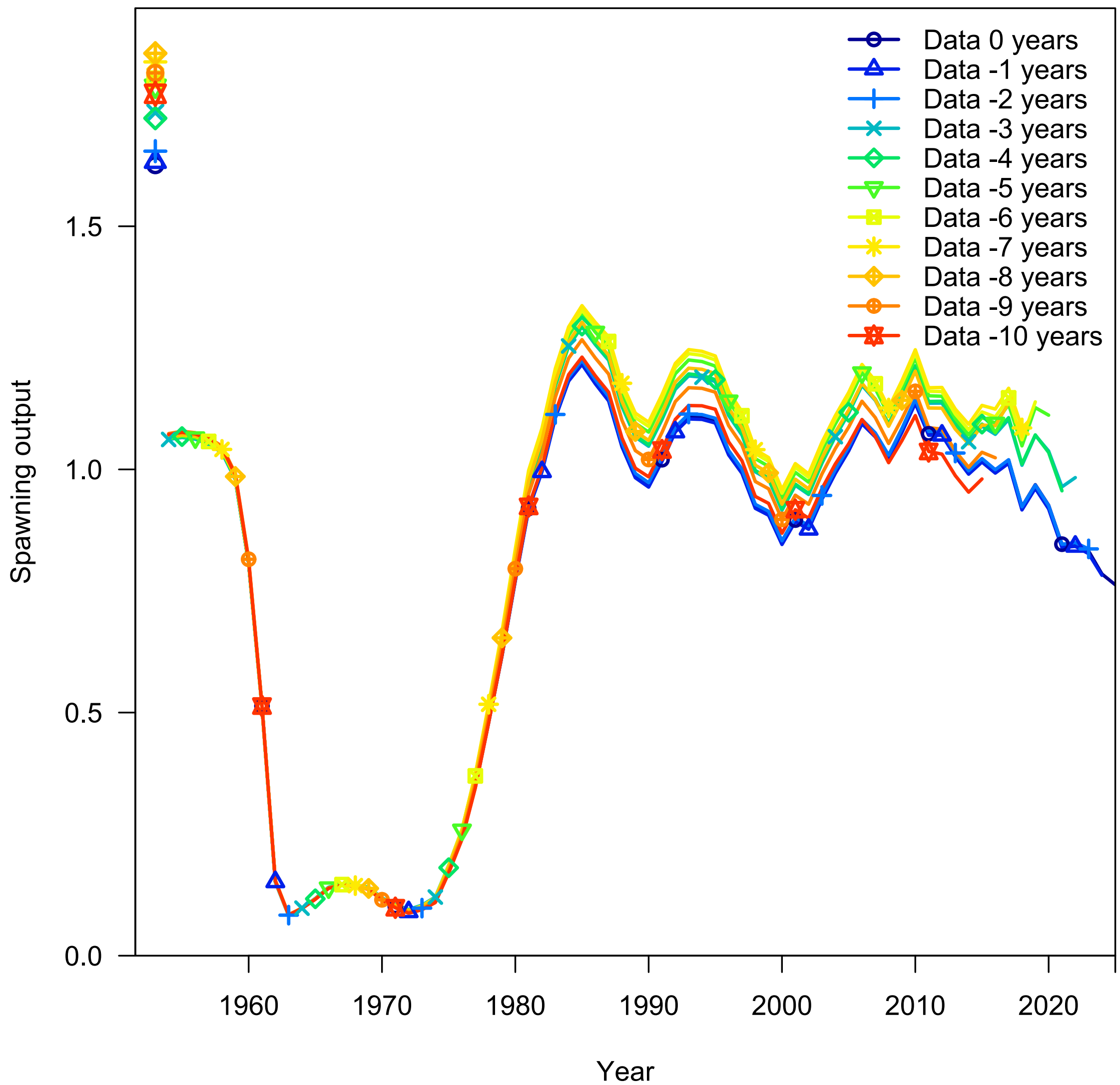


Figure 8c. Ten-year retrospective peel of spawning biomass for Model 25.0b.

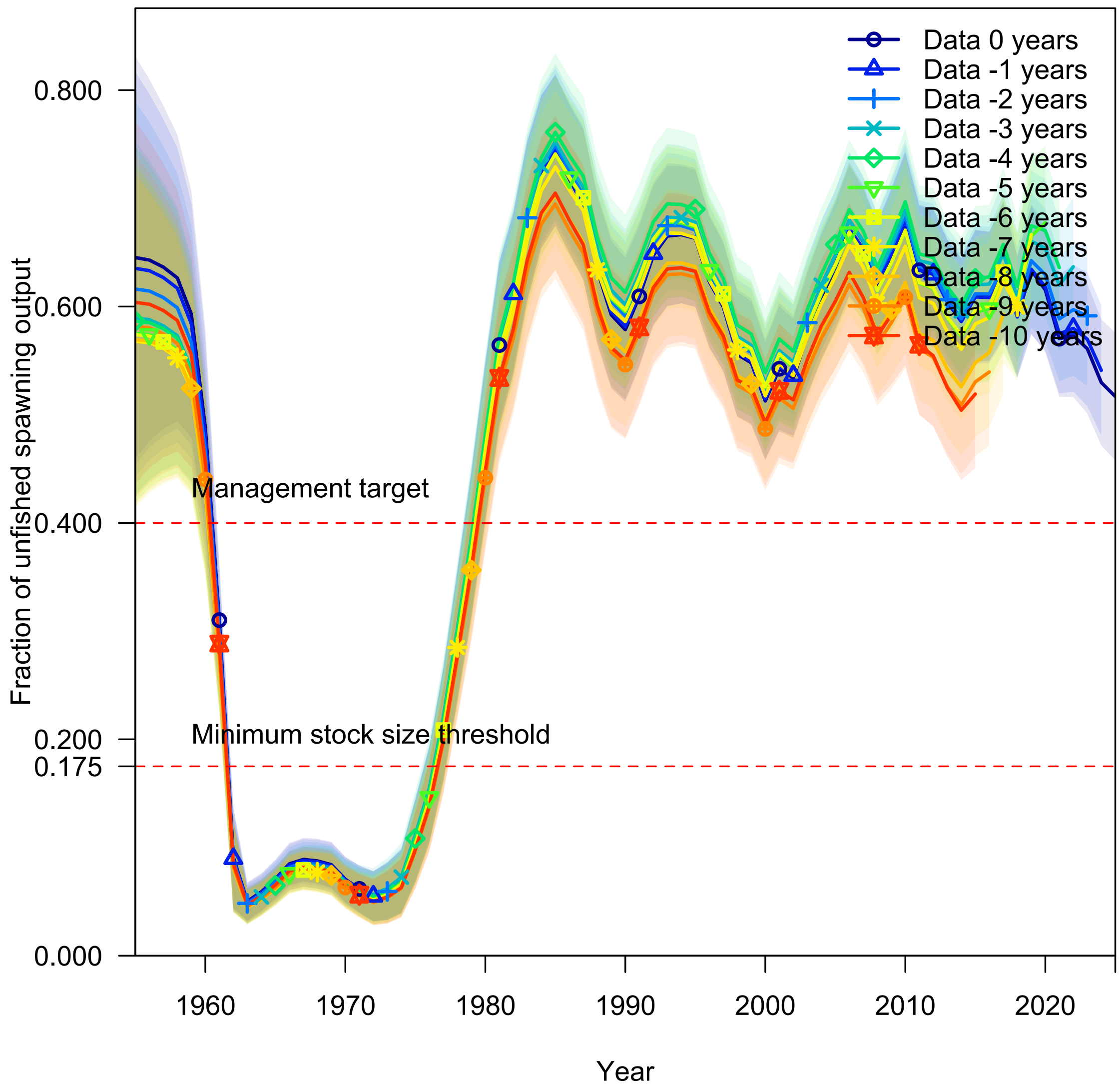


Figure 9a. Ten-year retrospective peel of the fraction of unfished spawning output for Model 25.0.

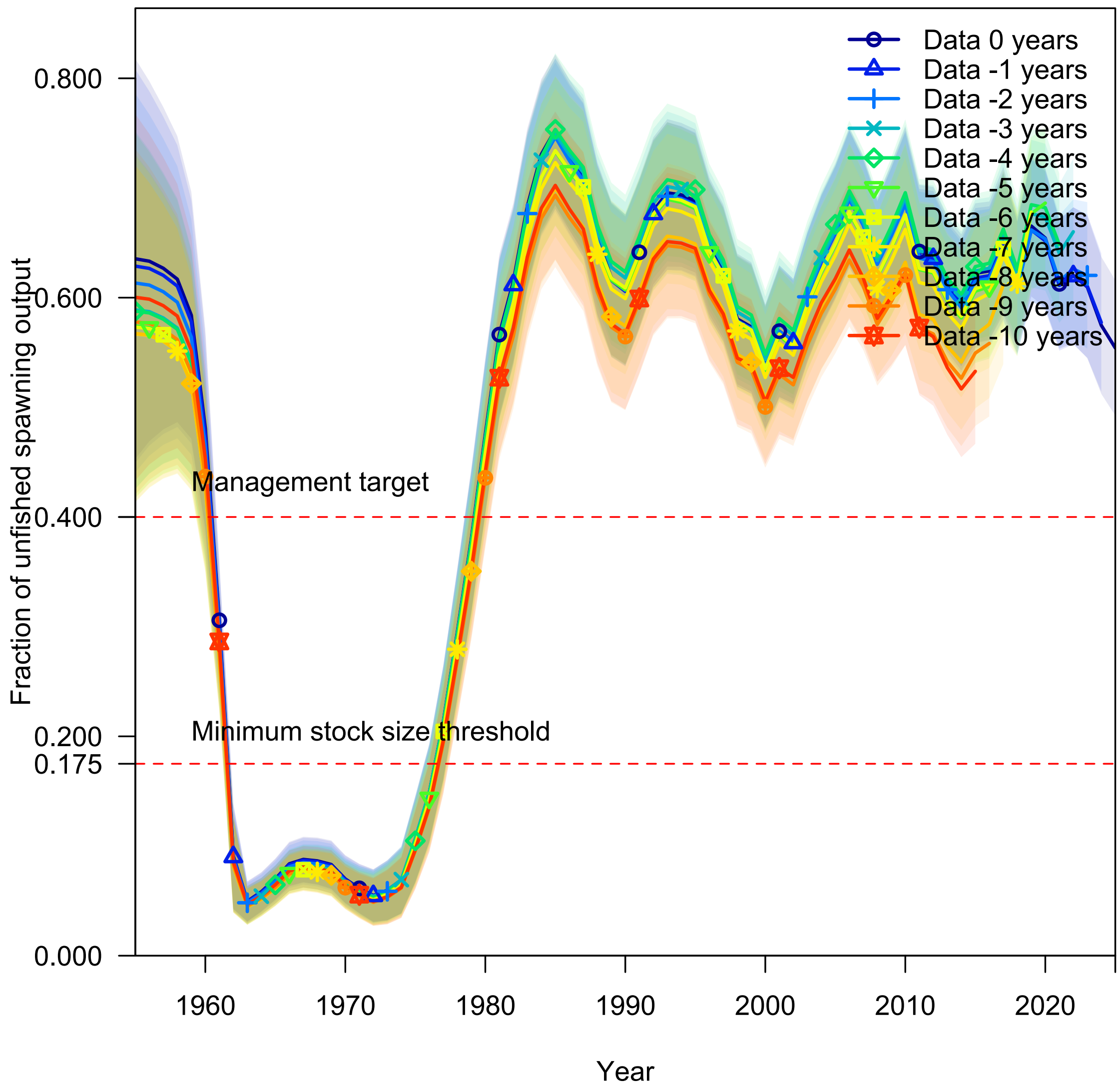


Figure 9b. Ten-year retrospective peel of the fraction of unfished spawning output for Model 25.0a.

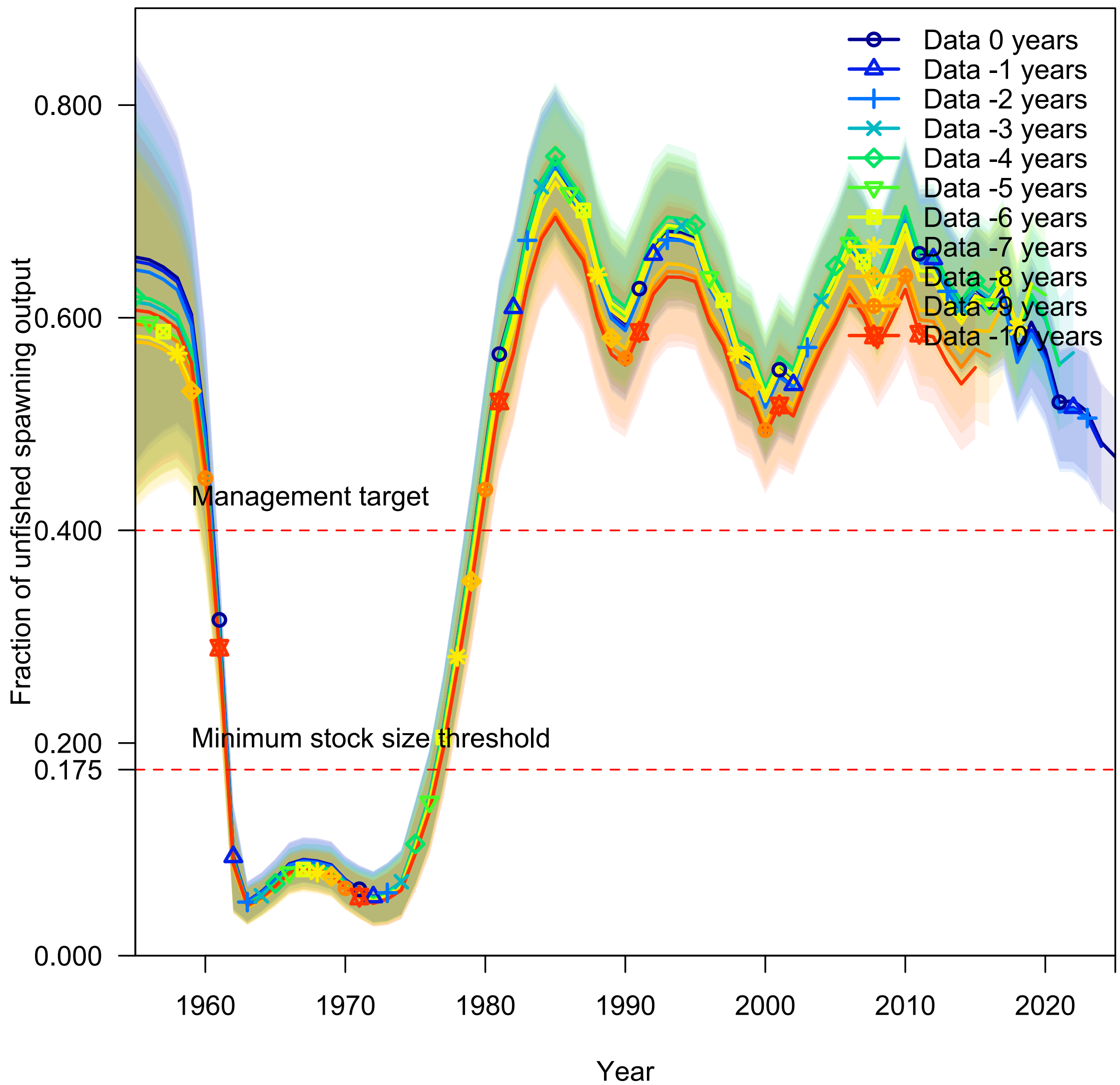


Figure 9c. Ten-year retrospective peel of the fraction of unfished spawning output for Model 25.0b.

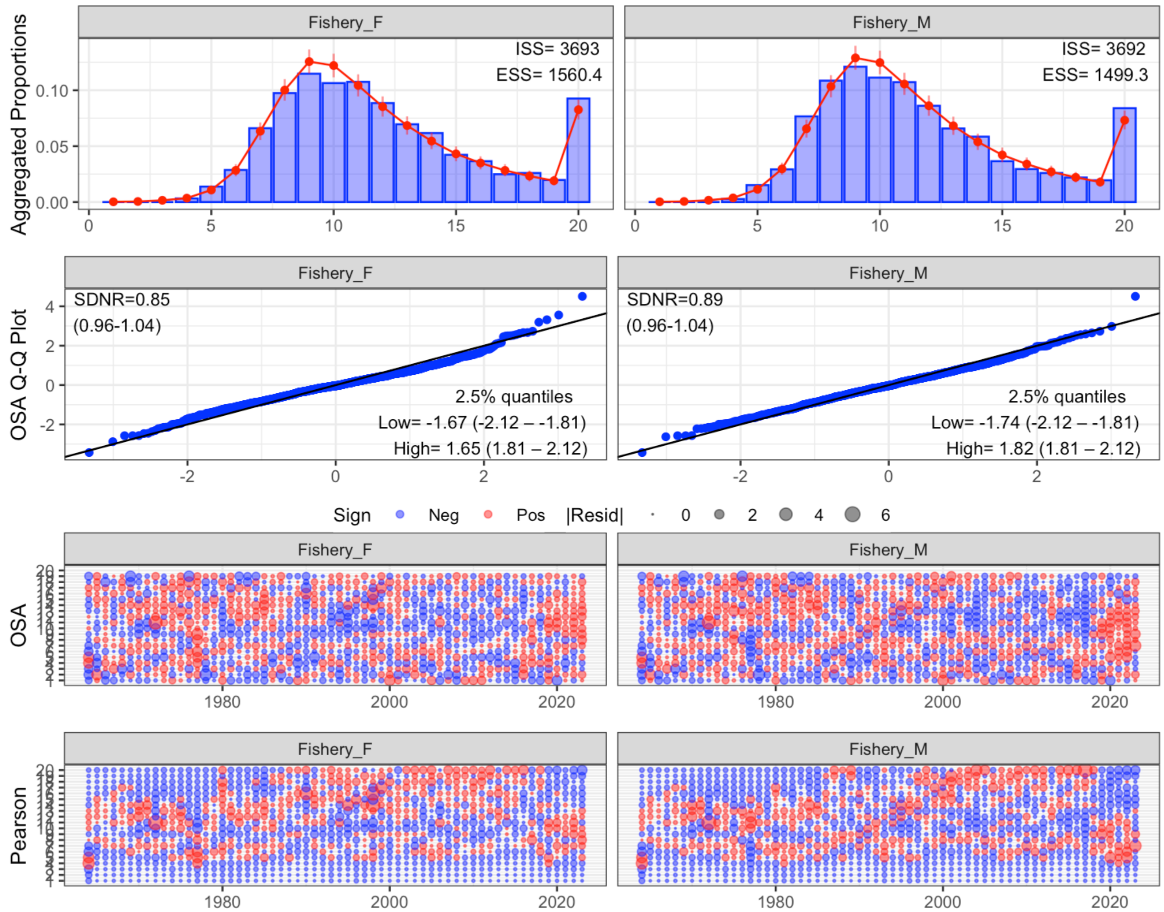


Figure 10. One-step ahead residuals for eastern Bering Sea yellowfin sole fishery, females (left), males (right) based on Model 25.0b. Lower panels depict OSA and Pearson residuals (blue positive, red negative) with year on the x-axis and age on the y-axis. Standardized deviation of normalized residuals: 0.85 females, 0.89 males.

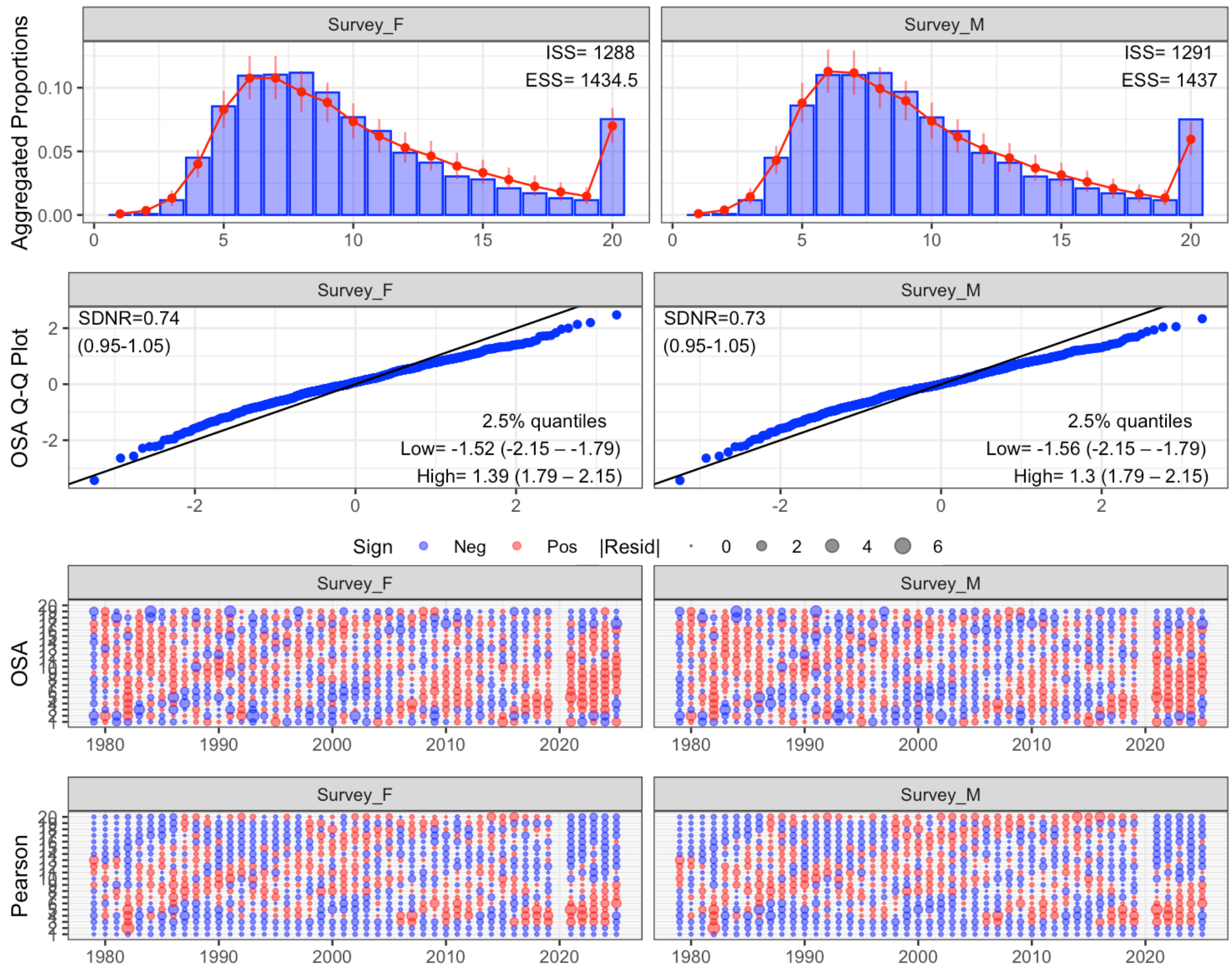


Figure 11. One-step ahead residuals for eastern Bering Sea yellowfin sole survey, females (left), males (right) based on Model 25.0b. Lower panels depict OSA and Pearson residuals (blue positive, red negative) with year on the x-axis and age on the y-axis. Standardized deviation of normalized residuals: 0.74 females, 0.73 males.