

Alternative late-period CPUE hypothesis & implications for the stock assessment of blue sharks in the North Pacific¹

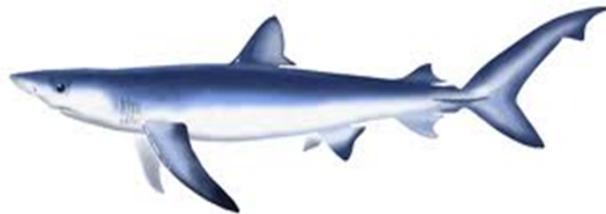
Nicholas D. Ducharme-Barth², Steven L. H. Teo³, Mikihiro Kai⁴, Felipe Carvalho²

2. NOAA National Marine Fisheries Service, Pacific Islands Fisheries Science Center, 1845 Wasp Boulevard, Building 176, Honolulu, Hawaii, USA 96818

3. NOAA Fisheries, Southwest Fisheries Science Center, La Jolla, California USA

4. Fisheries Resources Institute, Japan Fisheries and Education Agency, 5-7-1 Orido, Shimizu, Shizuoka, 424-8633, Japan

Email: nicholas.ducharme-barth@noaa.gov



¹Working document submitted to the ISC Shark Working Group Workshop, 26-28 April 2022 (JST), Web-meeting.
Document not to be cited without author's permission.

Contents

1	Introduction	3
2	Methods	5
2.1	Dynamic factor analysis	5
2.2	Stock assessment modeling	7
2.2.1	<i>data.ss</i> file	7
2.2.2	<i>control.ss</i> file	7
2.2.3	Sensitivity analyses	8
2.2.4	Diagnostics	9
3	Results	9
3.1	DFA	9
3.2	Stock Synthesis	10
3.2.1	Sensitivity analyses	10
3.2.2	Candidate model diagnostics	11
3.2.2.1	Convergence	11
3.2.2.2	Fit	11
3.2.2.3	ASPM	11
3.2.2.4	Retrospectives	11
3.2.2.5	R0 profile	12
3.2.3	Candidate model results	12
4	Discussion	12
5	Recommendations	13
6	Tables	18
7	Figures	21

Summary

This working paper details the approach used to construct an alternative CPUE hypothesis for the late model period from three candidate indices (Hawaii longline index, Taiwanese longline index, and Japanese research and training vessel index) using Dynamic Factor Analysis (DFA). Sensitivity analyses to the construction of the DFA composite index, along with how the DFA composite index is incorporated into the Stock Synthesis model are described. The effect of including the DFA composite index in the Stock Synthesis model appeared to be sensitive to the choice of fishery selectivity and the addition of extra variance for the index. DFA trends and Stock Synthesis outputs appeared relatively robust to changes to the input CV for each of three input indices. A baseline DFA composite index model was identified which assumed selectivity mirroring with the Japanese Enyo Deep longline, added extra variance to ensure an average CV of 0.2 for the index, and assumed the input CV for each index was the same as specified in the Stock Synthesis model fitting to the Japanese Kinkai shallow late index. Comparisons between models fitting to the baseline DFA composite index and the Japanese Kinkai shallow late index indicated marginally better fits to common likelihood components for the DFA composite index model. However, it was difficult to exclude either CPUE hypothesis based on model diagnostics alone. Both models showed retrospective bias which was identified in the DFA composite index model to be caused by the large observed sample size for the Taiwanese small scale longline length composition data in 2018 and 2020. Six additional sensitivities were run using the baseline DFA composite index model to explore eliminating the retrospective pattern via down-weighting or removing the length composition data in 2018 and 2020 for the Taiwanese small scale longline, and also investigate the effect of adding extra variance to ensure an average CV of 0.4. Evaluation of the six sensitivity models fitting to the baseline DFA composite index model did not identify a clear best model based on fits to the data or model diagnostics. Given this, and the clear difference in population trajectory and scale between the two alternative late CPUE hypotheses, a way forward would be to consider a model ensemble where each alternative CPUE hypothesis is weighted equally.

1 Introduction

Blue shark, *Prionace glauca*, are a productive, wide-ranging pelagic shark, that are globally distributed in temperate and sub-tropical waters (Nakano and Stevens, 2008). Blue shark are active in the epipelagic environment (<220m; Moyes et al., 2006; Heard et al., 2018; Fujinami et al., 2021) and are often captured in longline fishing operations, either as a targeted (Kai, 2021a) or bycatch species (Campana, 2016). Within the Pacific Ocean blue shark are considered to be comprised of two stocks with the equator dividing northern and southern hemisphere populations (King et al., 2015; Taguchi et al., 2015). Within the north Pacific Ocean, the International Scientific Committee (ISC) routinely conducts stock assessments to evaluate the population status and impacts of fishing activity on blue shark.

North Pacific Ocean blue shark was previously assessed in 2017 with the management advice based on the results of a single Stock Synthesis (Methot and Wetzel, 2013) *base-case* model (ISC, 2017). This base-case model was developed in an iterative process to produce a model consistent with data assumed to be most representative of the stock. Fisheries-dependent indices were developed to provide information on the population trend in two model periods (*early* ≤ 1993 , and *late* ≥ 1994) based on a change in species-specific data reporting in Japanese data. Only a single index based on the Japanese shallow-set longline was available in the early period, multiple indices were available for the late period. These included the Japanese shallow-set longline, Hawaii longline observer index, Taiwanese longline observer index, Pacific Community Regional Observer Program index, and the Mexican longline index. Fitting to these indices individually resulted in different population trajectories in the recent period from those indicated by the Japanese shallow-set index. This result was the topic of discussion at the Western and Central Pacific Fisheries Commission (WCPFC) Scientific Committee (SC) meeting (WCPFC, 2017).

In addition to the four alternative late period indices used in the 2017 ISC blue shark stock assessment, a fifth index was prepared from Japanese research and training vessel longline data for the current assessment (Kai, 2021b). Three of these indices; the Hawaii longline, Taiwanese large-scale longline, and Japanese research and training vessel; come from fisheries which predominantly target tunas typically via deep setting operations (Ducharme-Barth and Vincent, 2020). Though the Hawaii longline index is based on observer data which may more accurately reflect the rate of blue shark encounter relative to logbook records (Ducharme-Barth et al., 2022), use of this index was discounted in the previous assessment due to its reduced spatial extent. However, the Hawaii longline index appears to show similar trends in blue shark to the Taiwanese longline and Japanese research and training vessel indices, which have a much broader spatial extent (Kai, 2019; Liu et al., 2021). The apparent consistency in trend across three fisheries could present an alternative CPUE hypothesis from the Japanese shallow-set longline which seasonally targets blue shark.

Fitting to combined indices has seen increasing use in assessments conducted by tuna Regional Fisheries Management Organizations (RFMOs) including the International Commission for the Conservation of Atlantic Tunas (ICCAT) and the WCPFC (Ducharme-Barth et al., 2020; ICCAT, 2020, 2021), where the benefits include a reduction in data conflict between indices from individual fleets with varying levels of spatial and temporal coverage. When data from all fisheries is made available for analysis *a priori*; a combined, joint index is typically constructed from a single, unified standardization model (Hoyle et al., 2019; Ducharme-Barth and Vincent, 2020; Kitakado et al., 2021). *Post hoc* analytic techniques including hierarchical Bayesian analysis (Conn, 2010) and Dynamic Factor Analysis (Peterson et al., 2021) also exist for deriving a combined, composite index from existing indices or survey data.

Dynamic Factor Analysis (DFA) is frequently used for extracting common temporal trends from time series data (Zuur et al., 2003a). Additionally, DFA techniques can be applied to the oftentimes short, noisy time series with missing observations that are frequently encountered in analyses of

fisheries dependent data sets (Zuur et al., 2003b; Azevedo et al., 2008). Though DFA is generally able to reconcile conflicting trends and reduce statistical noise, it can be sensitive to the input variances associated with the conflicting trends (Peterson et al., 2021).

This working paper details the development of a composite index using DFA from the Hawaii longline index, Taiwanese longline index, and Japanese research and training vessel index. This paper also investigates key uncertainties with the composite DFA index: i) sensitivity to the input CV for each candidate index, ii) sensitivity to the assumed selectivity for the new composite index, and iii) sensitivity to the degree of extra variance added to the composite index. Lastly, we explore the performance and fit of six composite DFA index candidate models using standard diagnostics: residual analysis, age-structured production model (ASPM), retrospective analysis, and R0 profile

2 Methods

This working paper is divided into two sections. The first section (2.1) describes the DFA analysis used to combine the Hawaii longline (2002-2020; Ducharme-Barth et al., 2022), (2004-2020; Taiwanese longline Liu et al., 2021), and Japanese research and training vessel indices (1994-2020; Kai, 2019, 2021b) into a single combined, composite index spanning 1994-2020. The second section (2.2) describes how the composite index was incorporated into the existing Stock Synthesis (Methot and Wetzel, 2013) model for blue shark (Kai et al., 2022c,b,a).

2.1 Dynamic factor analysis

The DFA approach taken in this working paper was implemented in R using the state-space multivariate autoregressive modelling package *MARSS* (Holmes et al., 2012, 2021a,b), and replicated the methodology used by Peterson et al. (2021) to i) re-scale the indices for analysis and ii) fit the DFA model described by the following equations:

$$\mathbf{y}_t = \mathbf{\Gamma}\alpha_t + \boldsymbol{\epsilon}_t \text{ where } \boldsymbol{\epsilon}_t \sim MVN(0, \mathbf{H}) \quad (1)$$

$$\alpha_t = \alpha_{t-1} + \eta_t \text{ where } \eta_t \sim N(0, 1) \quad (2)$$

where in the case of the current blue shark analysis \mathbf{y}_t is a vector of length $j = 3$ representing the observations at time t for each index \mathbf{I}_j , α_t is the common trend across indices estimated as a random walk with *Normal* error η_t , $\mathbf{\Gamma}$ is a 3×1 matrix containing the factor loadings for each index, and \mathbf{H} is the 3×3 covariance matrix for the observations where diagonal elements were equal to the average CV for each index \mathbf{I}_j and off-diagonal elements were assumed to be zero (assuming independence across time series). In defining \mathbf{H} , the average CV was defined as specified in the Stock Synthesis model, including the additive variance adjustment factor, and rounded to the second decimal place. This resulted in average input CVs for the DFA analysis of 0.4 for the Hawaii longline index, 0.2 for the Taiwanese longline index, and 0.13 for the Japanese research and

training vessel index (shown in Figure 1).

Traditional DFA analysis necessitates de-trending and standardizing the indices by their standard deviation (converting indices to a *z-score*) before log-transforming the index in order to meet the DFA model assumptions of stationarity and a *Normal* error distribution. However, following such a transformation, the derived DFA trend is unit-less with mean 0, and unable to be back-transformed to the original units and scale. Peterson et al. (2021) developed an innovative approach to re-scaling the indices in such a way that the assumptions of the DFA model are met, and to allow for the recovery of the original scale and error structure of the index. An R code tutorial for implementing DFA following the re-scaling approach used by Peterson et al. (2021) is available at the author’s personal GitHub repository: https://github.com/cassidydpeterson/DFA_Simulation_and_Assessment/blob/main/DFA_TUTORIAL.R. Briefly, the process is described in the following 7 steps:

1. Multiply each index I_j by a constant c_j .
2. Apply a log-transformation to each index.
3. Center and de-mean each index by subtracting and dividing each index by its respective mean.
4. Calculate a global standard deviation (GSD) across all indices.
5. Divide each index by the GSD.
6. Fit the DFA model described in equations 1 & 2.
7. The DFA trend α_t and associated uncertainty were back-transformed and bias corrected according to equations 3-5.

$$\hat{\alpha}_t = \exp\left(GSD \times \alpha_t + \left(\frac{GSD \times SE_t^2}{2}\right)\right) \quad (3)$$

$$SD(\hat{\alpha}_t) = GSD \times SD(\alpha_t) \quad (4)$$

$$CV(\hat{\alpha}_t) = \frac{SD(\hat{\alpha}_t)}{\hat{\alpha}_t} \quad (5)$$

As mentioned previously, Peterson et al. (2021) developed an approach to ensure that the DFA trend α_t could be back-transformed such that the original scale was recoverable. This is accomplished in step 1 of the previous process by multiplying each index I_j by a constant c_j . The constant c_j can be any value such that the following constraints were satisfied:

1. The mean of $\log(I_j \times c_j) > 0$.
2. GSD was small ($GSD < 0.5$).
3. The SD of each index following division by the GSD (previous step 5) was $\cong 1$.
4. $\hat{\alpha}_t$ should follow changes in magnitude consistent with those of the survey inputs.

Peterson et al. (2021) iteratively tuned the constants c_j such that all four constraints were satisfied. In the current analysis a genetic optimization technique was implemented in R using the *rgenoud* package (Mebane and Sekhon, 2011) to solve for the set of c_j that satisfied constraints 1 - 3. Constraint 4 was satisfied via a *post hoc* assessment of the mean fit ratio across all surveys ($\text{ratio}_j = \sum_t y_{jt}^2 / \sum_t \epsilon_{jt}^2$), and a visual inspection of $\hat{\alpha}_t$ such that it showed a consistent trend to the input indices.

2.2 Stock assessment modeling

Model settings (e.g. *data.ss* & *control.ss*) for the Stock Synthesis model incorporating the DFA composite index were virtually identical to those described in Kai et al. (2022c,b,a). For that reason, only changes to settings made to incorporate the DFA composite index and run sensitivity analyses related to the DFA composite index model are described in this paper. Readers seeking greater detail are directed to those working papers for a complete description of overall model settings.

2.2.1 *data.ss* file

The data file used for DFA composite index model runs were identical to the one described in (Kai et al., 2022c) with the exception of the addition of a 31st fleet, the *S11_DFA_LATE* survey. This survey index spanned the years 1994 - 2020, assumed a lognormal error structure, and did not include any additional length or size composition data.

A retrospective analysis of the baseline Stock Synthesis model (Kai et al., 2022a) indicated significant retrospective steps after the removal of the 2020 & 2018 data years, and this retrospective pattern was also present for models fitting to the DFA composite index. This was identified to be caused by abnormally large input sample size² for Fleet 20 Taiwan small scale longline length composition data in those years (2018 sample size = 20,416 and 2020 sample size = 21,571). Two alternative *data.ss* files were prepared for sensitivity runs investigating the effect of those data on assessment outcomes:

1. (*drop*) Removal of the length composition data for Fleet 20 in 2018 and 2020.
2. (*ess*) Rescale the input sample size for Fleet 20 in 2018 and 2020 to the average sample size (sample size = 1,513) of the other years where length composition data was available (2012-2017 and 2019).

2.2.2 *control.ss* file

The control file used for DFA composite index model runs was identical to the one described in (Kai et al., 2022b) with the exception of the addition of and fitting to the *S11_DFA_LATE* survey described in Section 2.2.1. Fitting to this index required turning off the likelihood component

²~ 13× larger than average input sample sizes for Fleet 20 in years 2012-2017 and 2019

for *S6_JPN_LATE* (setting survey likelihood $\lambda = 0$ for this fleet), and turning on the likelihood component for *S11_DFA_LATE* (setting survey likelihood $\lambda = 1$ for this fleet).

Inserting a new survey fleet also required the specification of an associated selectivity curve. Selectivity was assumed to be length-based and sex-specific. Given that there was no composition data associated with the new survey fleet, selectivity was mirrored from another representative fleet. Two alternative *control.ss* files were prepared for sensitivity runs investigating the effect of selectivity choice on assessment outcomes:

1. (*Flt7*) Selectivity was mirrored from *F7_JPN_ENY_DP*, Japanese Enyo deep longline. This is the selectivity assumed for the Japanese research and training vessel survey index *S7_JPN_RTV*.
2. (*Flt17*) Selectivity was mirrored from *F17_US_Lonline_DP*, Hawaii deep-set longline. This is the selectivity assumed for the US Hawaii deep-set longline survey index *S1_HW_DP*.

Selectivity for the Taiwanese large scale longline fleet (*F19_TAIW_LG* & *S3_TAIW_LG*) was assumed to be time-varying. As such it can not be used to reliably index changes in population trend and so was not considered as a representative selectivity for the *S11_DFA_LATE* survey fleet (Kai et al., 2022b).

2.2.3 Sensitivity analyses

Two sets of sensitivity analyses were conducted in the development of the alternate late-period CPUE model. The first set of sensitivities consisted of 108 models relating to the construction of the DFA composite index and identification of an appropriate baseline Stock Synthesis model configuration for the DFA composite index. The 108 sensitivity models were a full-factorial combination of the following five model assumptions:

1. (*S1_CV*) Choice of input CV for the DFA for the US Hawaii longline index: **0.4³**, 0.2 or 0.13.
2. (*S3_CV*) Choice of input CV for the DFA for the Taiwanese large scale longline index: 0.4, **0.2** or 0.13.
3. (*S7_CV*) Choice of input CV for the DFA for the Japanese research and training vessel longline index: 0.4, 0.2 or **0.13**.
4. (*Selex mirror*) Choice of assumed selectivity mirror for the DFA survey fleet in the Stock Synthesis model: *Flt7* or *Flt17*.
5. (*Add extra variance*) Choice of adding extra variance to the output DFA index such that it had an average CV = 0.2 in the Stock Synthesis model: *TRUE* or *FALSE*.

Once an appropriate baseline Stock Synthesis model configuration for the DFA composite index model was identified from the initial 108 model runs, a second set of sensitivities was conducted to identify an optimal (or set of optimal) candidate model(s) to be considered as an alternate

³Bold values indicate the average CV as specified in the Stock Synthesis model.

late-period CPUE model. An additional 6 model runs were considered based on the baseline DFA composite index model, and were a full-factorial of the following two model assumptions:

1. (*Add extra variance*) Choice of adding extra variance to the output DFA index such that it had an average CV in the Stock Synthesis model of: 0.2 or 0.4.
2. (*Flt20_lf*) Choice of treatment of the Fleet 20 Taiwan small scale longline length composition data: no treatment, *drop*, or *ess*.

2.2.4 Diagnostics

Models in the first 108 sensitivity runs were evaluated on the basis of convergence (e.g. maximum gradient component $< 1e - 04$ and a positive definite Hessian solution), as well as with fits to common likelihood components (e.g. fit to *S5_JPN_EARLY* & total size composition likelihood).

Models in the second 6 sensitivity runs were evaluated based on standard stock assessment diagnostics [Carvalho et al. \(2017, 2021\)](#): residual analysis (e.g. survey and length composition runs tests), age-structured production model (e.g. ASPM), retrospective analysis (3 year peel), R0 profile, and model convergence (e.g. maximum gradient component $< 1e - 04$, and a positive definite Hessian solution).

3 Results

3.1 DFA

The common trend α_t estimated from the DFA models showed consistency with the input abundance indices, and was fairly robust to assumptions on the input CVs for the three input abundance indices ([Figure 2](#)). Overall the composite index showed a decline from 1994 to the mid-2000s before increasing through the mid-2010s, with clear noise reduction in the index. The trend appears stable over the last four years.

Focusing on the proposed candidate DFA index which assumed input CVs equal to those specified in the original Stock Synthesis model ([Kai et al., 2022b,a](#)), the estimated factor loadings $\mathbf{\Gamma}$ were 0.368 for the Hawaii deep-set longline index, 0.216 for the Taiwanese large scale longline index and 0.418 for the Japanese research and training vessel index. Factor loadings sum to 1 and represent the relative contribution of each input index to the common trend α_t . In the current analysis the Japanese research and training vessel index was the most influential, followed by the Hawaii deep-set longline index, and Taiwanese large scale longline index. Factor loadings greater than $|0.2|$ are an indication that the common DFA trend α_t is strongly coherent with that particular input trend ([Zuur et al., 2003b](#)), which is the case in the current analysis. Furthermore, the proposed candidate DFA index showed a good mean fit ratio (0.415) calculated across all indices. [Peterson et al. \(2021\)](#) suggested that a mean fit ratio $\lesssim 0.4$ is indicative of a very well fit model, a mean fit

ratio $\lesssim 0.5$ is indicative of a good fitting model, and a mean fit ratio $\gtrsim 0.5$ could be an indication of a poorly fitted model.

3.2 Stock Synthesis

3.2.1 Sensitivity analyses

Results from the first set of sensitivity analyses indicated that the Stock Synthesis model outputs were most sensitive to 1) the choice of fleet to use as a mirror for the *S11_DFA_LATE* selectivity curve (Figure 3) and 2) the decision to add extra variance to the DFA index when input into Stock Synthesis (Figure 4). The Hawaii deep-set longline *F17_US_Lonline_DP* encounters slightly larger individuals than the Japanese research and training vessels *F7_JPN_ENY_DP* (see Kai et al., 2022a Figure 13), and shifting the selectivity to index the largest individuals could result in the lower estimate of population scale and more depleted stock status. The back-transformed estimates of uncertainty coming out of the DFA were small ($CV \simeq 0.05$) so adding an extra variance adjustment factor to get the average $CV = 0.2$ pulled the model solution away from the *S11_DFA_LATE* index and towards the other likelihood components (e.g. Catch, Length composition, Size Frequency data, and Recruitment). Noticeably, fit to the early index *S5_JPN_EARLY* did not meaningfully improve when the *S11_DFA_LATE* was down-weighted indicating 1) that the model solution was strongly informed by the fit to the *S5_JPN_EARLY* index or that 2) the *S11_DFA_LATE* was not inconsistent with the *S5_JPN_EARLY* index. As mentioned previously (Section 3.1), Stock Synthesis model outputs indicated that the model was reasonably robust to the choice of input CV for the three indices: Hawaii longline (*S1_HW_DP*; Figure 5), Taiwanese large scale longline (*S3_TW_LG*; Figure 6), and the Japanese research and training vessels (*S7_JPN_RTV*; Figure 7).

Selection of the baseline DFA candidate model out of the 108 models evaluated was done on the basis of convergence (maximum gradient $< 1e-04$ and positive definite hessian solution) and improved fit relative to an equivalent model fitting to the *S6_JPN_LATE* index described in Kai et al. (2022a). Improved fit was judged based on having a lower negative log-likelihood for key likelihood components that were common between the two models: the *S5_JPN_EARLY* index and the length composition data. There were 27 models that met the selection criteria, all of which assumed selectivity mirrored by the *F7_JPN_ENY_DP* and adding an extra variance adjustment factor to the *S11_DFA_LATE* index so that the mean $CV = 0.2$. The only difference between models being the assumed input CVs for each index in the DFA. Estimated model trends were very similar across the 27 models which met the criteria (Figure 8) and the model which preserved the input CVs as assumed in Kai (2021a); Kai et al. (2022a) was selected as the baseline DFA candidate model: 0.4 for the Hawaii longline index, 0.2 for the Taiwanese longline index, and 0.13 for the Japanese research and training vessel index.

3.2.2 Candidate model diagnostics

3.2.2.1 Convergence All six candidate models fitting to the *S11_DFA_LATE* index appeared to converge with gradients $< 1e - 04$ except for the model which dropped the 2018 & 2020 for Fleet 20 and had a variance adjustment factor for *S11_DFA_LATE* resulting in an average CV of 0.2 (Table 1). This model also had several warnings indicating that one of the time-varying parameters for *F19_TAIW_LG* was close to a bound during minimization. However, an investigation into the estimated time varying selectivity for *F19_TAIW_LG* by sex (females Figure 9, and males Figure 10) did not find anomalies with the estimated selectivities for this model relative to the other candidates indicating that this warning could be benign.

3.2.2.2 Fit All six candidate models fitting to the *S11_DFA_LATE* index exhibited better fits (smaller negative log likelihood) to comparable length composition components (Table 2) than the model fitting to the S6 index (Kai et al., 2022a). Within the six candidate models, fit to comparable length composition components improved when models dropped or down-weighted the 2018 & 2020 length composition for Fleet 20. Additionally, two candidate models which did not modify the Fleet 20 length composition data showed a lower NLL for the *S5_JPN_EARLY* than the model fitting to the S6 index (Kai et al., 2022a). However, the model fitting to the S6 index (Kai et al., 2022a) had a lower NLL for the catch, generalized size frequency and recruitment components of the likelihood.

Fit to the CPUE indices (*S5_JPN_EARLY* & *S11_DFA_LATE*) was generally quite good (RMSE values < 0.11 for all indices; Table 3). However, despite the small amount of residual error, runs tests showed a non-random pattern in the residuals for the fit to the *S11_DFA_LATE* index across the board (Figures 11 - 16).

Runs testing of the mean length residual also showed a consistent pattern (Figures 17 - 22). All models fitting to the *S11_DFA_LATE* index showed non-random residual patterns for the same five fisheries with non-random residual pattern in the *S6_JPN_LATE* model (Kai et al. (2022a) Figure 2): *F1_MEX*, *F3_CHINA*, *F4_JPN_KK_SH*, *F15_USA_GILL*, *F18_USA_Lonline_SH*. However, models with larger down-weighting for the *S11_DFA_LATE* index (variance adjustment = 0.4) also showed a non-random residual pattern for *F17_USA_Lonline_DP*

3.2.2.3 ASPM All ASPM diagnostic models achieved a positive definite hessian solution and converged to low gradients (the *S11_drop_0.2* had the largest maximum gradient of $1.28e-04$). Additionally, visual inspection indicated the presence of a production function producing a consistent trend and estimates of virgin biomass with the full model for four of the six candidate models (Figure 23), the two outliers in terms of virgin biomass being *S11_base_0.4* and *S11_ess_0.2*.

3.2.2.4 Retrospectives Most models showed some degree of retrospective bias for spawning biomass (Figures 24-30) and fishing mortality (Figures 31-37), of comparable direction and magnitude to the S6 index model (Kai et al., 2022a), when quantities were not re-scaled to a mean of 1.

The exception was the *S11_drop-0.2* model which showed virtually no retrospective bias. In general models which reduced the effect of the Fleet 20 data in years 2018 & 2020, either by removing or down-weighting this data, indicated less retrospective bias.

3.2.2.5 R0 profile Likelihood profiles of R0 for total likelihood components (Figures 38-44) were dominated by the recruitment component, and asymmetric with a greater support for a larger range of $\log(R0)$. When broken out by fleet (Figures 45-51) there was a clear conflict between the early and late CPUE indices. This conflict was smallest for the *S11_drop-0.2* and *S11_ess-0.2* models which minimized the effect of the Fleet 20 data without downweighting the *S11* index. The length composition data profiles for two major fleets (*F4_JPN_KK_SH* & *F7_JPN_ENY_DP*) generally showed consistency with the total likelihood.

3.2.3 Candidate model results

All six S11 candidate models estimated lower estimates of population scale than the S6 index model (Kai et al., 2022a), and also estimated a flatter or slightly declining trend over the last 20 years of the model period (Figure 52). This reflects the pattern of estimated fishing mortality in the six S11 models which do not continue to decline after the drift gillnet ban (Figure 53). Patterns in recruitment are largely similar, all models predict a high spike in recruits in the late 1980s, coinciding with the highest estimates of F. However, recruitment deviates for the six S11 candidate models were more negative from the mid-1990s through the mid-2000s relative to the S6 index model (Figure 54). Estimates of fisheries selectivity were fairly consistent across models though the six S11 candidate models typically estimated slightly greater selectivity at larger sizes (Figures 9, 10, 55, & 56).

Within the six S11 candidate models, models that reduced the influence of the Fleet 20 Taiwan small scale longline length composition data (down-weighting or removing) tended to estimate lower population scale and a corresponding higher increase in fishing mortality. Selectivity estimates for Fleet 20 did not noticeably differ across runs (Figures 55, & 56), indicating that changes in model outputs are not likely related to differences among the Fleet 20 composition data between years. Rather, by down-weighting (or removing) the Fleet 20 composition data it effectively up-weights the other data components which may prefer lower estimates of population scale. Conversely, models that added extra variance to the S11 index (mean CV = 0.4) down-weighted this index in the likelihood and pulled estimates towards the *S5_JPN_EARLY* which preferred larger estimates of population scale.

4 Discussion

Dynamic Factor Analysis was able to produce a composite index that generally captured the trend seen in the three candidate indices (the Hawaii longline, Taiwanese large-scale longline, and Japanese research and training vessel) and can serve as an alternative model hypothesis for the

2020 ISC blue shark stock assessment. However, there are a number of ways that this approach could be further explored. Though the composite index appeared to be relatively insensitive to the choice of input CV, the output CVs were quite small and could be sensitive to the choice of re-scaling constant c_j Peterson et al. (2021). Future analyses should explore alternative choices for c_j (that also meet all 4 selection criteria) to evaluate if this has an impact on the output CV for the composite index. Regardless of the choice of c_j , this working paper evaluated the effects of down-weighting the composite DFA index (adding extra variance adjustment such that the average CV was 0.2 or 0.4) and found that down-weighting the composite DFA index tended to produce larger estimates of population scale and more optimistic stock status. Lastly, though the three candidate indices for the DFA were chosen *a priori* based on similarity of trends in the overlapping period and similarity in fishing behavior, DFA could accommodate all indices in a single analysis and this could be explored in the future. However, combining all indices into a single trend would require careful thought for an appropriate selectivity curve in the Stock Synthesis model.

Stock Synthesis model outputs appeared sensitive to the choice of assumed selectivity curve for the index. However, the *F17_US_Lonline_DP* was not subsequently considered as a viable candidate on the basis of not improving model the fit to key likelihood components. This indicated that assuming the *F17_US_Lonline_DP* selectivity could be introducing conflict between the existing data components relative to the *F7_JPN_ENY_DP* selectivity pattern. Furthermore, the *F17_US_Lonline_DP* selectivity curve was based on a smaller number of observations from a more restricted spatial area which could mean that it may not be as representative of the indexed population.

5 Recommendations

Evaluation of the six S11 candidate models relative to the S6 index model (Kai et al., 2022a) did not identify a clear best model based on fits to the data or model diagnostics. Even within the six S11 candidate models it was difficult to identify a single model that was superior across all diagnostic categories. Given this, and the clear difference in population trajectory and scale between the two alternative late CPUE hypotheses, a way forward would be to consider a model ensemble where each alternative CPUE hypothesis is weighted equally. For the alternative late CPUE hypothesis two models could be considered with equal weighting: *S11_base_0.2* & *S11_ess_0.2* in order to account for the model configuration which alleviates retrospective bias without arbitrarily removing the Taiwanese small scale longline length composition data for years 2018 & 2020 from the model. This would result in a proposed ensemble of three models: 50% *S6_base_0*, 25% *S11_base_0.2* & 25% *S11_ess_0.2*.

Acknowledgments

Thank you very much to members of the ISC SHARKWG and participants at the pre-assessment meetings who contributed to discussions which shaped this analysis. Additionally, thank you to Meg

Oshima (NOAA Fisheries) for assistance with setting up, running, and plotting model diagnostics with the `ss3diags` R package.

References

- Azevedo, M., Duarte, R., Cardador, F., Sousa, P., Fariña, C., Sampedro, P., Landa, J., and Costas, G. (2008). Application of dynamic factor analysis in the assessment of Iberian anglerfish stocks. *ICES Journal of Marine Science*, 65(7):1362–1369.
- Campana, S. E. (2016). Transboundary movements, unmonitored fishing mortality, and ineffective international fisheries management pose risks for pelagic sharks in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(10):1599–1607. Section: 1599.
- Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., and Piner, K. R. (2017). Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fisheries Research*, 192:28–40.
- Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K. R., Maunder, M. N., Taylor, I., Wetzell, C. R., Doering, K., Johnson, K. F., and Methot, R. D. (2021). A cookbook for using model diagnostics in integrated stock assessments. *Fisheries Research*, 240:105959.
- Conn, P. B. (2010). Hierarchical analysis of multiple noisy abundance indices. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(1):108–120.
- Ducharme-Barth, N., Siders, Z., and Ahrens, R. (2022). Blue shark catch and CPUE for the US Hawaii longline fleet. Technical Report ISC/22/SHARKWG-1/02.
- Ducharme-Barth, N. and Vincent, M. (2020). Analysis of Pacific-wide operational longline dataset for bigeye and yellowfin tuna catch-per-unit-effort (CPUE). Technical Report SC16-SA-IP-07.
- Ducharme-Barth, N., Vincent, M., Hampton, J., Hamer, P., Williams, P., and Pilling, G. (2020). Stock assessment of bigeye tuna in the western and central Pacific Ocean (30July) - Rev.03. Technical Report SC16-SA-WP-03.
- Fujinami, Y., Shiozaki, K., Hiraoka, Y., Semba, Y., Ohshimo, S., and Kai, M. (2021). Seasonal migrations of pregnant blue sharks (*Prionace glauca*) in the northwestern Pacific. *Marine Ecology Progress Series*, 658:163–179. Section: 163.
- Heard, M., Rogers, P. J., Bruce, B. D., Humphries, N. E., and Huvaneers, C. (2018). Plasticity in the diel vertical movement of two pelagic predators (*Prionace glauca* and *Alopias vulpinus*) in the southeastern Indian Ocean. *Fisheries Oceanography*, 27(3):199–211. Section: 199.
- Holmes, E. E., Ward, E. J., and Scheuerell, M. D. (2021a). Analysis of multivariate time series using the MARSS package. Version 3.11.4. Publisher: Zenodo Version Number: 3.11.4.
- Holmes, E. E., Ward, E. J., Scheuerell, M. D., and Wills, K. (2021b). MARSS: Multivariate Autoregressive State-Space Modeling.

- Holmes, E. E., Ward, E. J., and Wills, K. (2012). MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal*, 4(1):30.
- Hoyle, S. D., Lauretta, M., Lee, M. K., Matsumoto, T., Sant’Ana, R., Yokoi, H., and Su, N.-J. (2019). Collaborative study of yellowfin tuna cpue from multiple atlantic ocean longline fleets in 2019. ICCAT SCRS/2019/081.
- ICCAT (2020). Report of the 2019 ICCAT yellowfin tuna stock assessment meeting. *Collect. Vol. Sci. Pap. ICCAT*, 76(6):244–515.
- ICCAT (2021). Report of the 2021 bigeye stock assessment meeting. *Collect. Vol. Sci. Pap. ICCAT*, 78(2):335–485.
- ISC (2017). Stock assessment and future projections of blue shark in the North Pacific Ocean through 2015. Technical Report ISC 17 Plenary report and document.
- Kai, M. (2019). Spatio-temporal changes in catch rates of pelagic sharks caught by Japanese research and training vessels in the western and central North Pacific. *Fisheries Research*, 216:177–195.
- Kai, M. (2021a). Spatio-temporal model for CPUE standardization: Application to blue shark caught by Japanese offshore and distant water shallow-set longliner in the western North Pacific. Technical Report ISC/21/SHARKWG-2/01.
- Kai, M. (2021b). Spatio-temporal model for CPUE standardization: Application to blue shark caught by longline of Japanese research and training vessels in the western and central North Pacific. Technical Report ISC/21/SHARKWG-2/03.
- Kai, M., Teo, S. L., Ducharme-Barth, N. D., and Carvalho, F. (2022a). Main outcomes and model diagnostics of the Stock Synthesis base-case model candidate in the stock assessment for blue sharks in the North Pacific. Technical Report ISC/22/SHARKWG-2/6.
- Kai, M., Teo, S. L., Ducharme-Barth, N. D., and Carvalho, F. (2022b). Stock Synthesis settings of control file for the stock assessment of blue shark in the North Pacific. Technical Report ISC/22/SHARKWG-2/5.
- Kai, M., Teo, S. L., Ducharme-Barth, N. D., and Carvalho, F. (2022c). Stock Synthesis settings of data file for the stock assessment of blue shark in the North Pacific. Technical Report ISC/22/SHARKWG-2/4.
- King, J., Wetklo, M., Supernault, J., Taguchi, M., Yokawa, K., Sosa-Nishizaki, O., and Withler, R. (2015). Genetic analysis of stock structure of blue shark (*Prionace glauca*) in the north Pacific ocean. *Fisheries Research*, 172:181–189.

- Kitakado, T., Satoh, K., Lee, S. I., Su, N.-J., Matsumoto, T., Yokoi, H., Okamoto, K., Lee, M. K., Lim, J.-H., Kwon, Y., Wang, S.-P., Tsai, W.-P., Chang, S.-T., and Chang, F.-C. (2021). Update of the trilateral collaborative study among Japan, Korea, and Chinese Taipei for producing joint abundance indices for the atlantic bigeye tunas using longline fisheries data up to 2019. *Collect. Vol. Sci. Pap. ICCAT*, 78(2):169–196.
- Liu, K.-M., Su, K.-Y., Tsai, W.-P., and Chin, C.-P. (2021). Updated standardized CPUE and catch estimation of the blue shark caught by the Taiwanese large scale tuna longline fishery in the North Pacific Ocean. Technical Report ISC/21/SHARKWG-2/14.
- Mebane, W. R. and Sekhon, J. S. (2011). Genetic Optimization Using Derivatives: The **rgenoud** Package for *R*. *Journal of Statistical Software*, 42(11).
- Methot, R. D. and Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142:86–99.
- Moyes, C. D., Fragoso, N., Musyl, M. K., and Brill, R. W. (2006). Predicting Postrelease Survival in Large Pelagic Fish. *Transactions of the American Fisheries Society*, 135(5):1389–1397. Section: 1389.
- Nakano, H. and Stevens, J. D. (2008). The Biology and Ecology of the Blue Shark, *Prionace glauca*. In *Sharks of the Open Ocean*, pages 140–151.
- Peterson, C. D., Wilberg, M. J., Cortés, E., and Latour, R. J. (2021). Dynamic factor analysis to reconcile conflicting survey indices of abundance. *ICES Journal of Marine Science*, 78(5):1711–1729.
- Taguchi, M., King, J. R., Wetklo, M., Withler, R. E., and Yokawa, K. (2015). Population genetic structure and demographic history of Pacific blue sharks (*Prionace glauca*) inferred from mitochondrial DNA analysis. *Marine and Freshwater Research*, 66(3):267.
- WCPFC (2017). Thirteenth Regular Session of the Scientific Committee: Summary Report. Technical Report SC13.
- Zuur, A. F., Fryer, R. J., Jolliffe, I. T., Dekker, R., and Beukema, J. J. (2003a). Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*, 14(7):665–685.
- Zuur, A. F., Tuck, I. D., and Bailey, N. (2003b). Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(5):542–552.

6 Tables

Table 1: Model convergence metrics (positive definite hessian solution, maximum gradient component, and number of warnings) for the model fitting to the S6 index (Kai et al., 2022a) and the 6 candidate DFA models fitting to the S11 index.

Late_index	Flt20_LF	Extra_var	PDH	mgc	warnings
S6	base	0	TRUE	5.585E-05	0
S11	base	0.2	TRUE	7.909E-06	0
S11	base	0.4	TRUE	7.674E-05	0
S11	drop	0.2	TRUE	1.276E-04	4
S11	drop	0.4	TRUE	8.836E-05	0
S11	ess	0.2	TRUE	8.594E-05	0
S11	ess	0.4	TRUE	5.656E-05	0

Table 2: Negative log-likelihood (NLL) for key components for the model fitting to the S6 index (Kai et al., 2022a) and the 6 candidate DFA models fitting to the S11 index. The NLL components are: Catch, Total Length Composition, Common Length Composition (excluding Flt 20), Size Frequency, Recruitment, S5_JPN_EARLY, S6_JPN_LATE, S11_DFA_LATE.

Late_index	Flt20_LF	Extra_var	Catch	LF	Common_LF	SizeFreq	Recruitment	S5	S6	S11
S6	base	0	0.0002	114.8570	110.2239	0.1058	-8.5890	-27.6532	-48.5806	
S11	base	0.2	0.0008	114.2100	109.4018	0.1464	-7.6920	-27.7780		-42.7064
S11	base	0.4	0.0005	113.6940	108.9602	0.1293	-8.7313	-27.7229		-23.8244
S11	drop	0.2	0.0025	108.6270	103.7778	0.1685	-8.0460	-27.3072		-42.9086
S11	drop	0.4	0.0015	107.9140	103.1492	0.1548	-8.3612	-27.4696		-24.0697
S11	ess	0.2	0.0021	109.2850	104.4301	0.1657	-7.8268	-27.4783		-42.8557
S11	ess	0.4	0.0010	108.5870	103.8443	0.1444	-8.4864	-27.6079		-23.9373

Table 3: Root-mean squared error (RMSE) for fitted indices for the model fitting to the S6 index (Kai et al., 2022a) and the 6 candidate DFA models fitting to the S11 index. The fitted indices are: S5_JPN_EARLY, S6_JPN_LATE, S11_DFA_LATE.

Late_index	Flt20_LF	Extra_var	S5	S6	S11
S6	base	0	0.0773	0.0849	
S11	base	0.2	0.0738		0.0473
S11	base	0.4	0.0754		0.1038
S11	drop	0.2	0.0869		0.0405
S11	drop	0.4	0.0826		0.0888
S11	ess	0.2	0.0824		0.0424
S11	ess	0.4	0.0787		0.0971

7 Figures

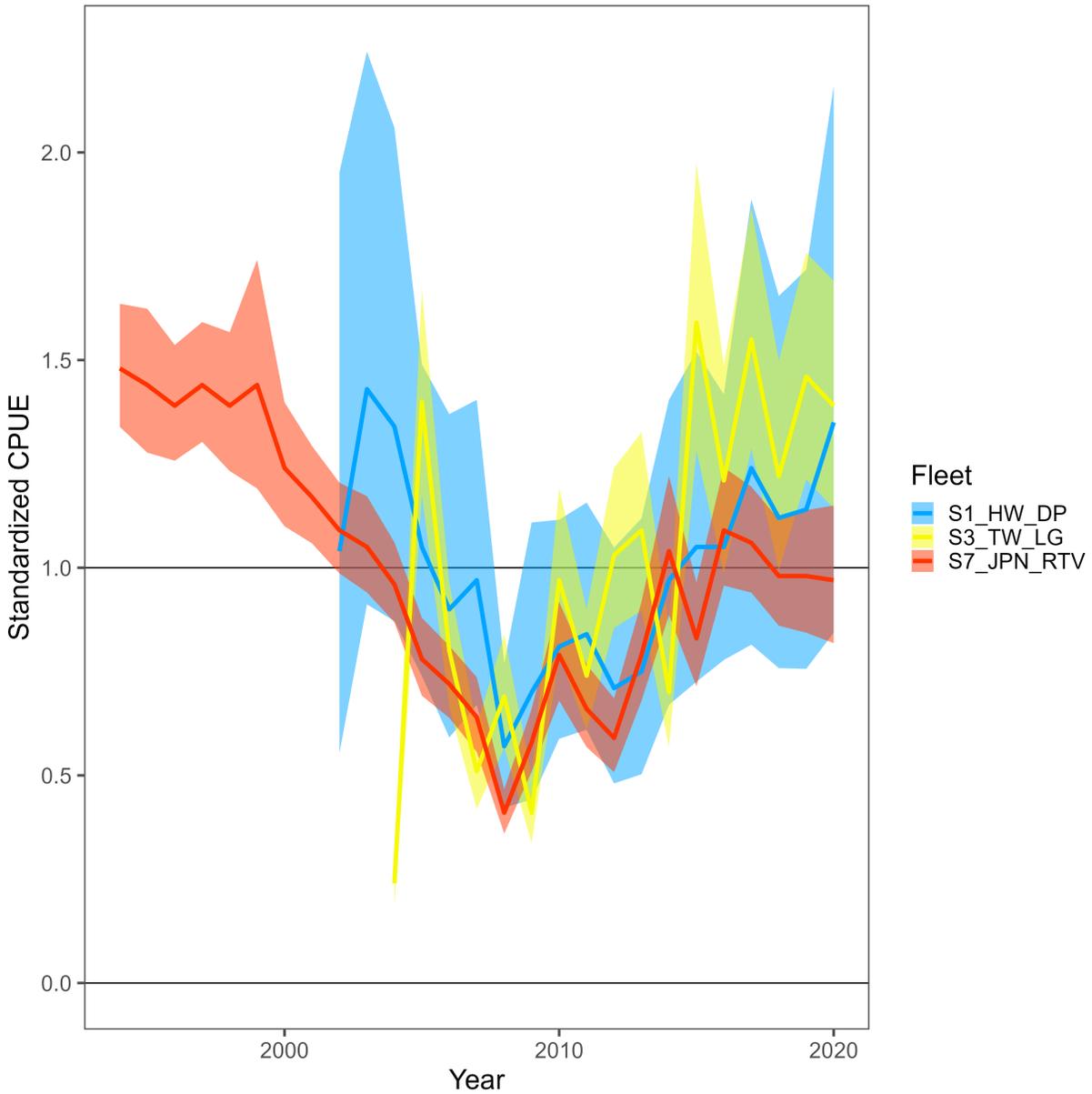


Figure 1: Standardized CPUE indices for the Hawaii longline (blue; $S1_HW_DP$), Taiwanese large scale longline (yellow; $S3_TW_LG$), and the Japanese research and training vessels (red; $S7_JPN_RTV$). Uncertainty, denoted as ± 1 SE including the additive variance adjustment factor, is shown by the shaded polygon.

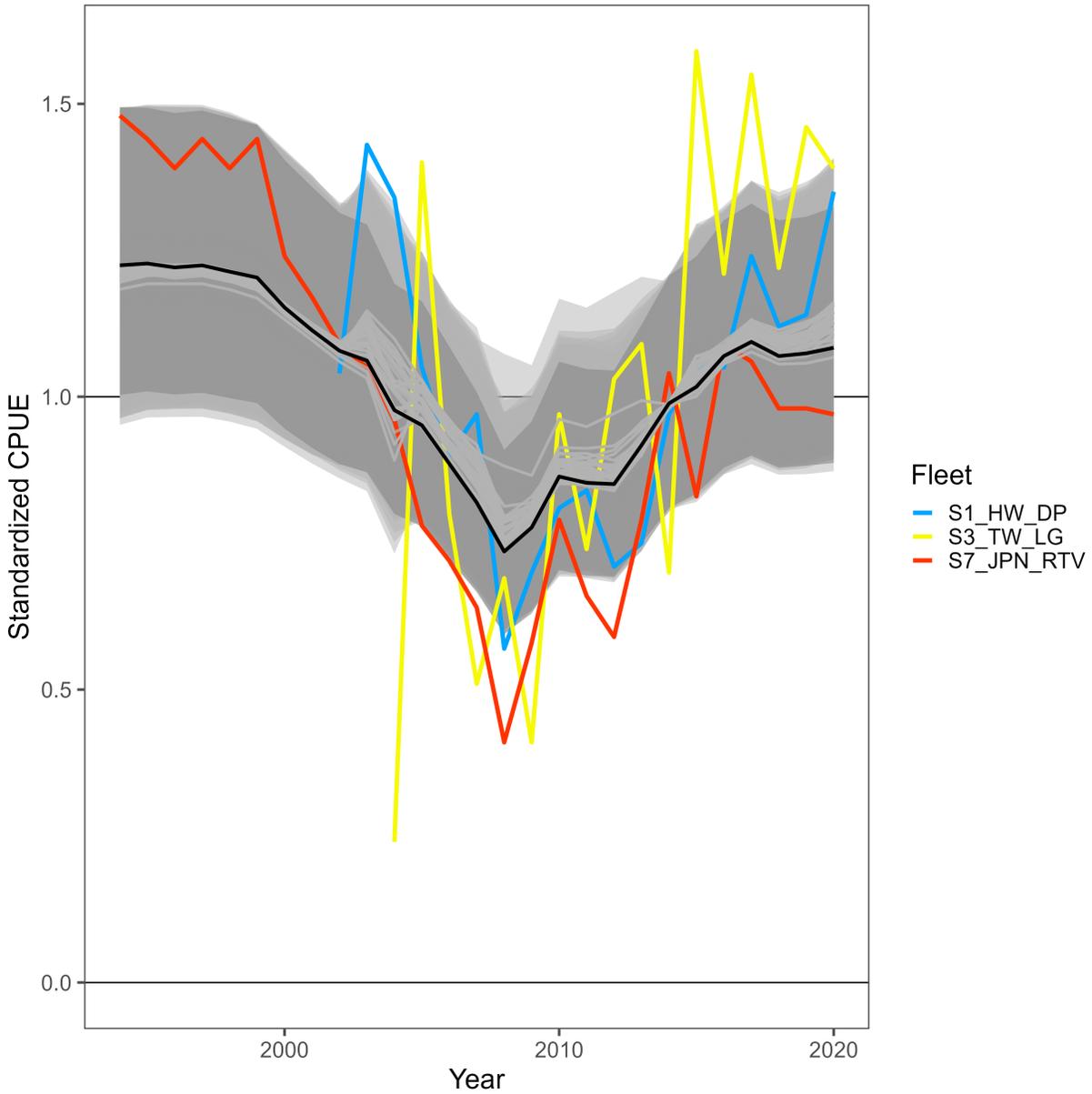


Figure 2: Estimated DFA common trends (light gray lines) across all unique combinations of input CVs. The associated uncertainty, ± 1 SE including the additive variance adjustment factor, is shown by the shaded gray polygons. The proposed candidate DFA index is shown in black. The standardized CPUE indices for the Hawaii longline (blue; *S1_HW_DP*), Taiwanese large scale longline (yellow; *S3_TW_LG*), and the Japanese research and training vessels (red; *S7_JPN_RTV*) are also shown.

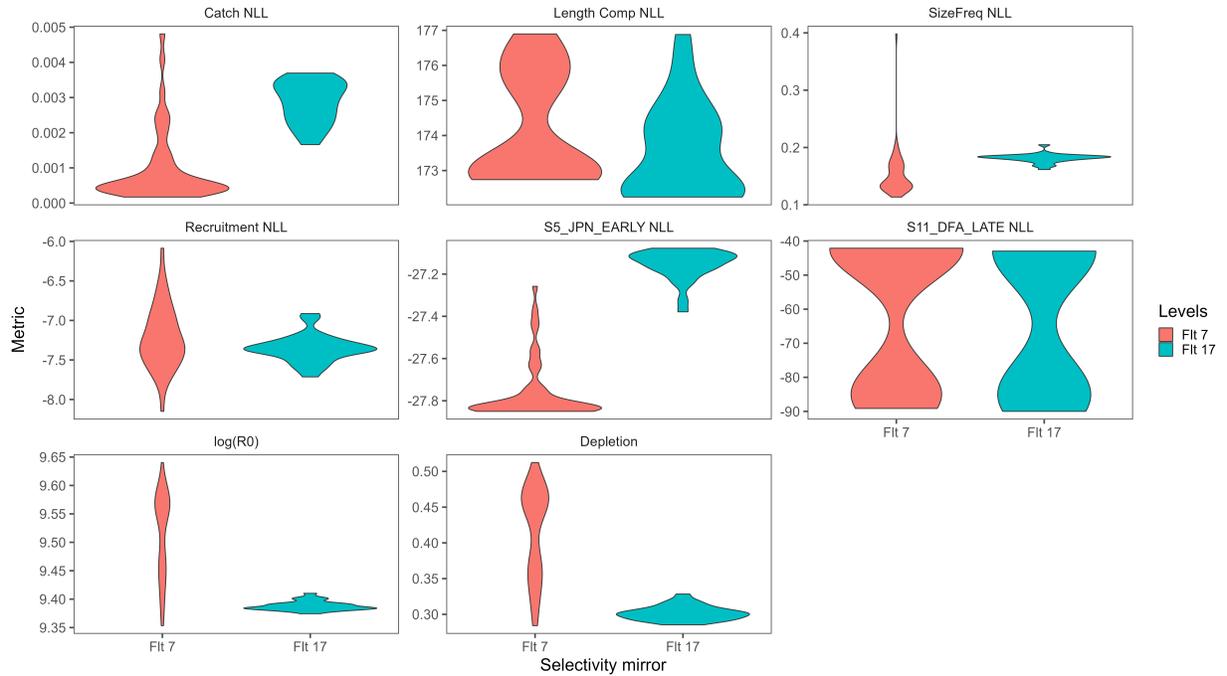


Figure 3: Violin plots showing the fit (Catch likelihood, Length composition likelihood, Size frequency likelihood, Recruitment likelihood, *S5_JPN_EARLY* likelihood and *S11_DFA_LATE* likelihood) and estimated quantities ($\log(R_0)$ and depletion) across all converged models (maximum gradient $< 1e-04$ and positive definite hessian solution) for different *S11_DFA_LATE* selectivity assumptions.

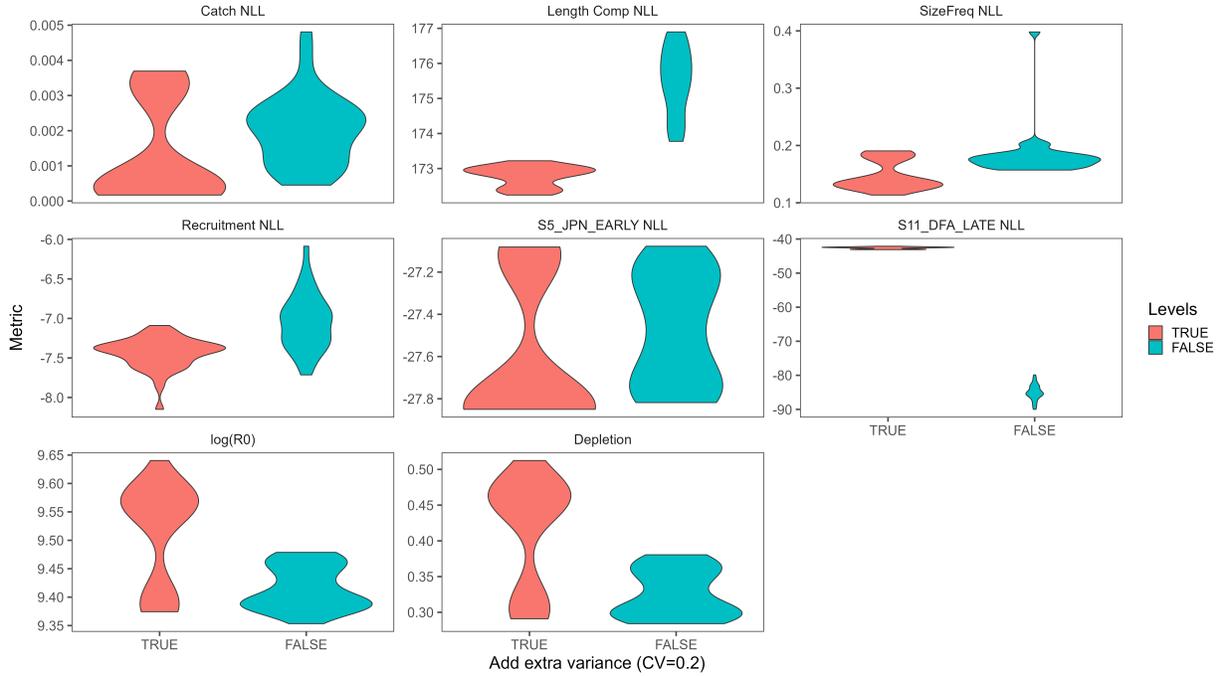


Figure 4: Violin plots showing the fit (Catch likelihood, Length composition likelihood, Size frequency likelihood, Recruitment likelihood, *S5_JPN_EARLY* likelihood and *S11_DFA_LATE* likelihood) and estimated quantities ($\log(R_0)$ and depletion) across all converged models (maximum gradient $< 1e-04$ and positive definite hessian solution) for models with and without added variance adjustment for *S11_DFA_LATE*.

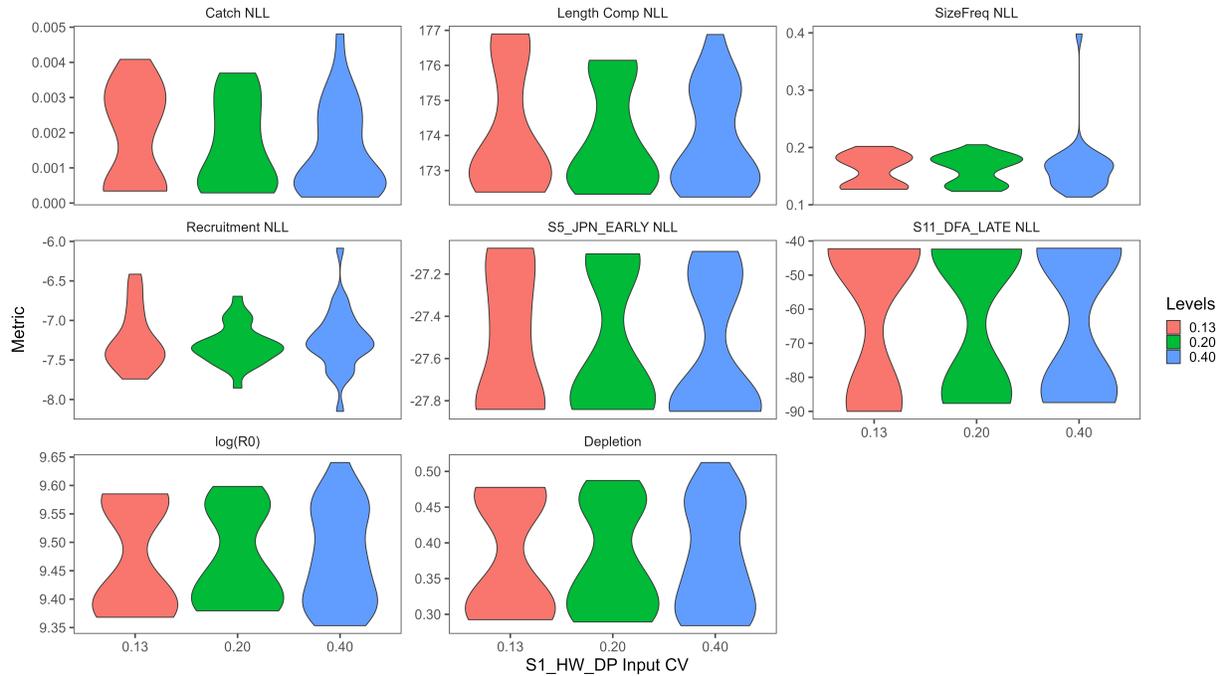


Figure 5: Violin plots showing the fit (Catch likelihood, Length composition likelihood, Size frequency likelihood, Recruitment likelihood, *S5_JPN_EARLY* likelihood and *S11_DFA_LATE* likelihood) and estimated quantities ($\log(R_0)$ and depletion) across all converged models (maximum gradient $< 1e-04$ and positive definite hessian solution) for different levels of input CV assumed for *S1_HW_DP* in the DFA.

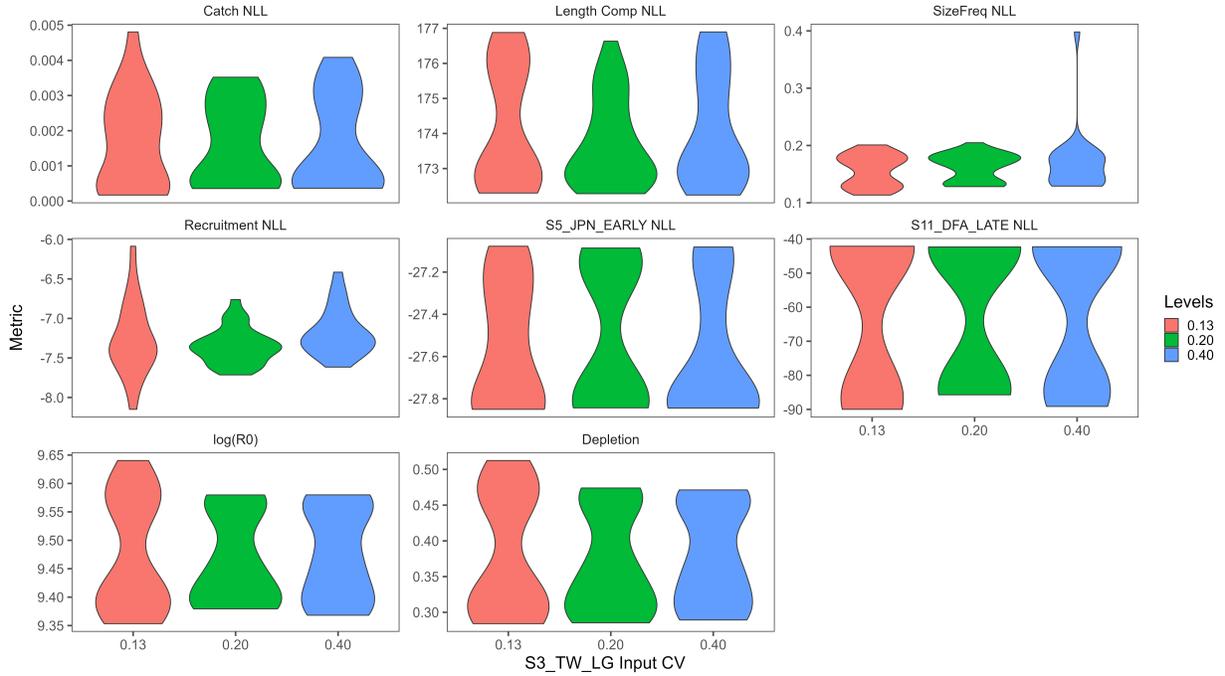


Figure 6: Violin plots showing the fit (Catch likelihood, Length composition likelihood, Size frequency likelihood, Recruitment likelihood, *S5_JPN_EARLY* likelihood and *S11_DFA_LATE* likelihood) and estimated quantities ($\log(R_0)$ and depletion) across all converged models (maximum gradient $< 1e-04$ and positive definite hessian solution) for different levels of input CV assumed for *S3_TW_LG* in the DFA.

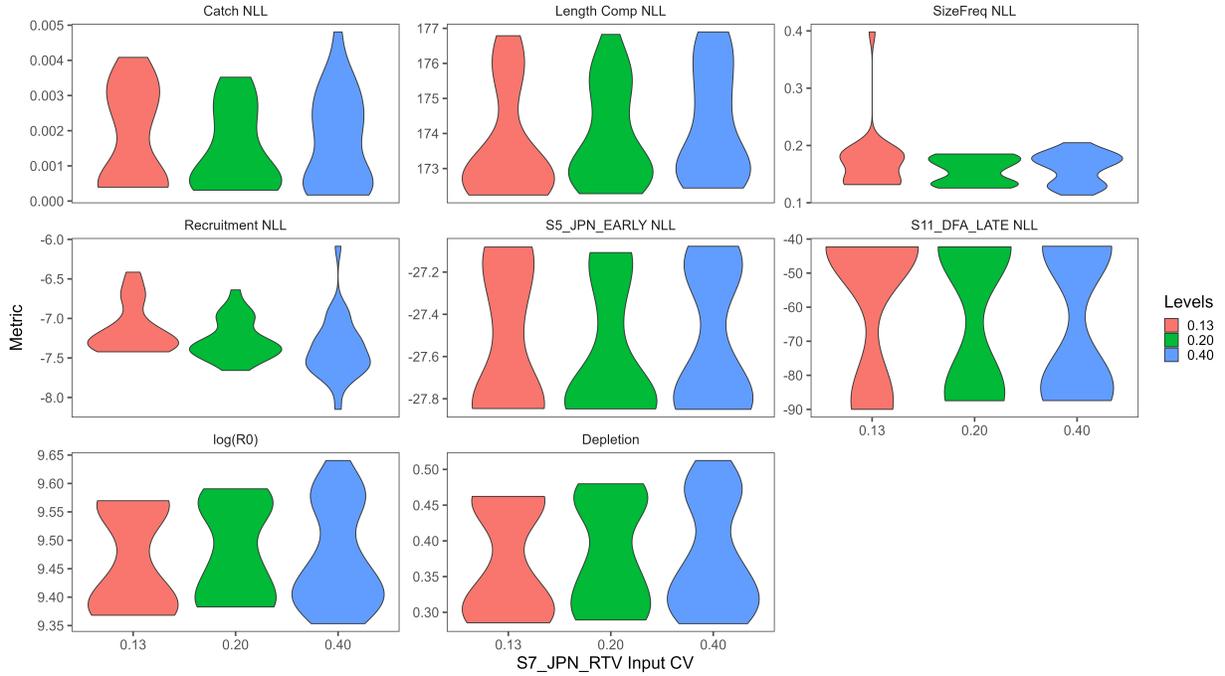


Figure 7: Violin plots showing the fit (Catch likelihood, Length composition likelihood, Size frequency likelihood, Recruitment likelihood, *S5_JPN_EARLY* likelihood and *S11_DFA_LATE* likelihood) and estimated quantities ($\log(R_0)$ and depletion) across all converged models (maximum gradient $< 1e-04$ and positive definite hessian solution) for different levels of input CV assumed for *S7_JPN_RTV* in the DFA.

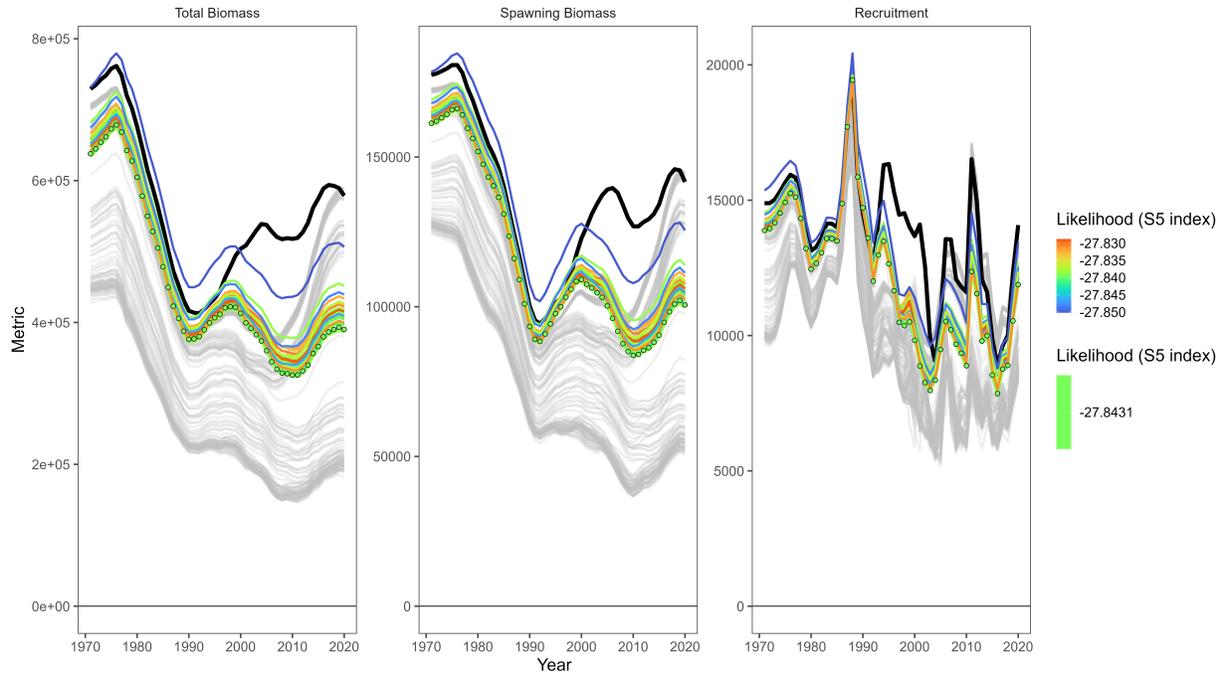


Figure 8: Time series of total biomass (left panel), spawning biomass (center panel), and recruitment (right panel) shown for all converged models considered in the first sensitivity (gray lines). Models which met the selection criteria are colored with respect to their fit to the $S5_JPN_EARLY$ index. The baseline DFA candidate model is shown by the green dots.

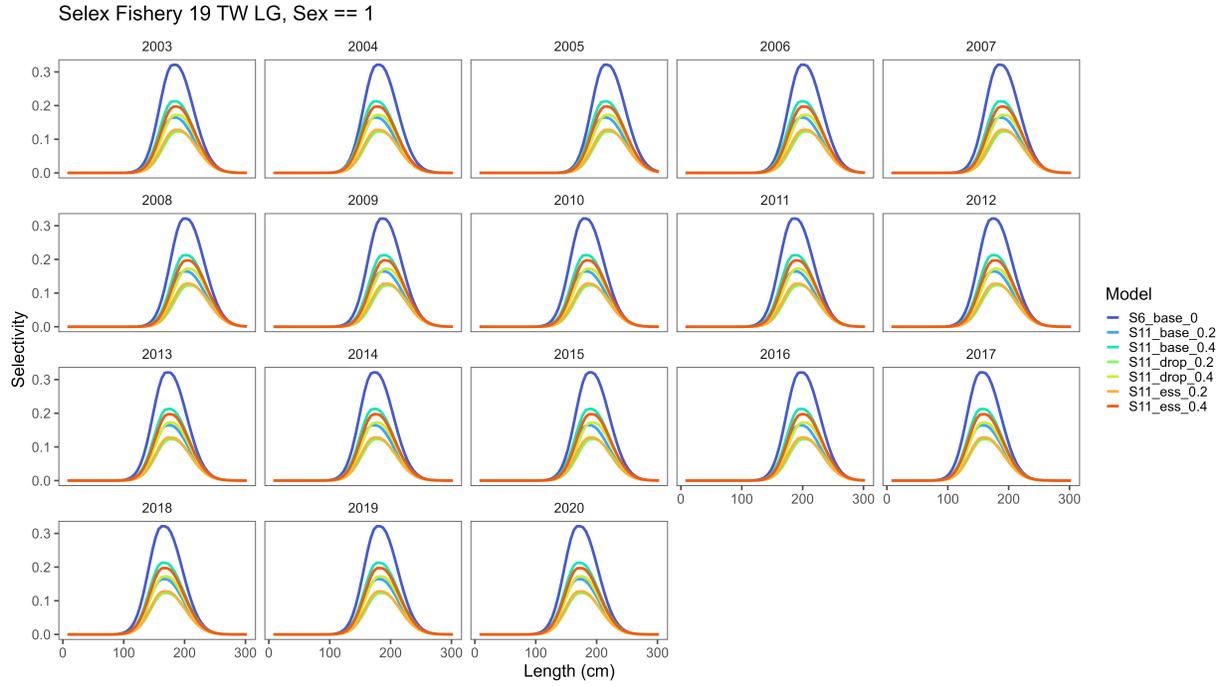


Figure 9: Time-varying selectivity for sex 1 (females) for the *F19_TAIW_LG* by candidate model.

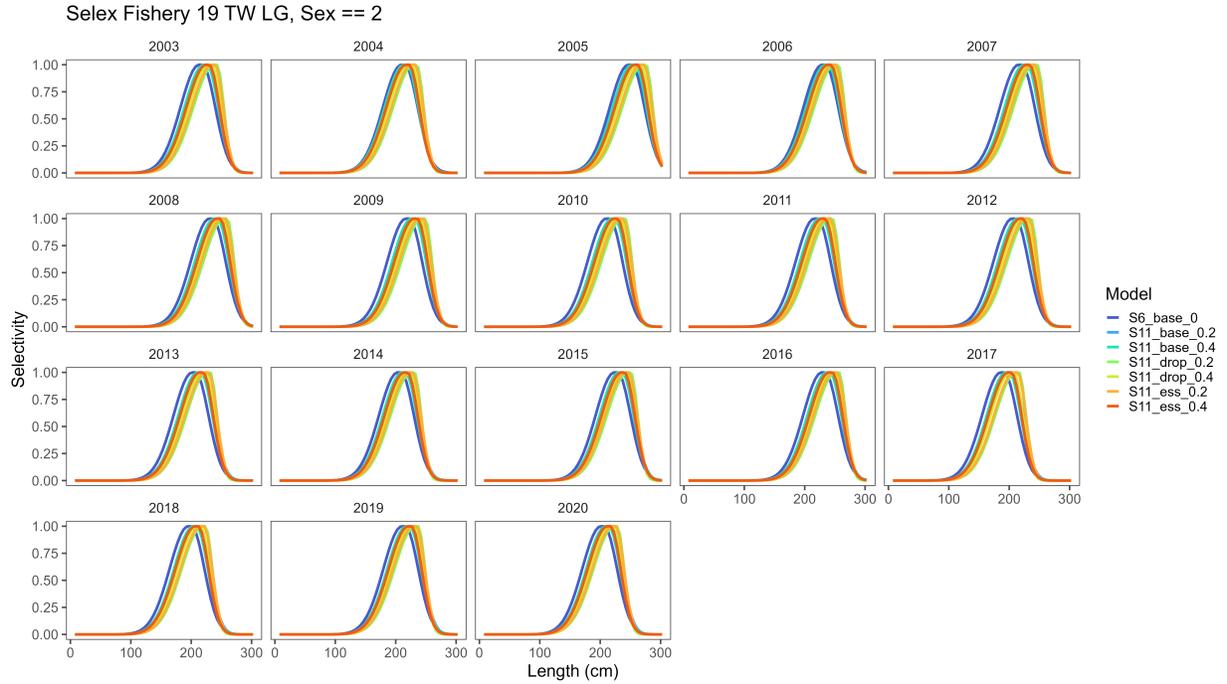


Figure 10: Time-varying selectivity for sex 2 (males) for the *F19-TAIW-LG* by candidate model.

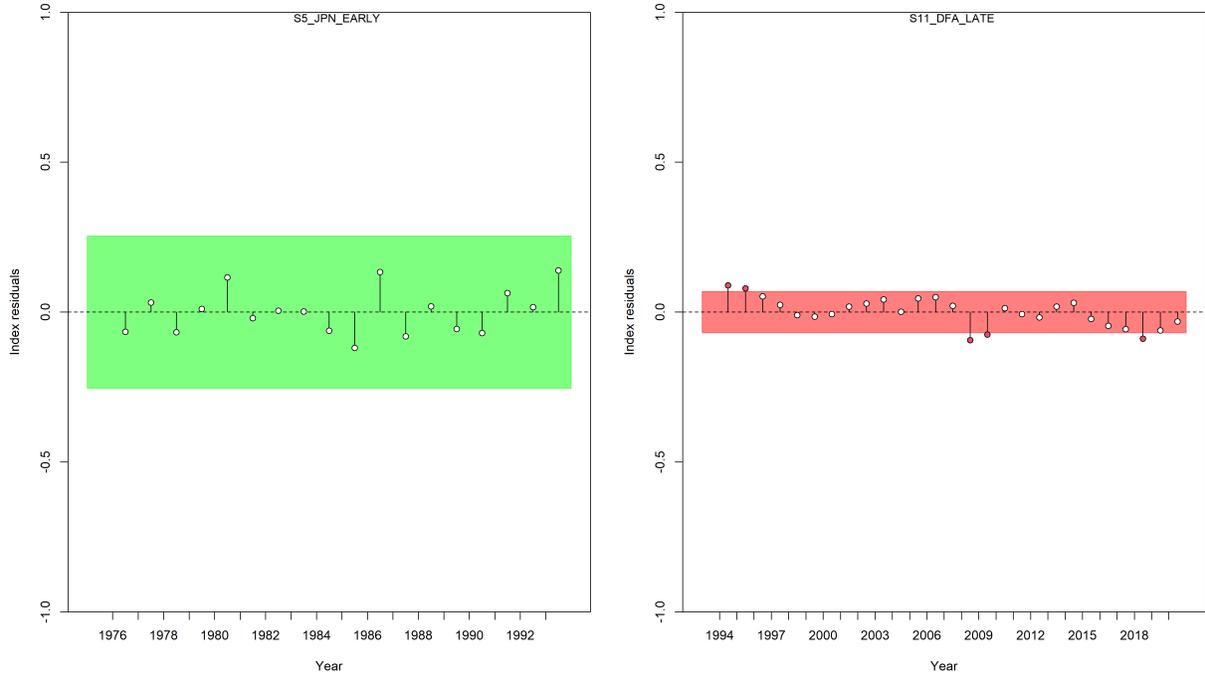


Figure 11: Model name: S11_base_0.2: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

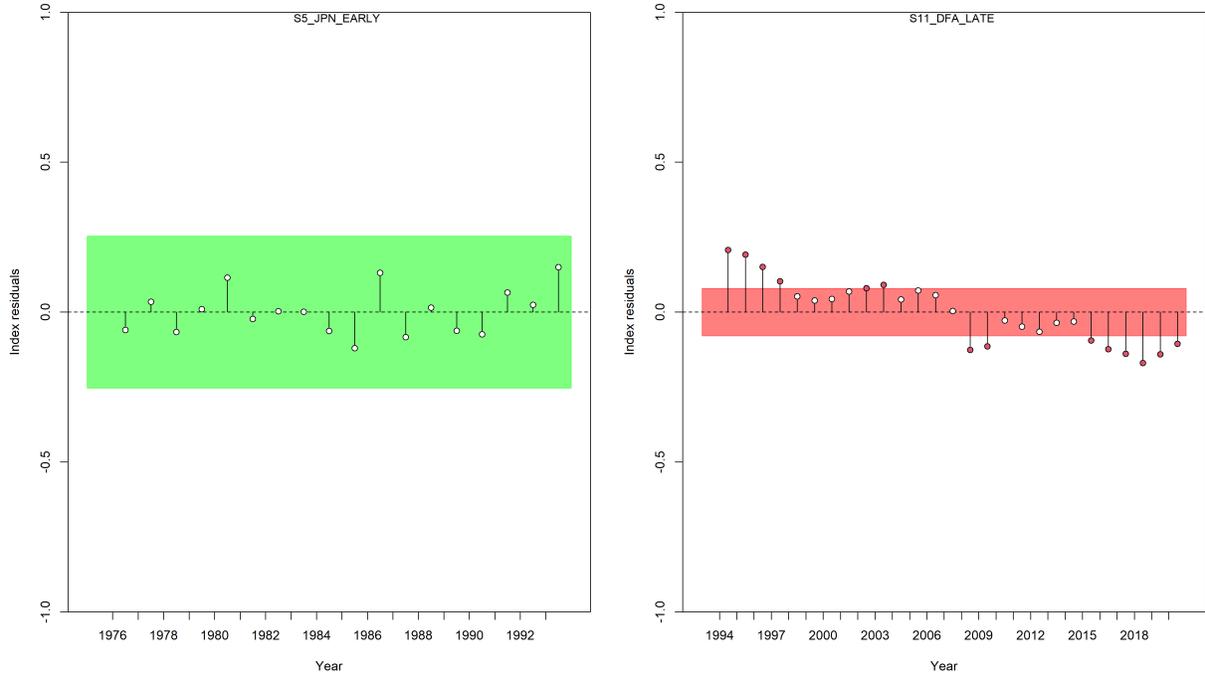


Figure 12: Model name: S11_base_0.4: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

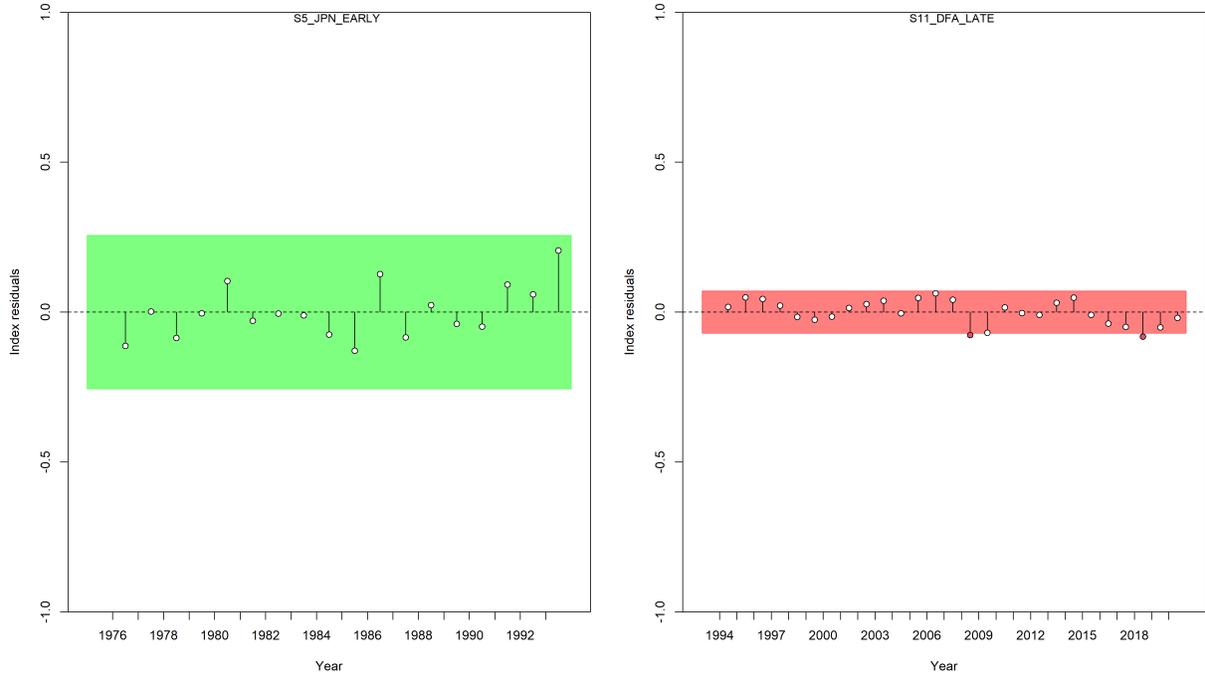


Figure 13: Model name: S11_drop_0.2: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

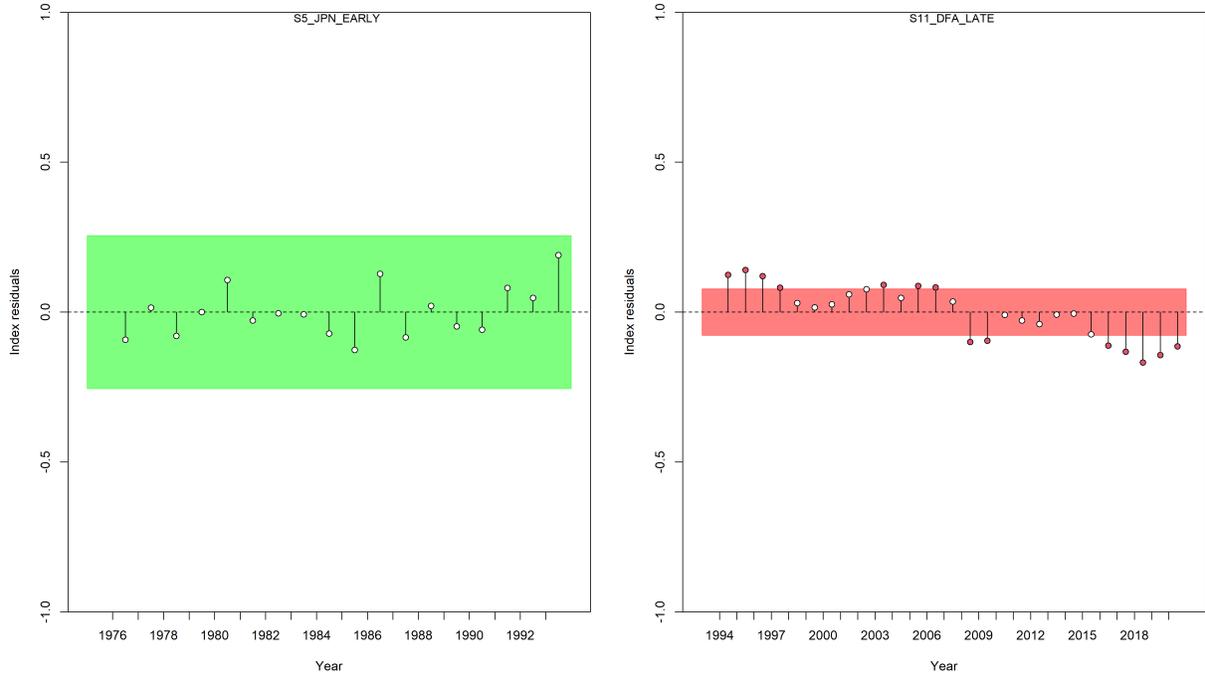


Figure 14: Model name: S11_drop_0.4: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

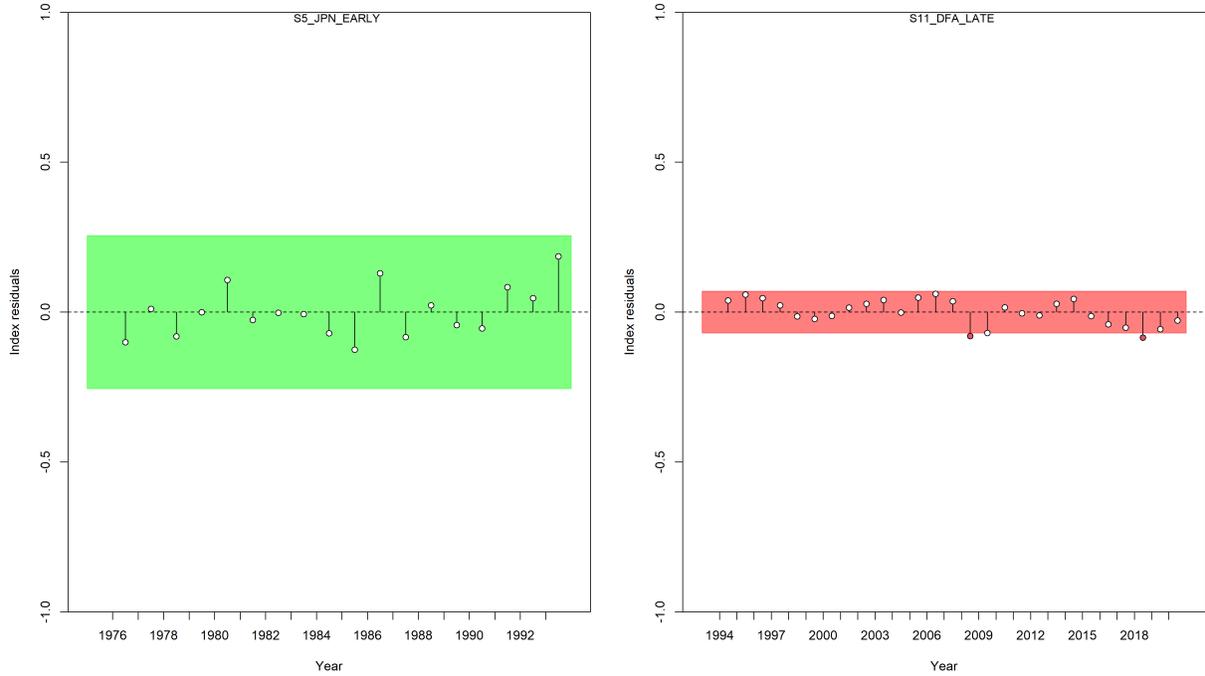


Figure 15: Model name: S11_ess.0.2: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

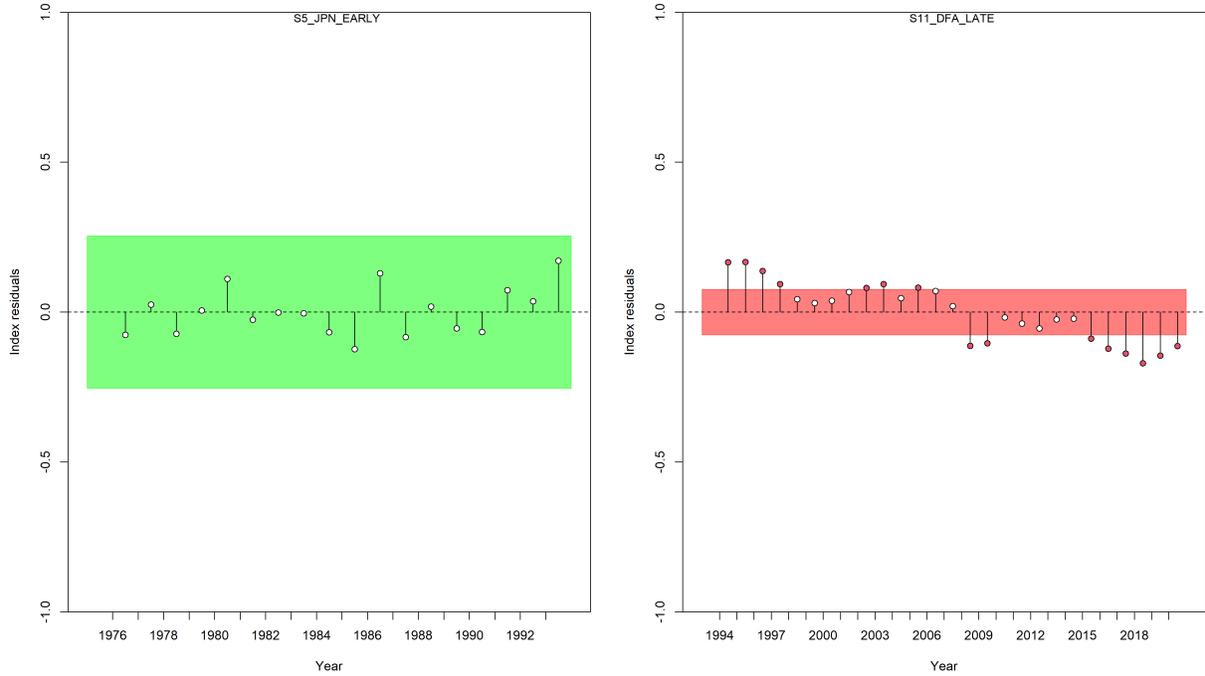


Figure 16: Model name: S11_ess.0.4: CPUE residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

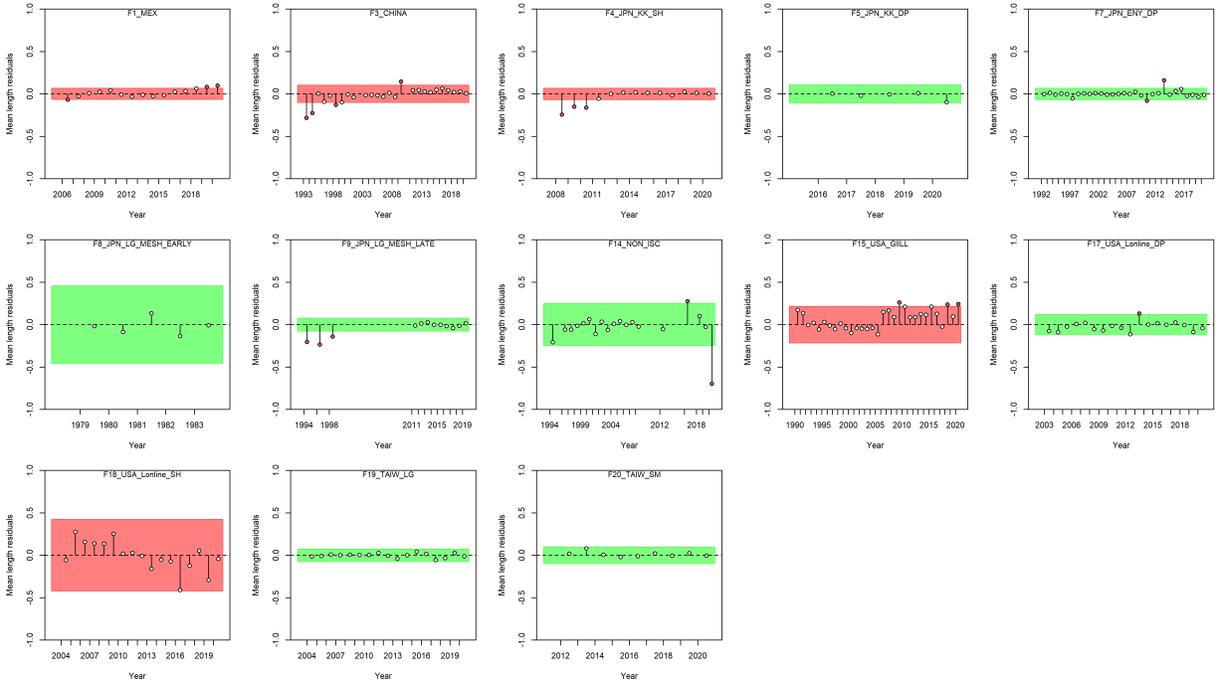


Figure 17: Model name: S11_base.0.2: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

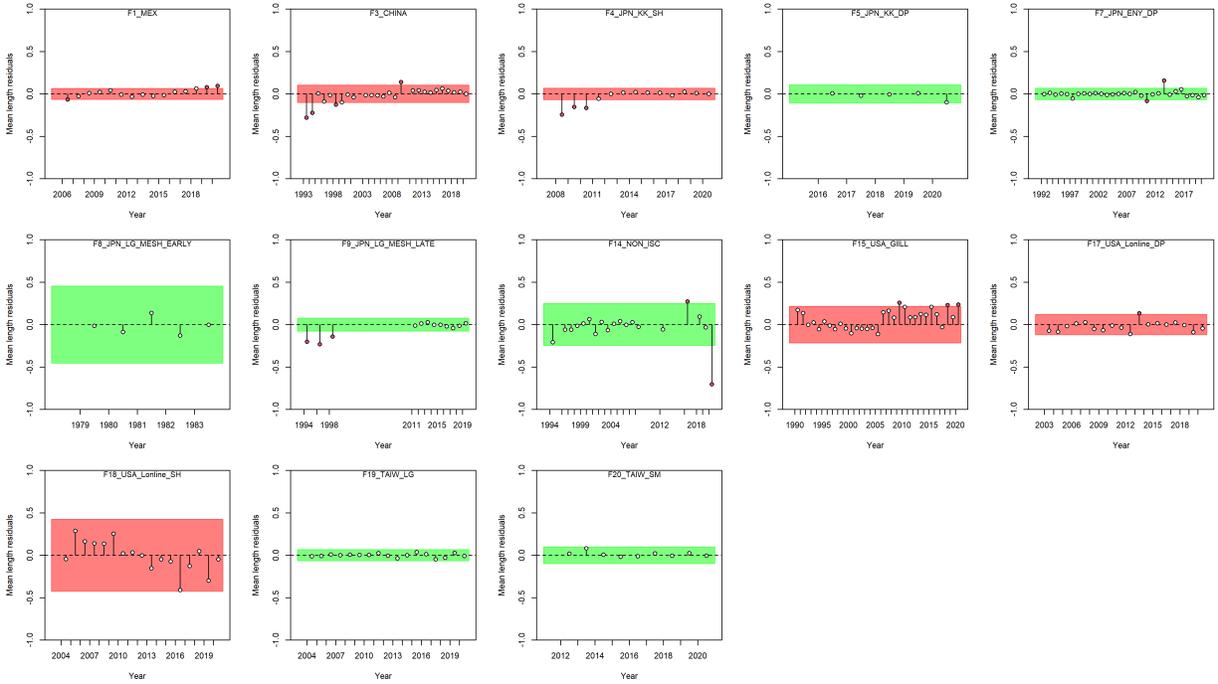


Figure 18: Model name: S11_base_0.4: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

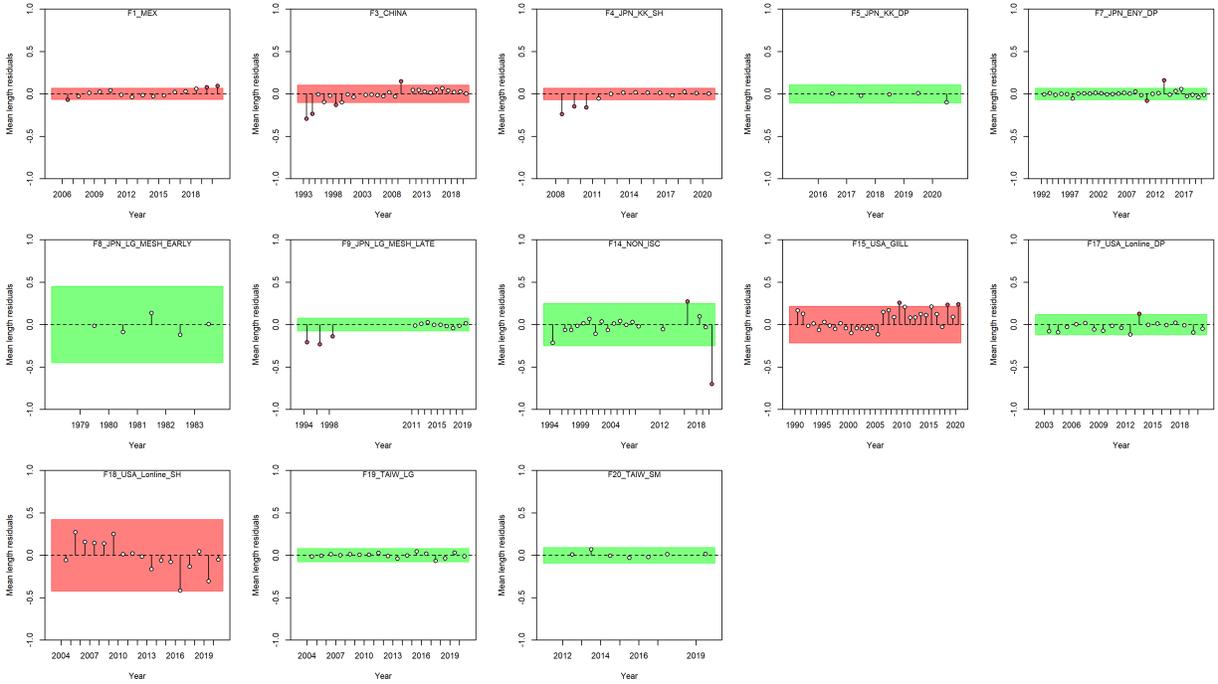


Figure 19: Model name: S11_drop.0.2: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

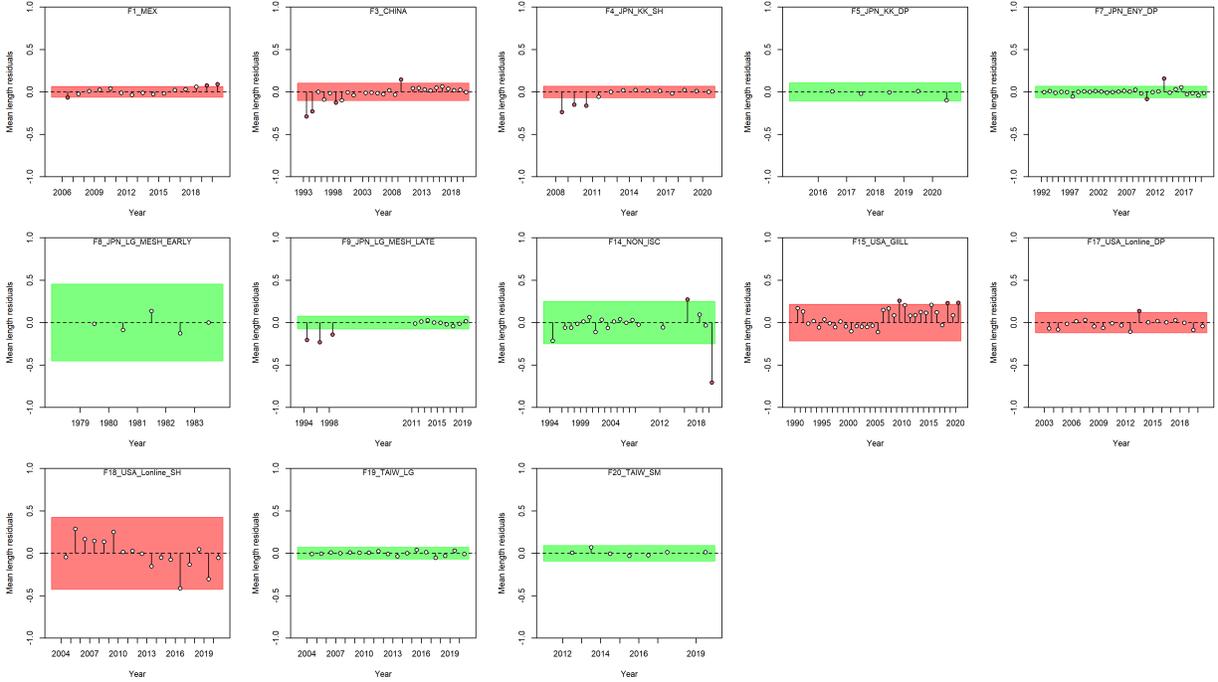


Figure 20: Model name: S11_drop.0.4: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

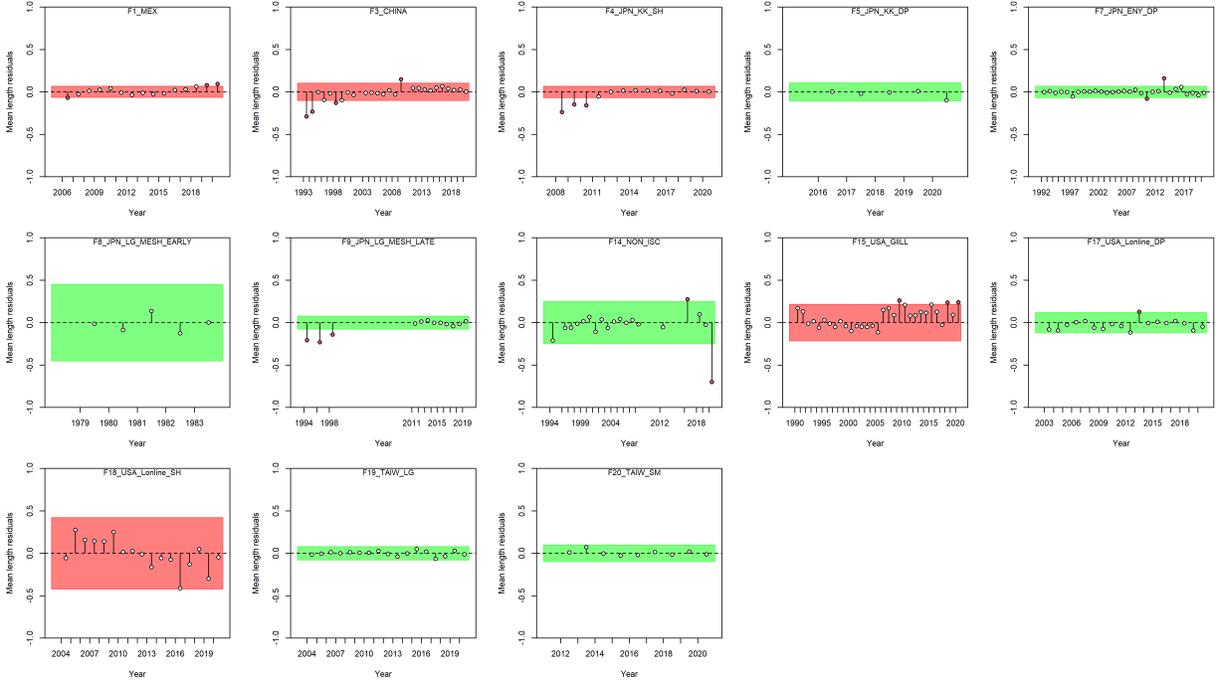


Figure 21: Model name: S11_ess.0.2: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

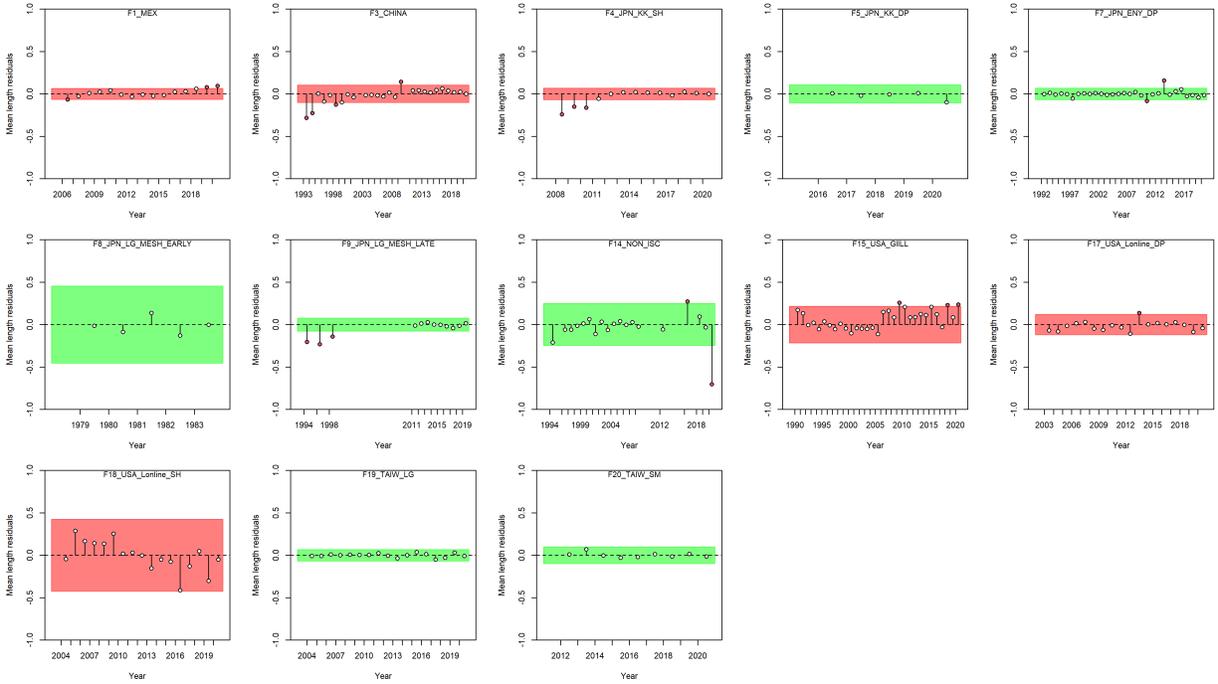


Figure 22: Model name: S11_ess.0.4: Mean length residuals runs test for the two main indices *S5_JPN_EARLY* and *S11_DFA_LATE*. Red points indicate outliers based on 3 standard deviations. The colored polygon indicates a passing (green) or failing (red) score for the residual runs test.

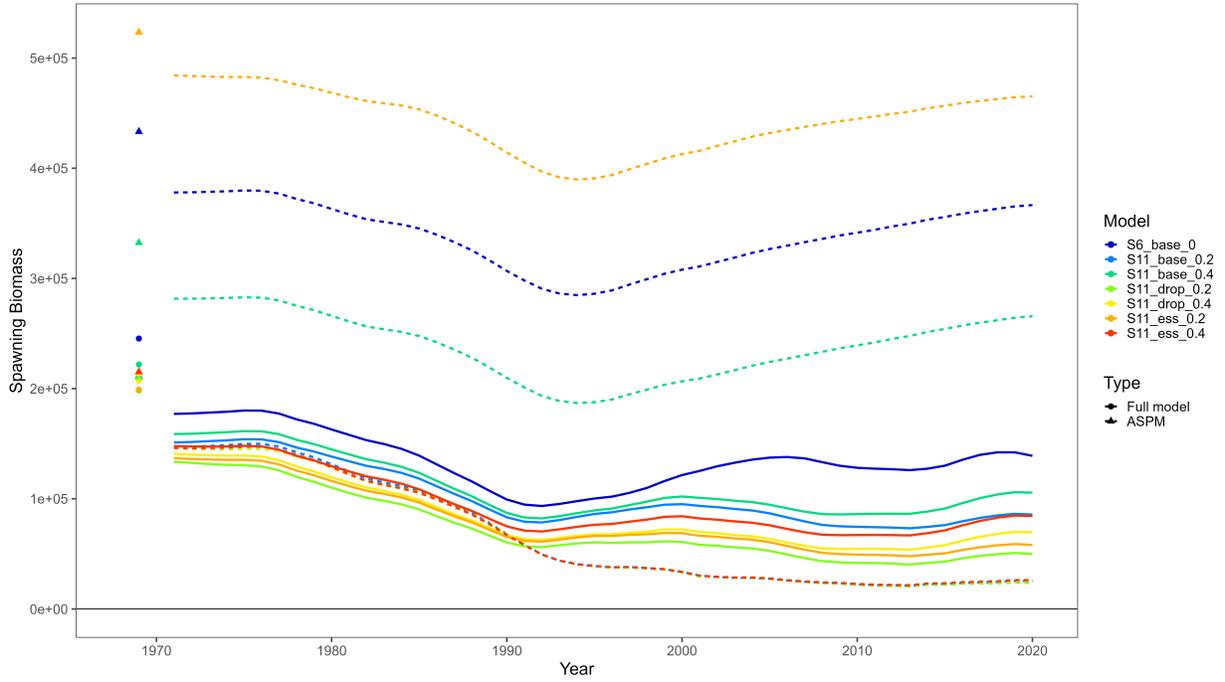


Figure 23: Summary plot of Age-structured production model (ASPM) diagnostic. Models are coded by color and type, where dotted lines show the ASPM version of the full model (solid line).

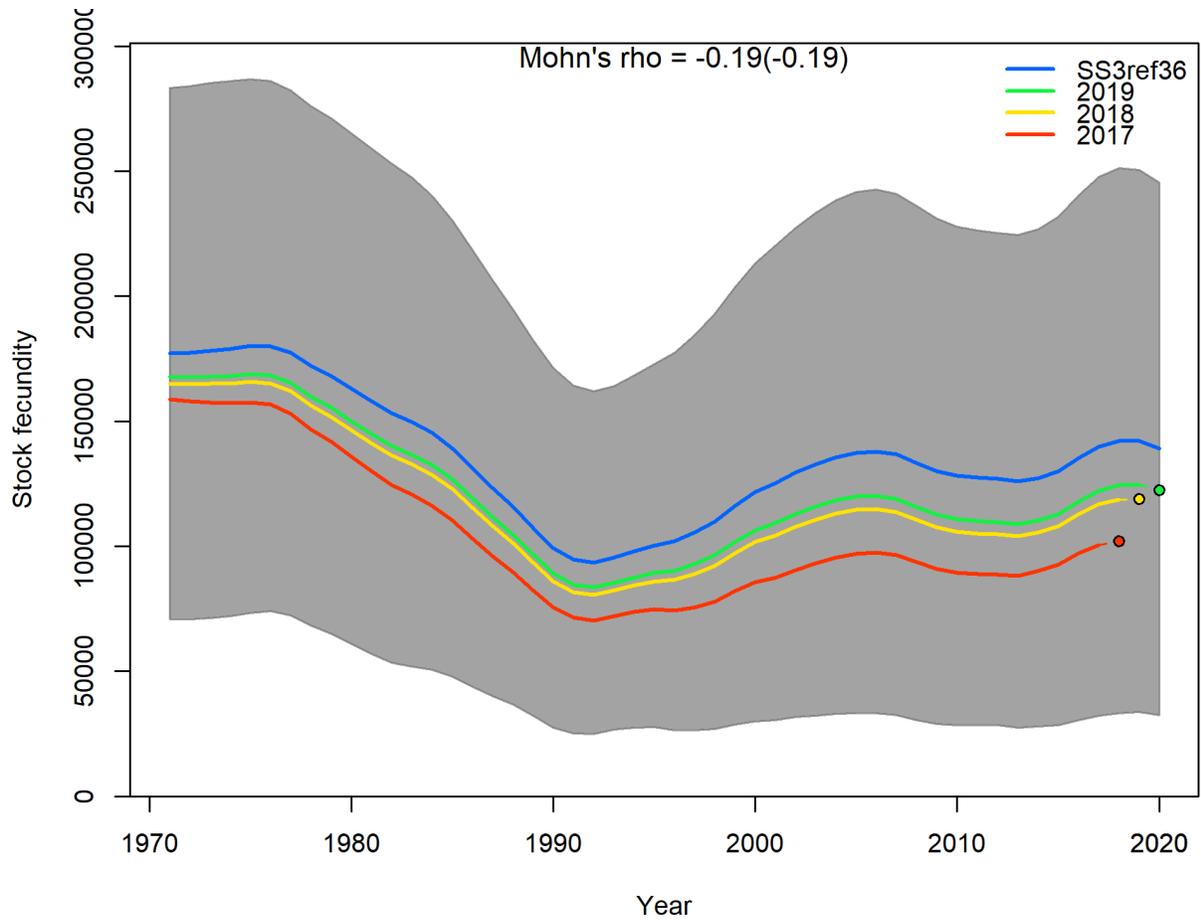


Figure 24: Model name: S6_base_0: Three year retrospective analysis of spawning biomass.

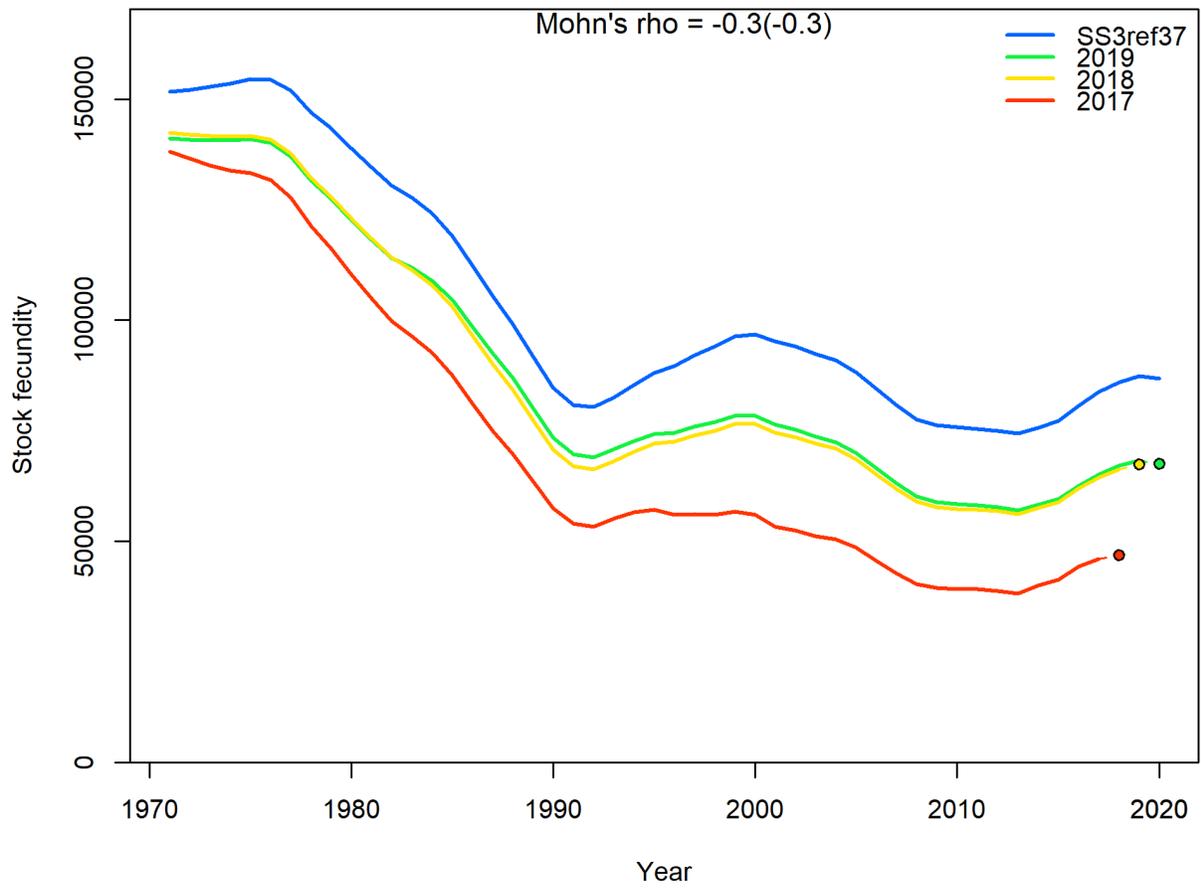


Figure 25: Model name: S11_base_0.2: Three year retrospective analysis of spawning biomass.

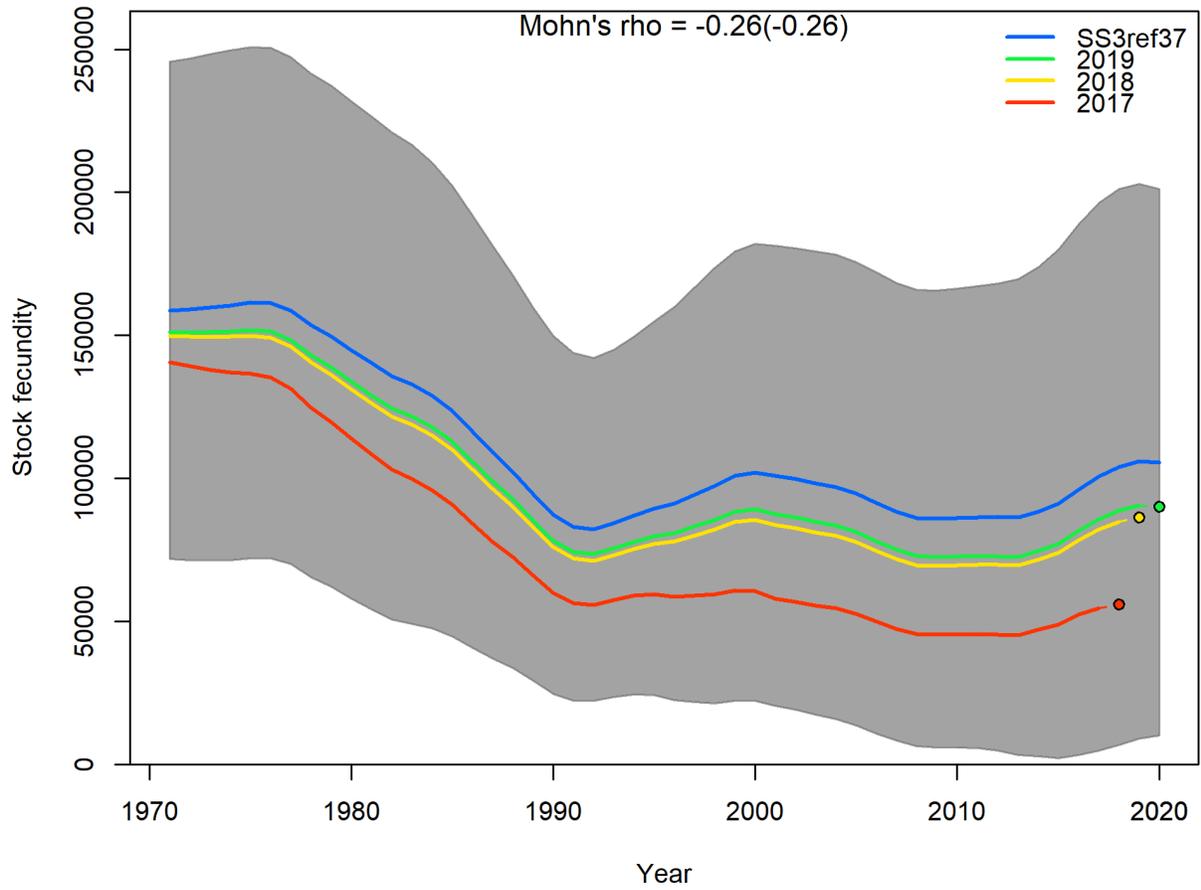


Figure 26: Model name: S11_base_0.4: Three year retrospective analysis of spawning biomass.

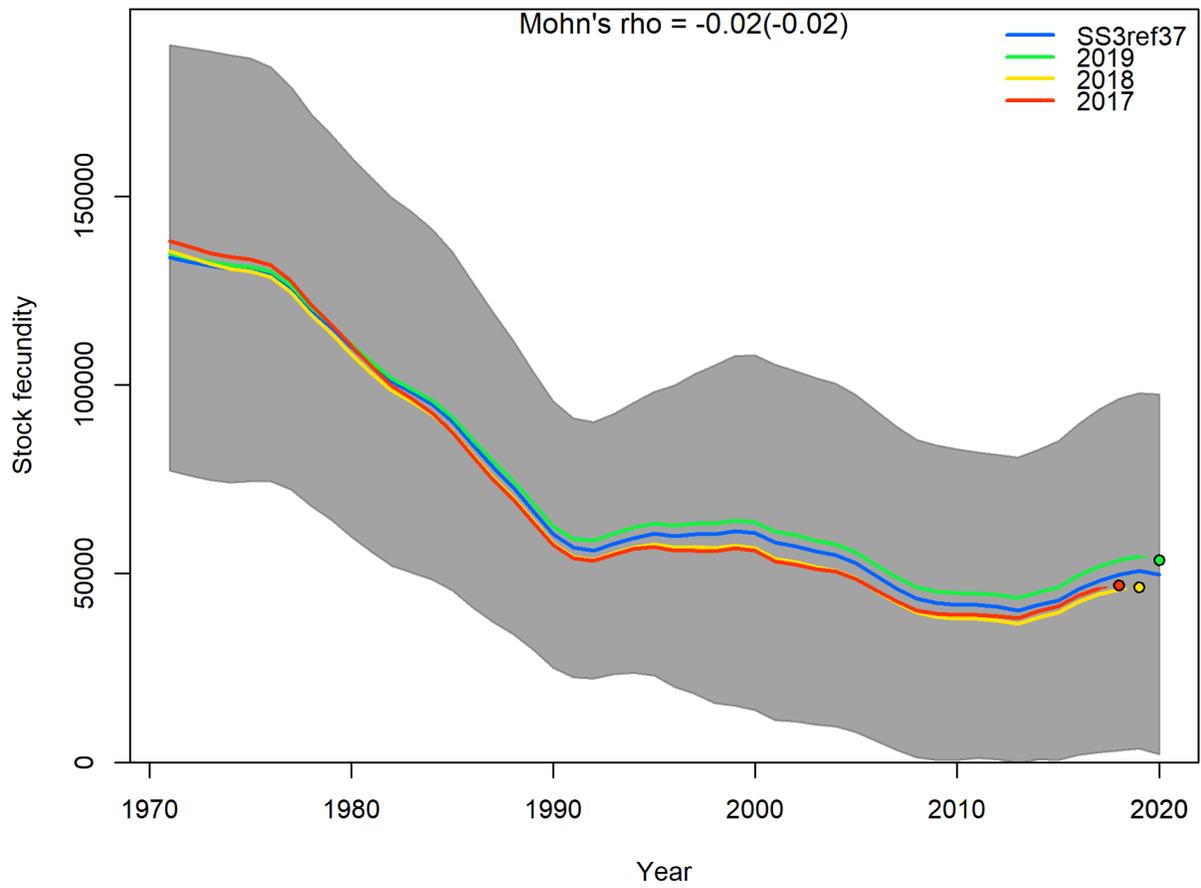


Figure 27: Model name: S11_drop_0.2: Three year retrospective analysis of spawning biomass.

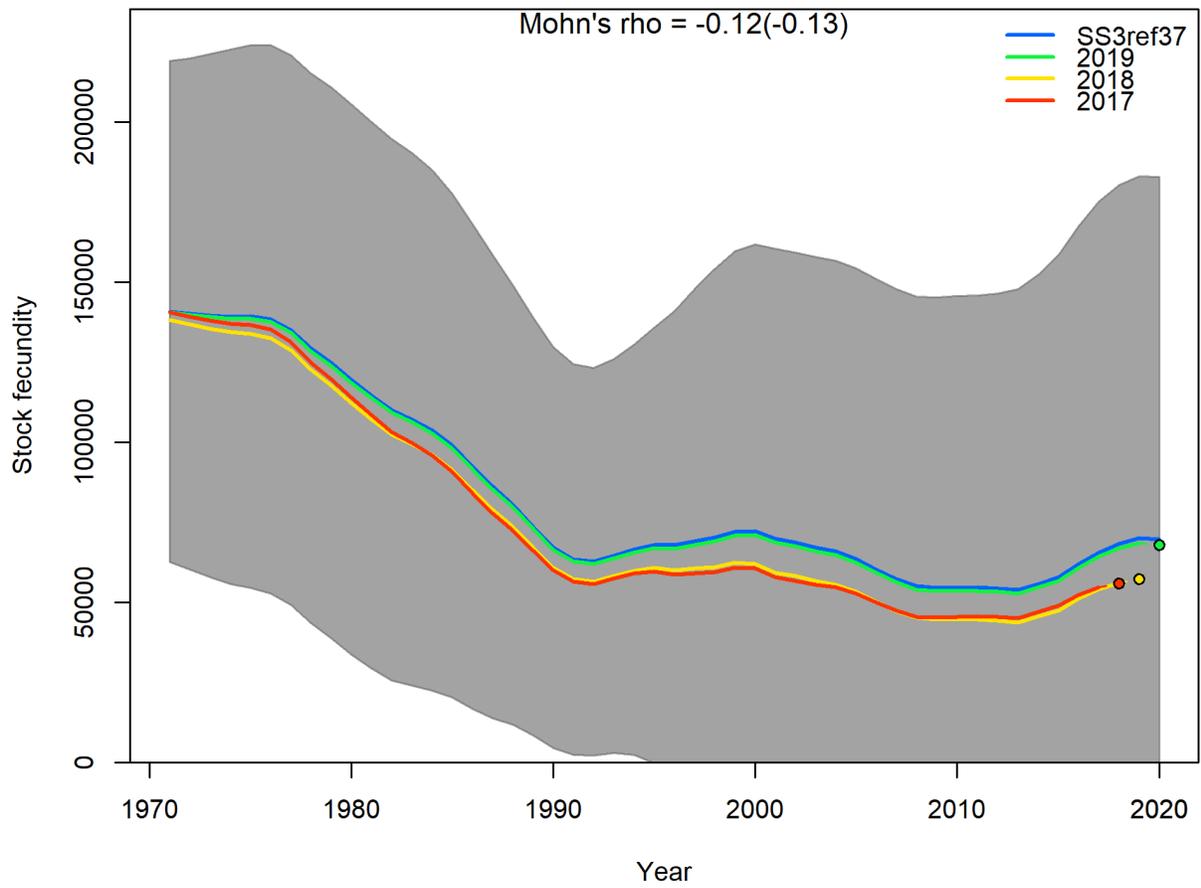


Figure 28: Model name: S11_drop_0.4: Three year retrospective analysis of spawning biomass.

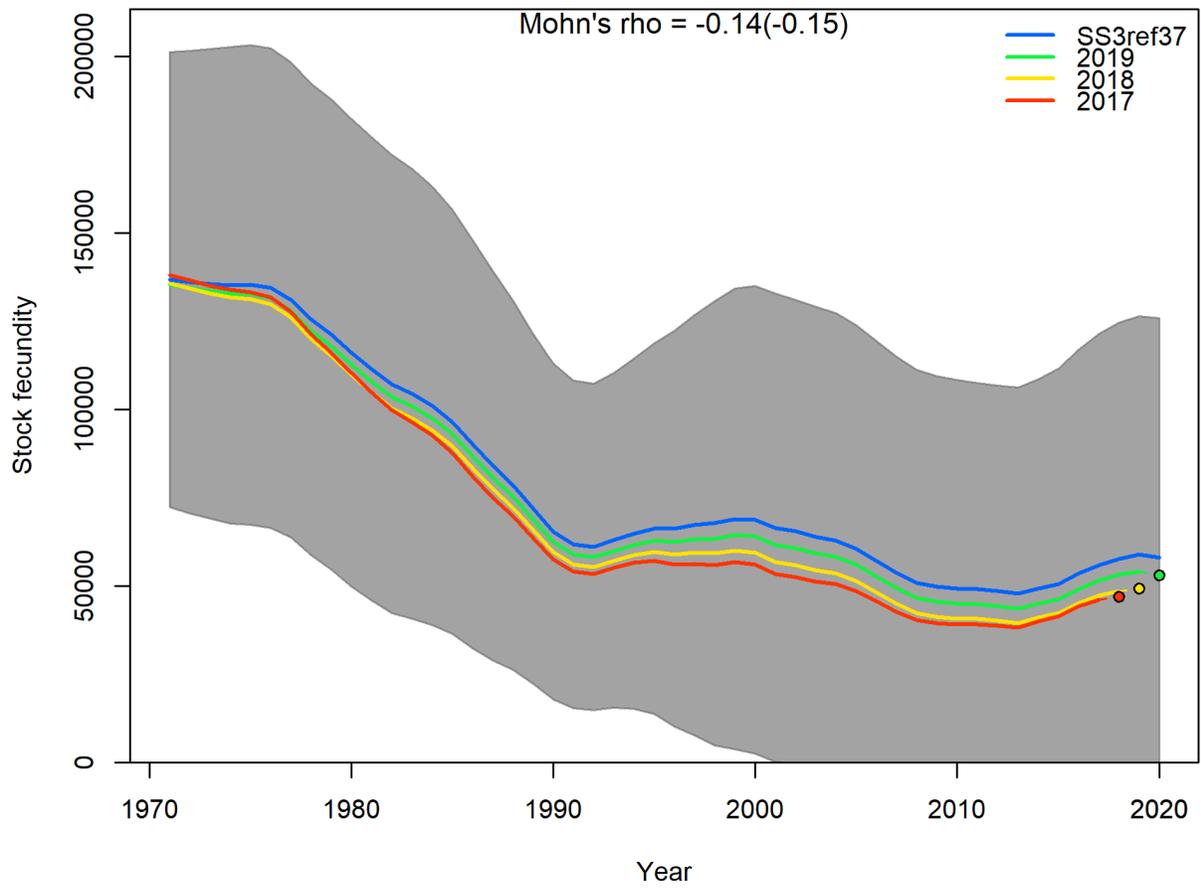


Figure 29: Model name: S11_ess_0.2: Three year retrospective analysis of spawning biomass.

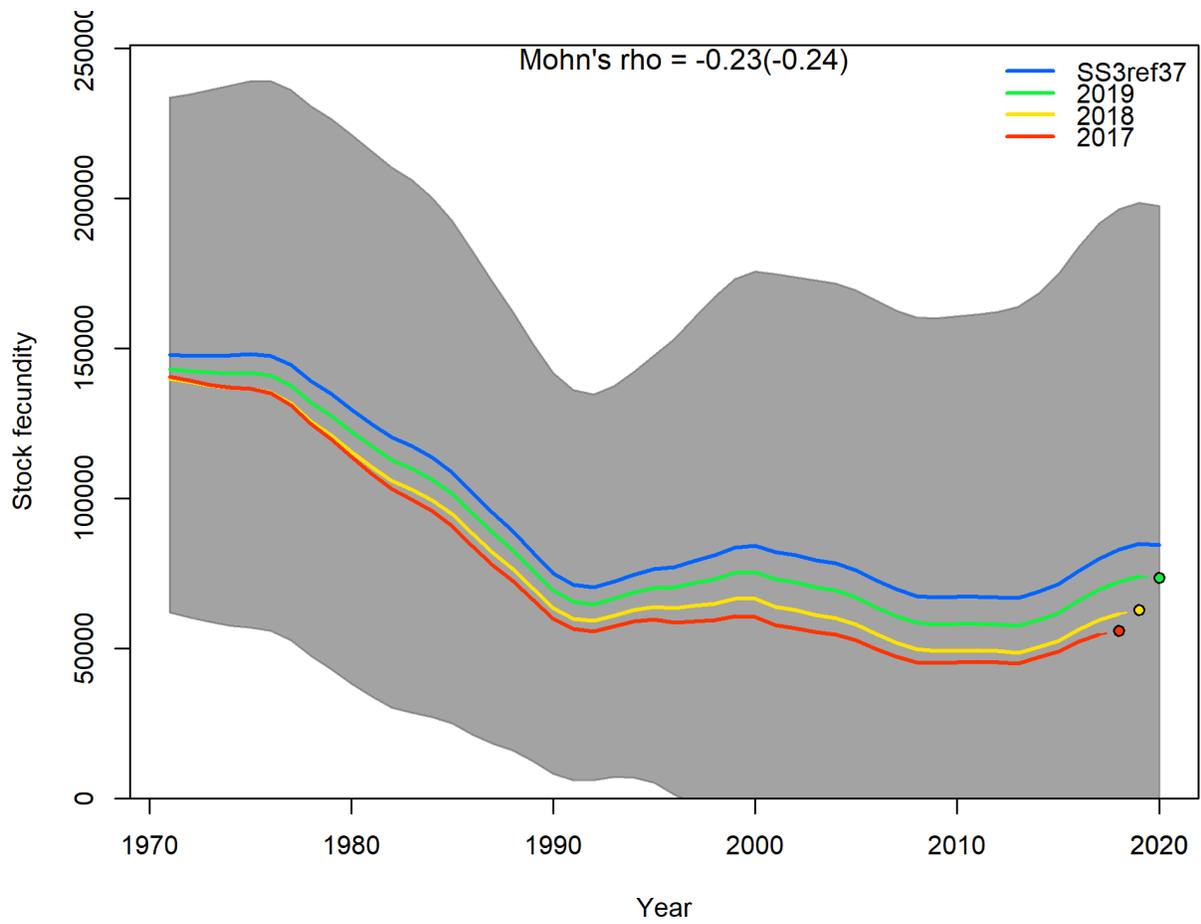


Figure 30: Model name: S11_ess_0.4: Three year retrospective analysis of spawning biomass.

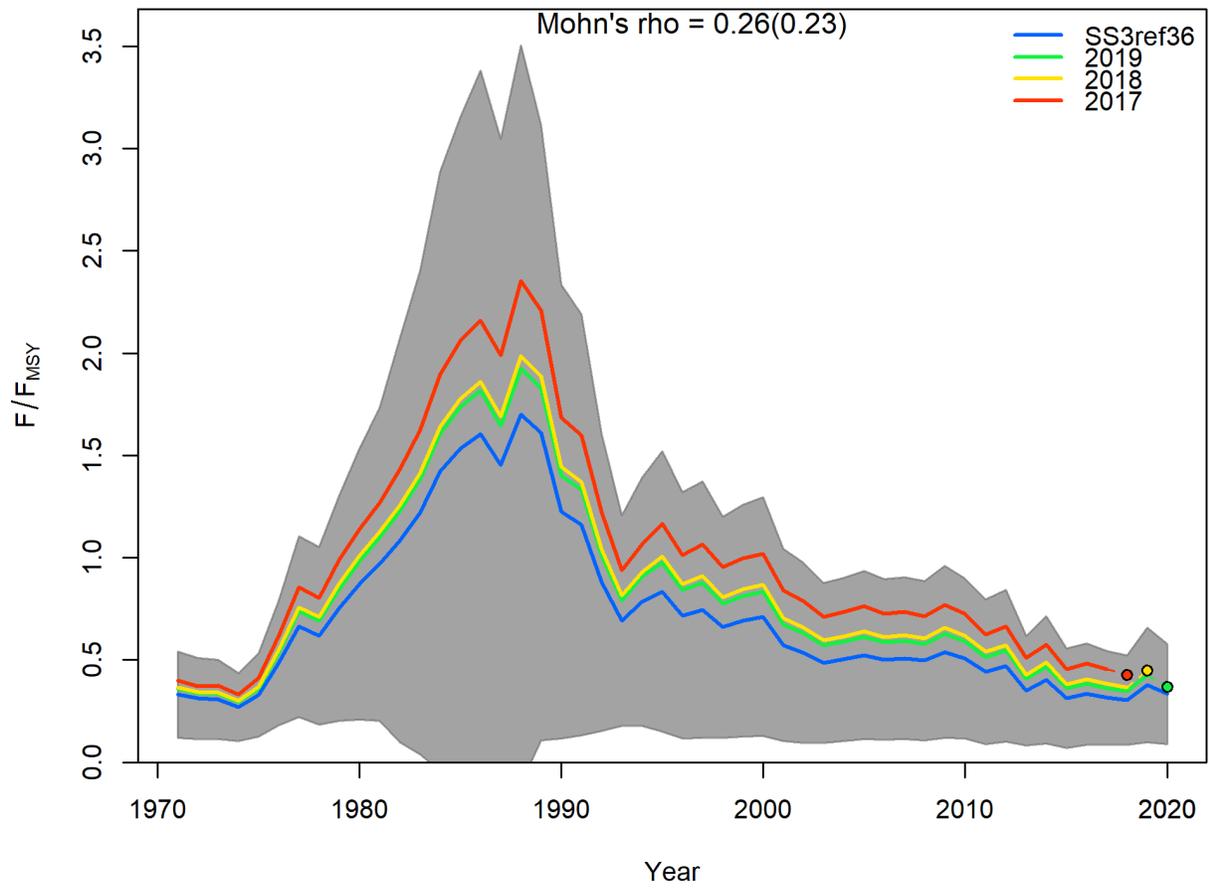


Figure 31: Model name: S6_base_0: Three year retrospective analysis of fishing mortality.

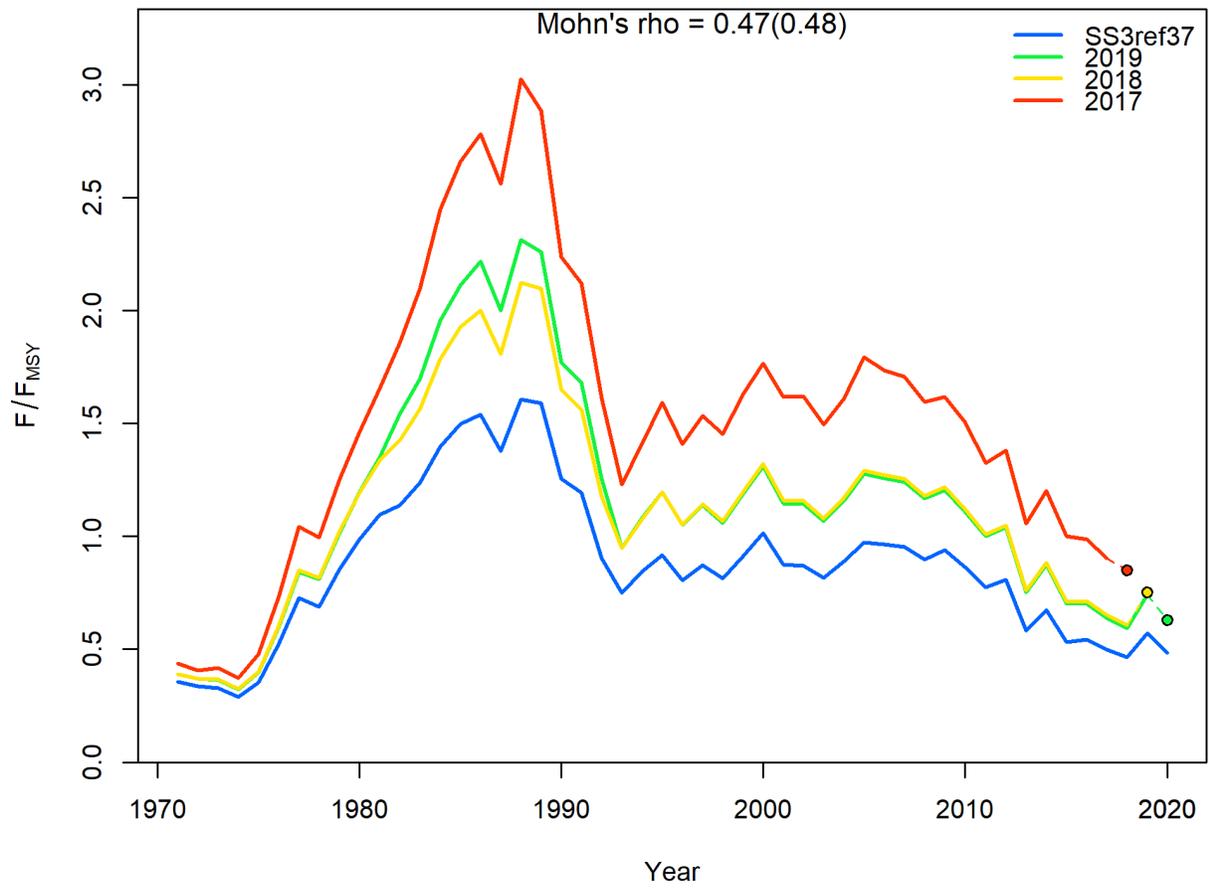


Figure 32: Model name: S11.base.0.2: Three year retrospective analysis of fishing mortality.

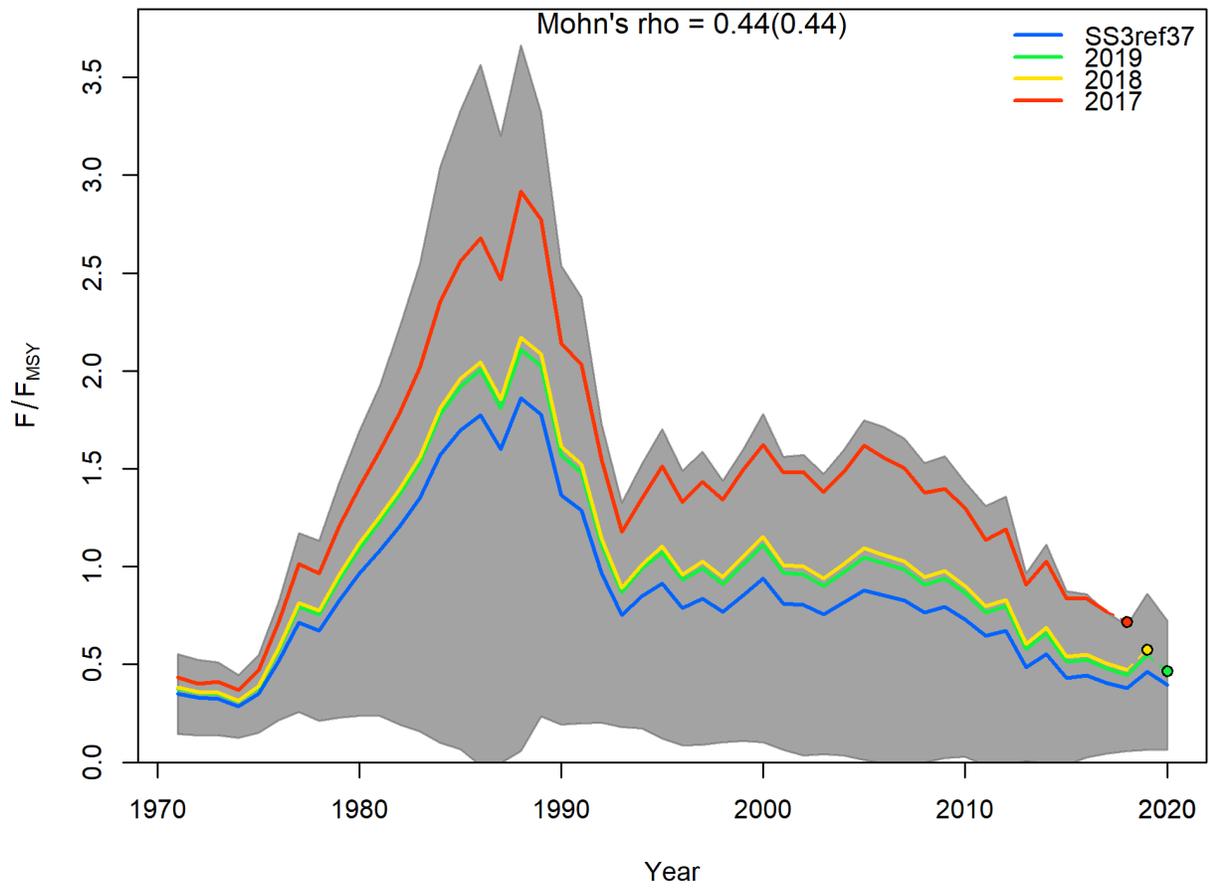


Figure 33: Model name: S11_base.0.4: Three year retrospective analysis of fishing mortality.

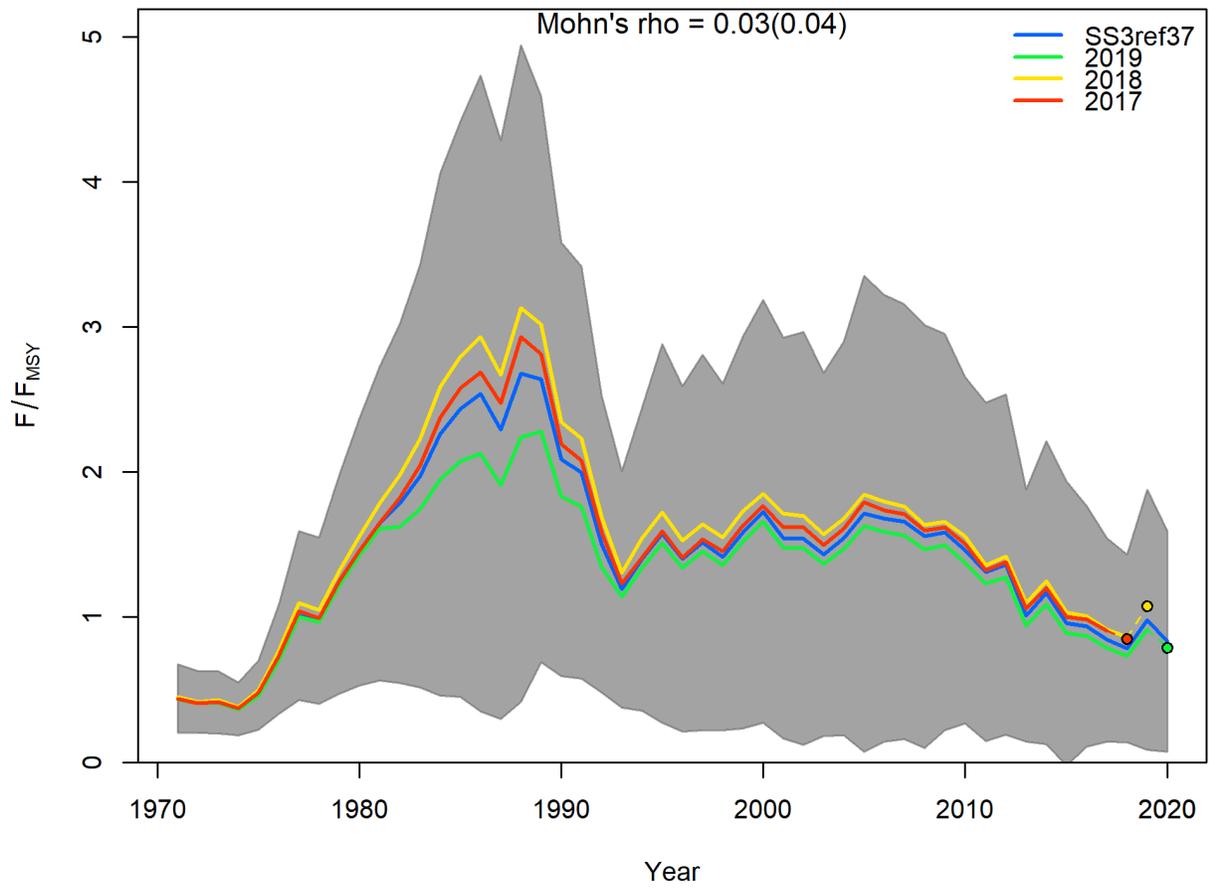


Figure 34: Model name: S11_drop_0.2: Three year retrospective analysis of fishing mortality.

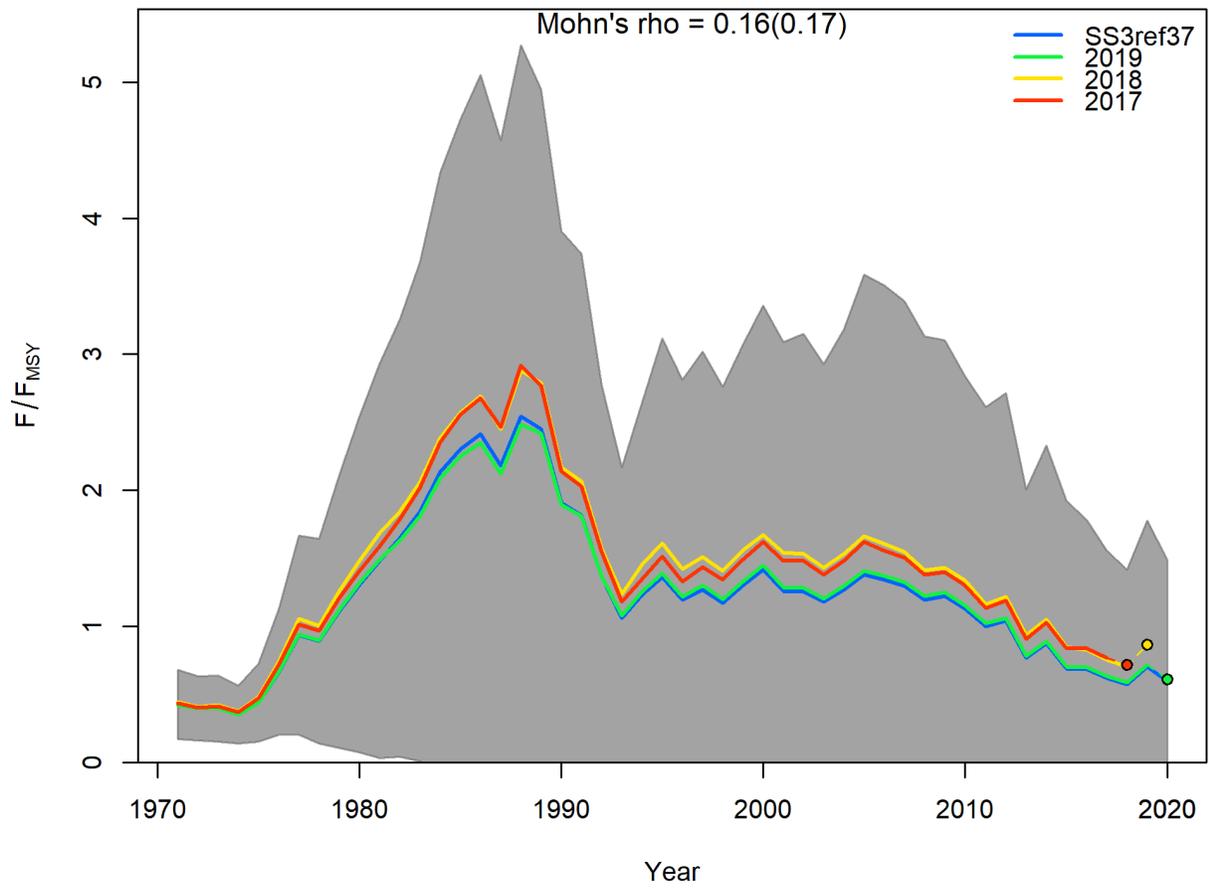


Figure 35: Model name: S11_drop_0.4: Three year retrospective analysis of fishing mortality.

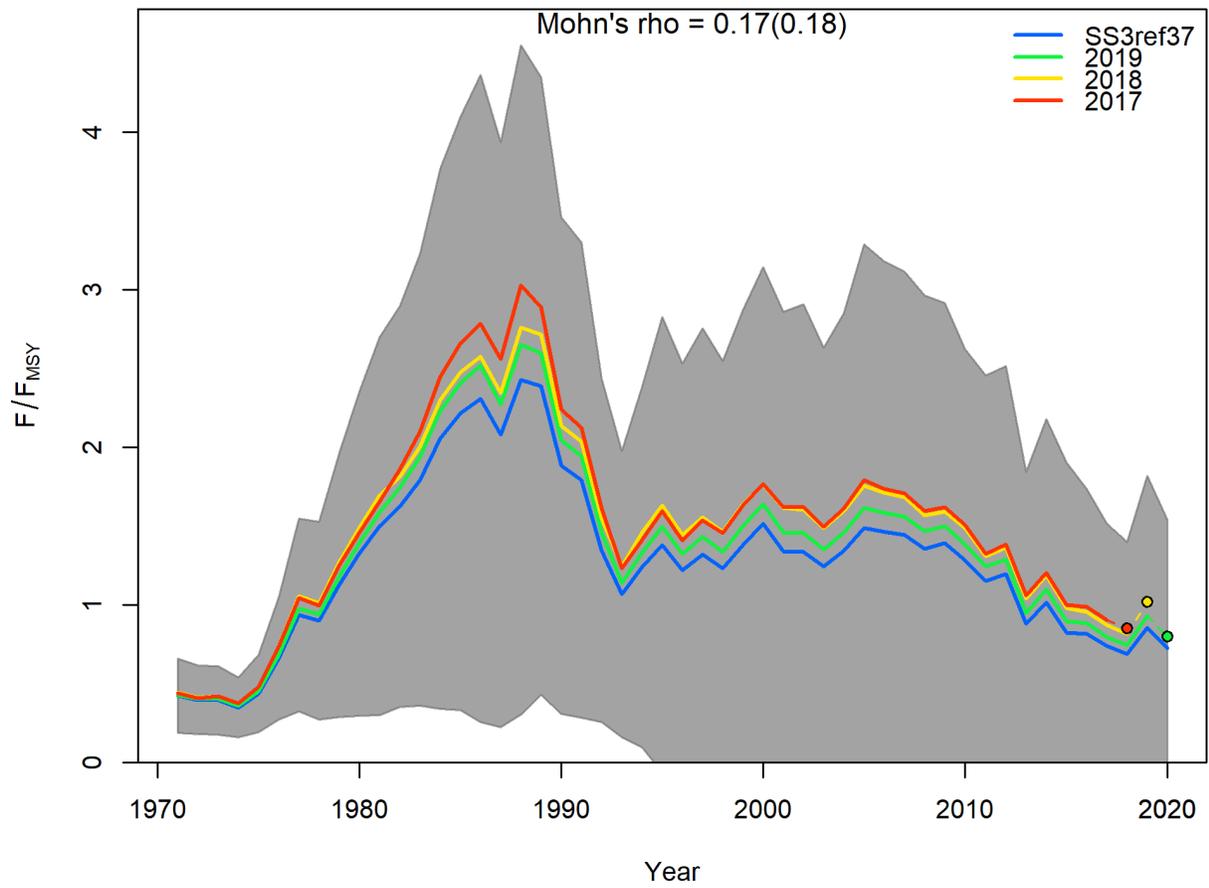


Figure 36: Model name: S11_ess_0.2: Three year retrospective analysis of fishing mortality.

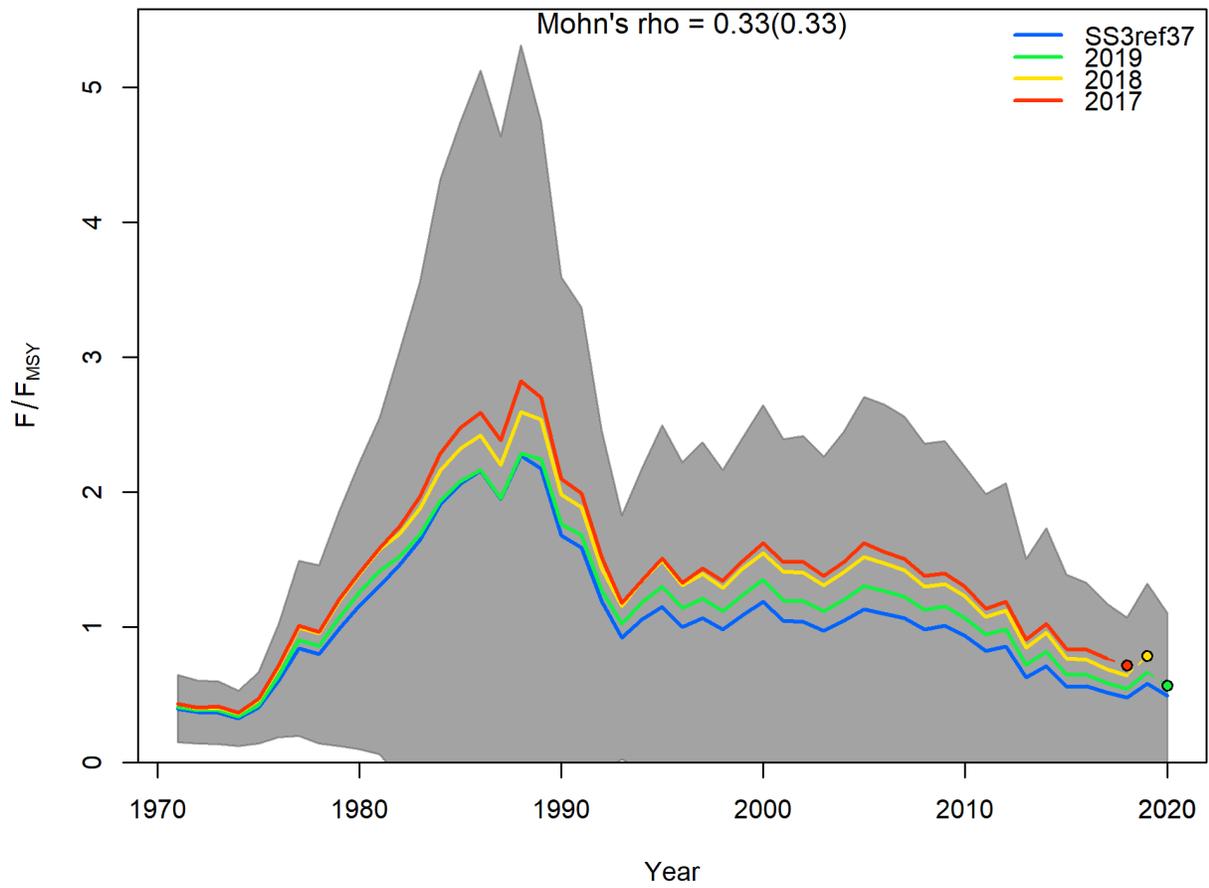


Figure 37: Model name: S11_ess_0.4: Three year retrospective analysis of fishing mortality.

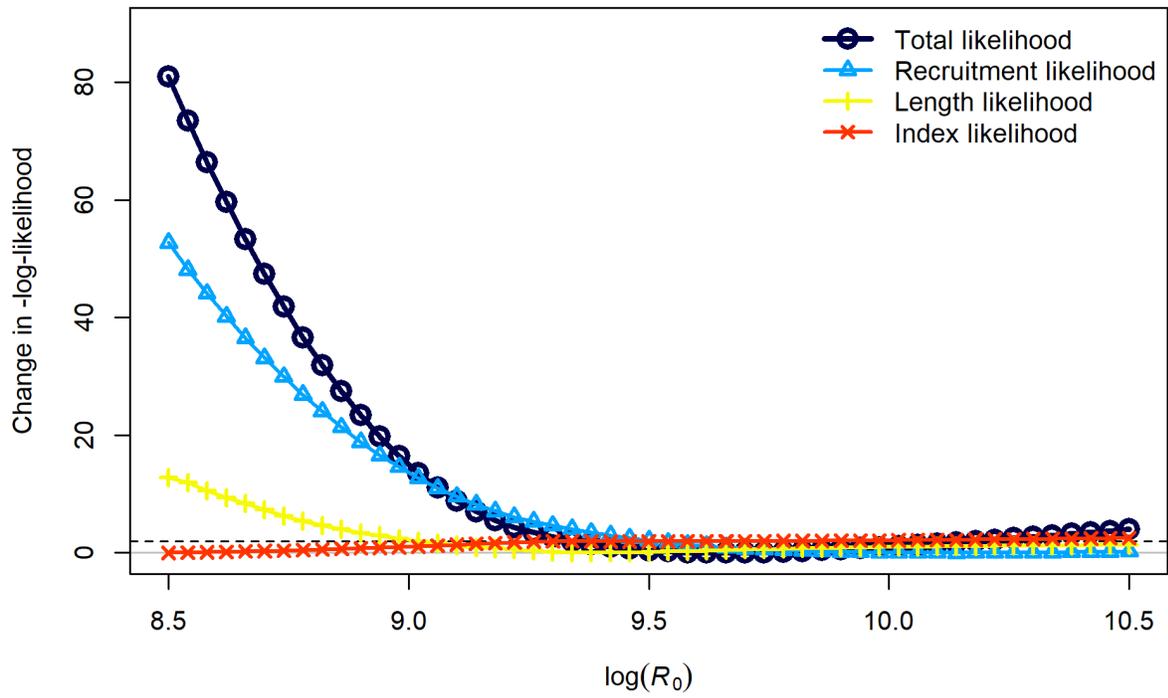


Figure 38: Model name: S6_base_0: R0 profile for three main likelihood components: recruitment, index, and length composition.

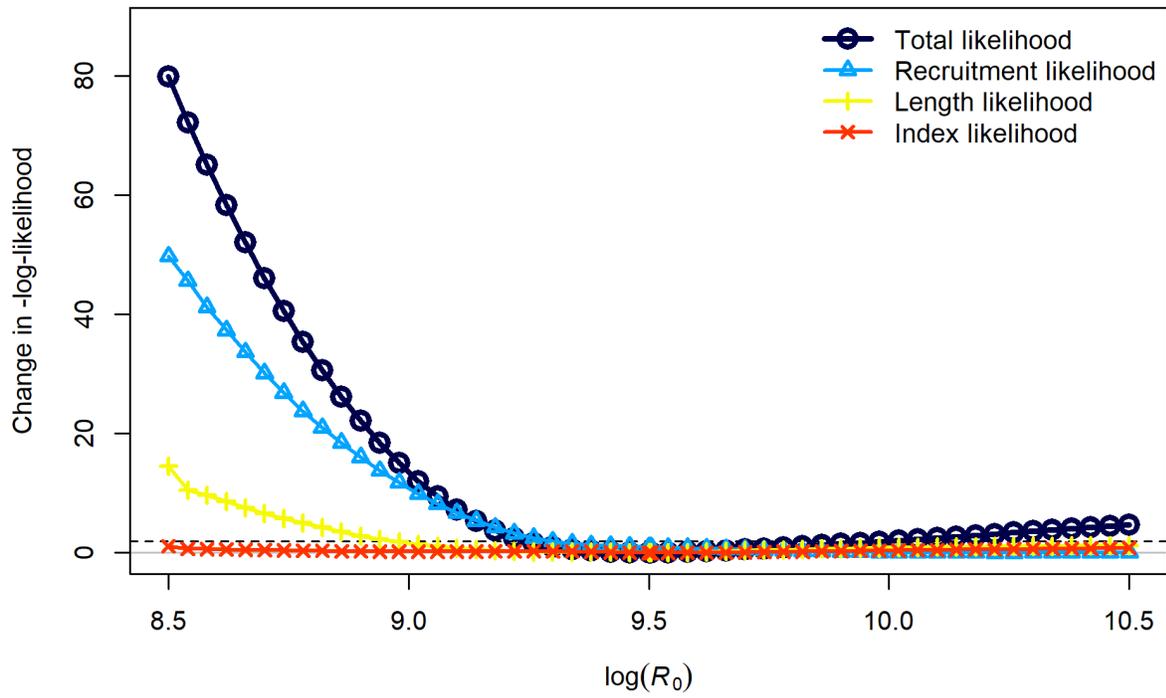


Figure 39: Model name: S11_base-0.2: R0 profile for three main likelihood components: recruitment, index, and length composition.

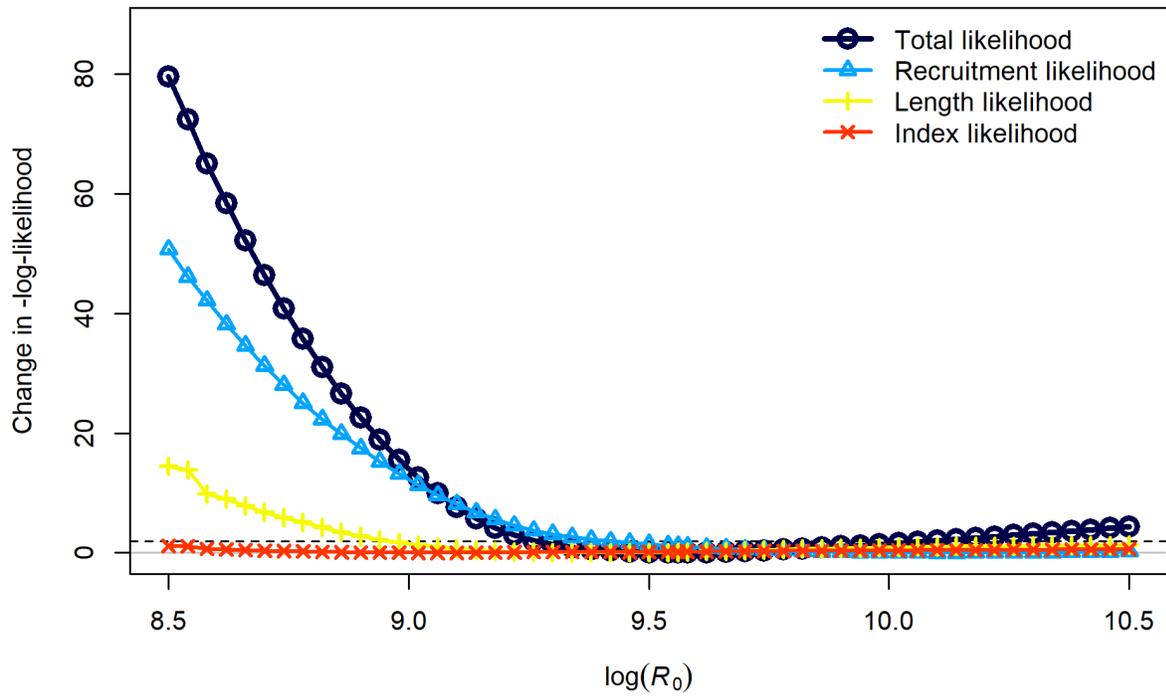


Figure 40: Model name: S11_base_0.4: R0 profile for three main likelihood components: recruitment, index, and length composition.

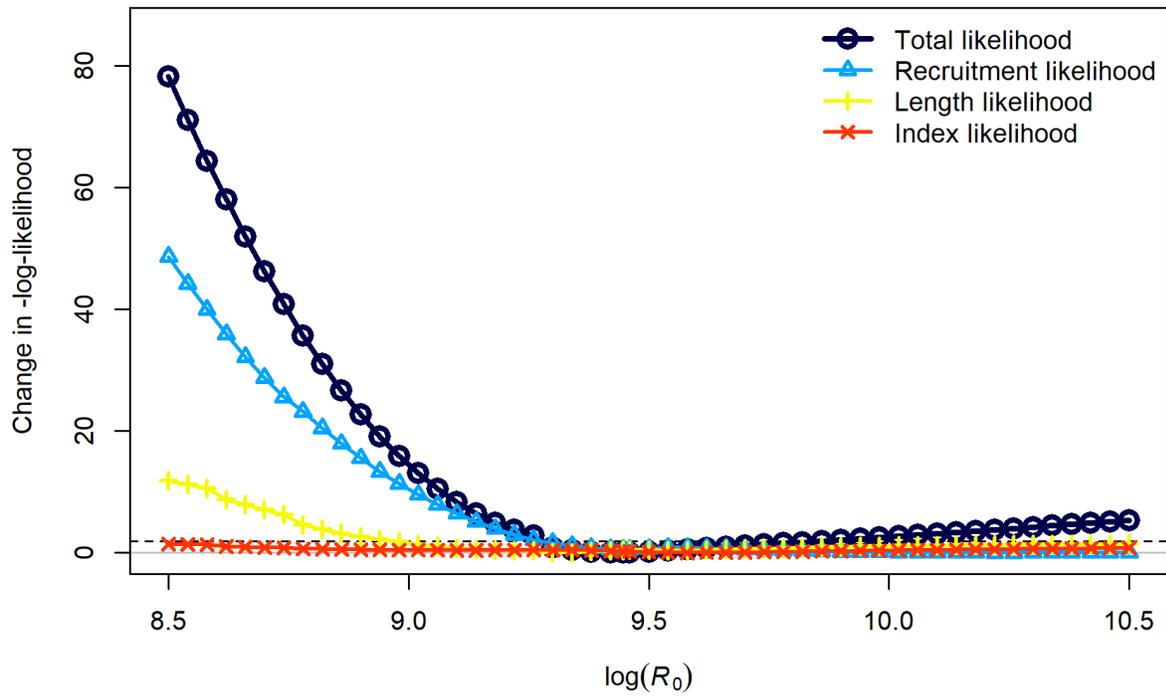


Figure 41: Model name: S11_drop_0.2: R_0 profile for three main likelihood components: recruitment, index, and length composition.

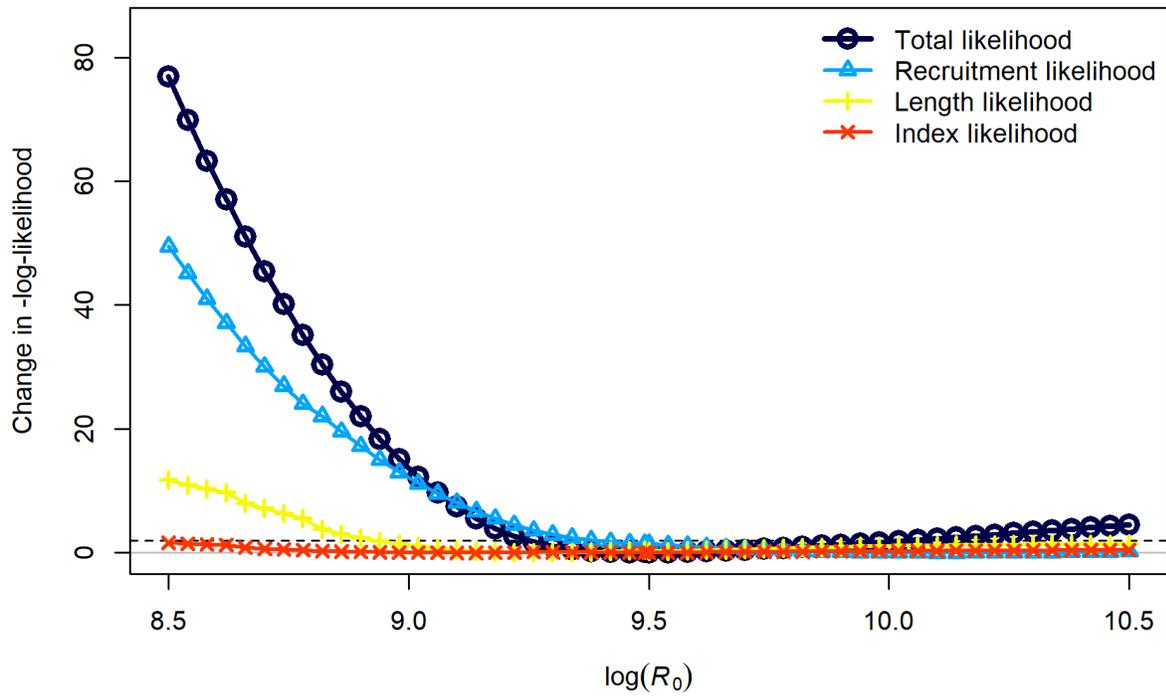


Figure 42: Model name: S11_drop_0.4: R0 profile for three main likelihood components: recruitment, index, and length composition.

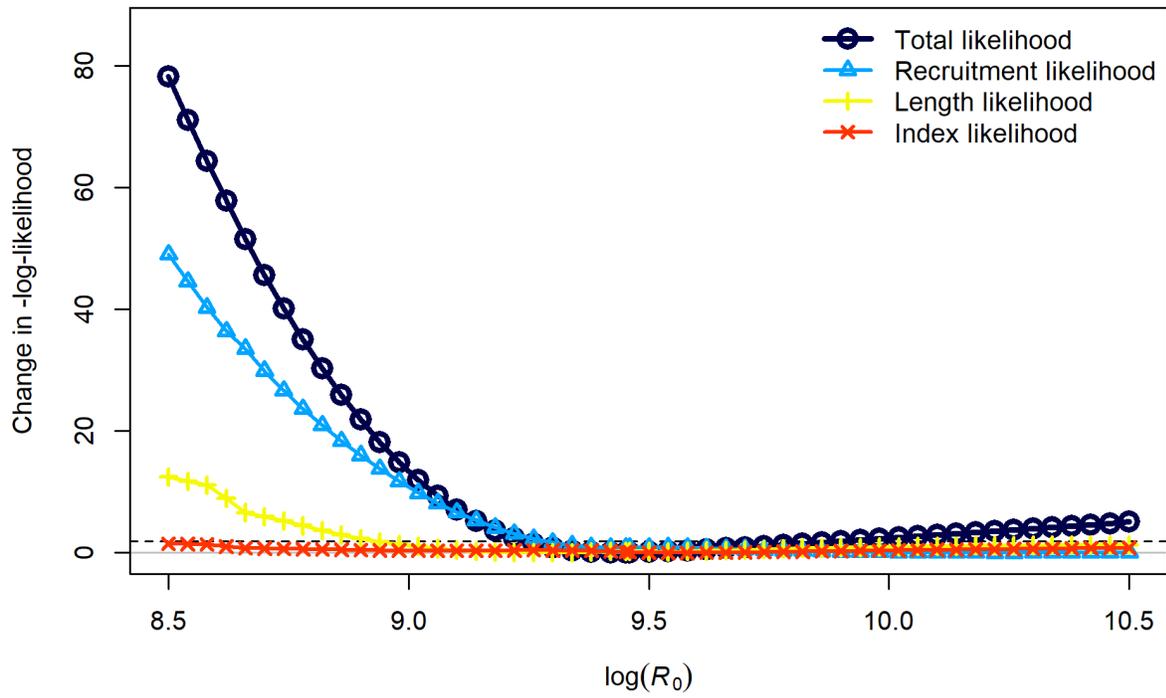


Figure 43: Model name: S11_ess_0.2: R0 profile for three main likelihood components: recruitment, index, and length composition.

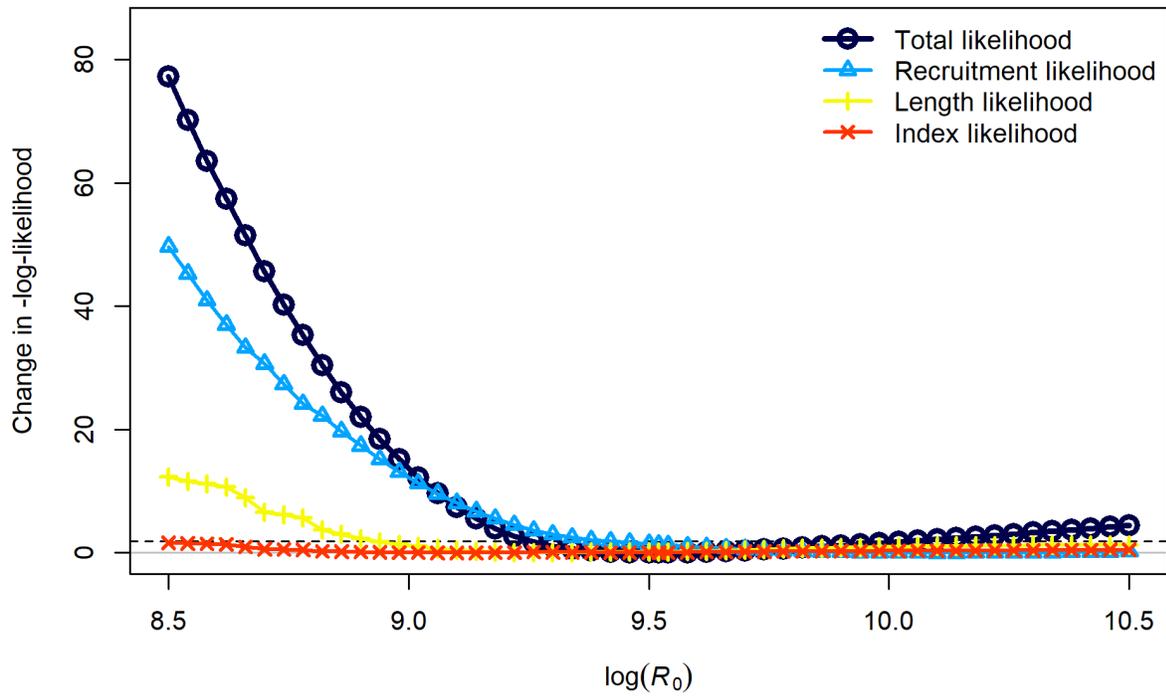


Figure 44: Model name: S11_ess_0.4: R0 profile for three main likelihood components: recruitment, index, and length composition.

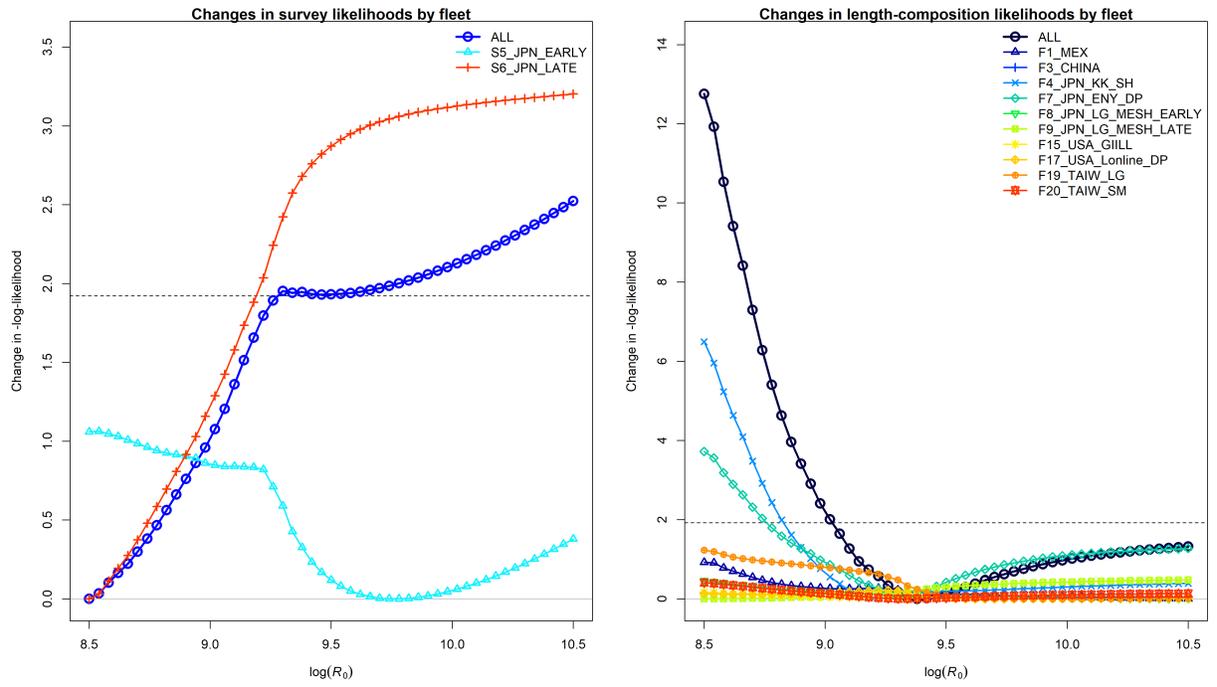


Figure 45: Model name: S6_base_0: R0 profile for two main likelihood components by fleet: survey index and length composition.

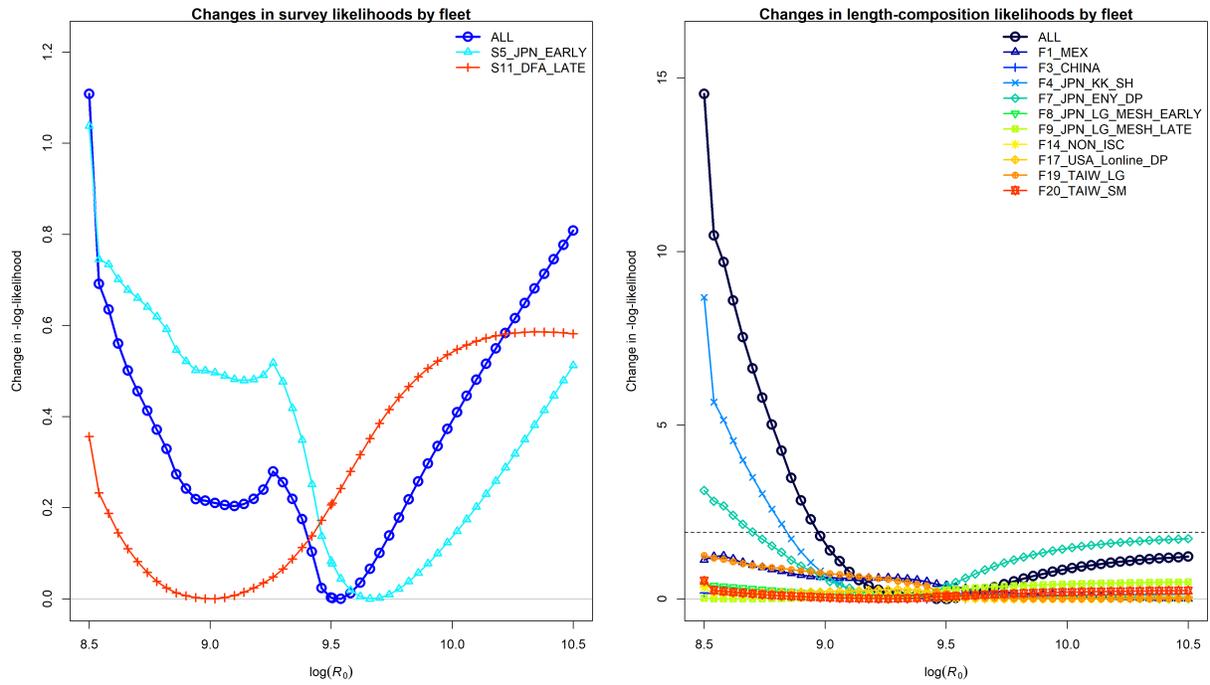


Figure 46: Model name: S11_base_0.2: R_0 profile for two main likelihood components by fleet: survey index and length composition.

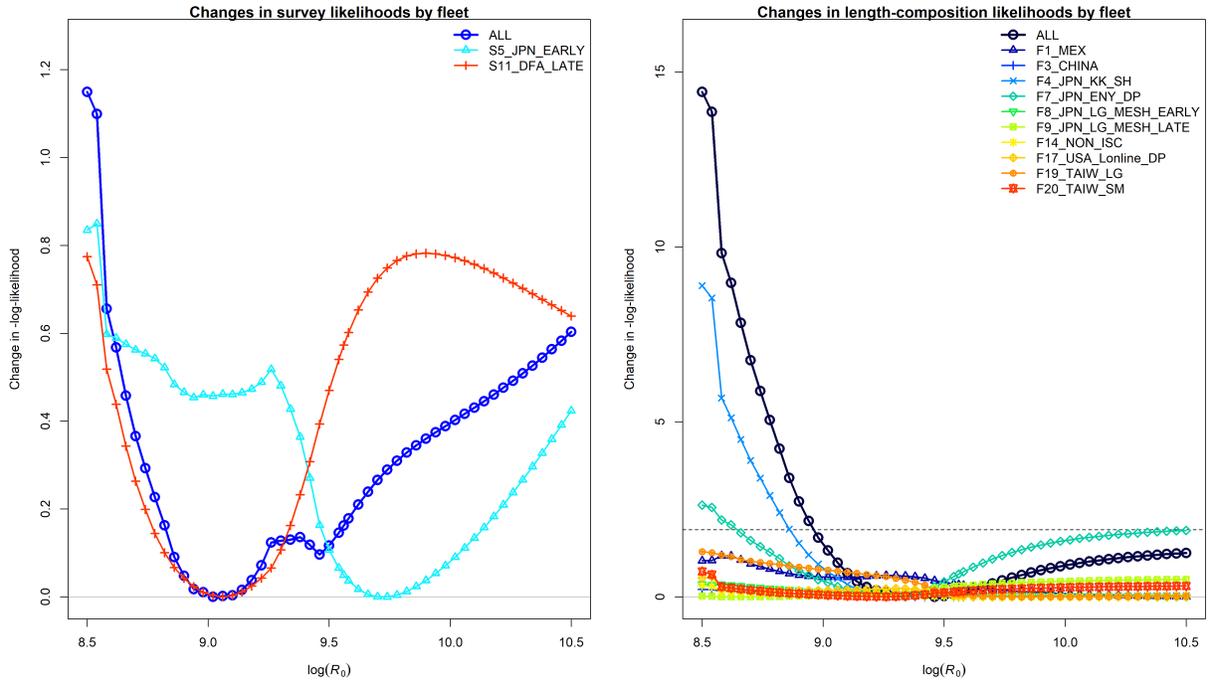


Figure 47: Model name: S11_base_0.4: R₀ profile for two main likelihood components by fleet: survey index and length composition.

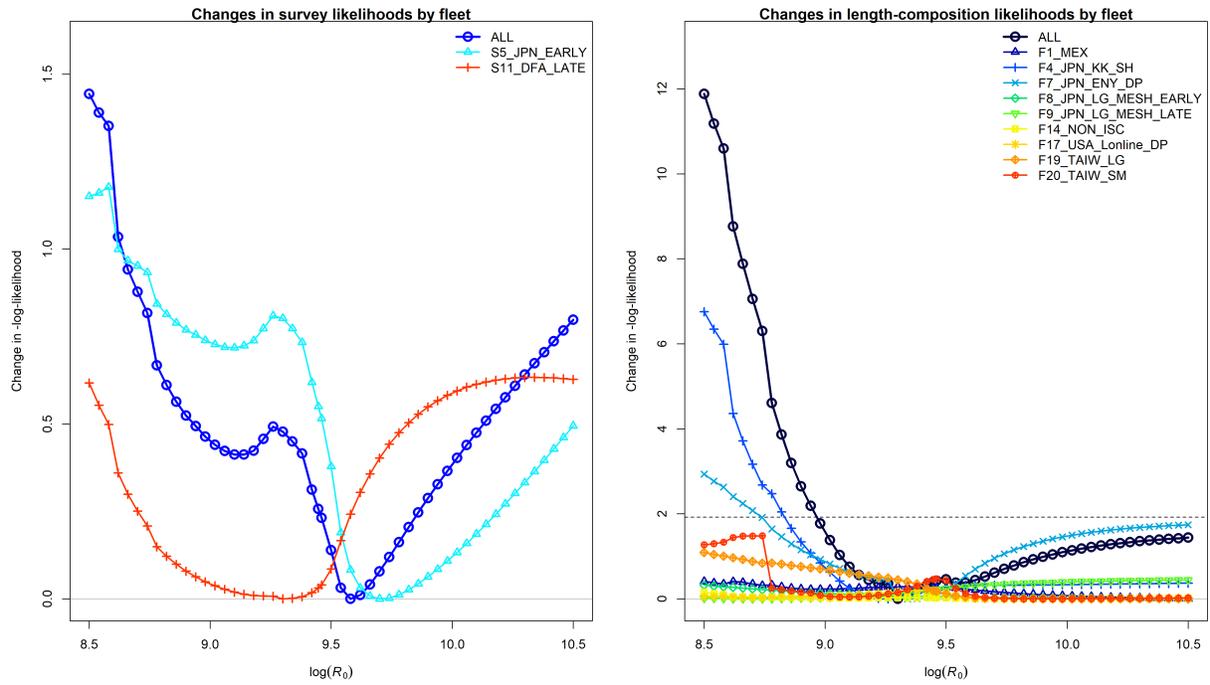


Figure 48: Model name: S11_drop_0.2: R_0 profile for two main likelihood components by fleet: survey index and length composition.

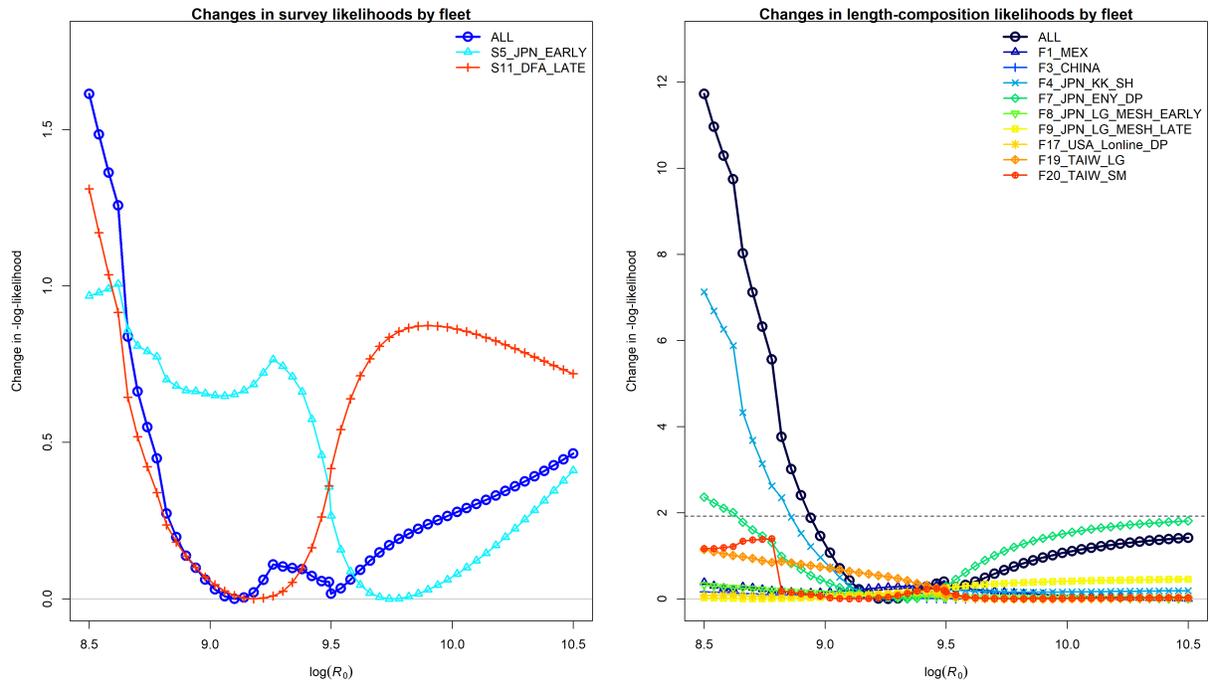


Figure 49: Model name: S11_drop_0.4: R_0 profile for two main likelihood components by fleet: survey index and length composition.

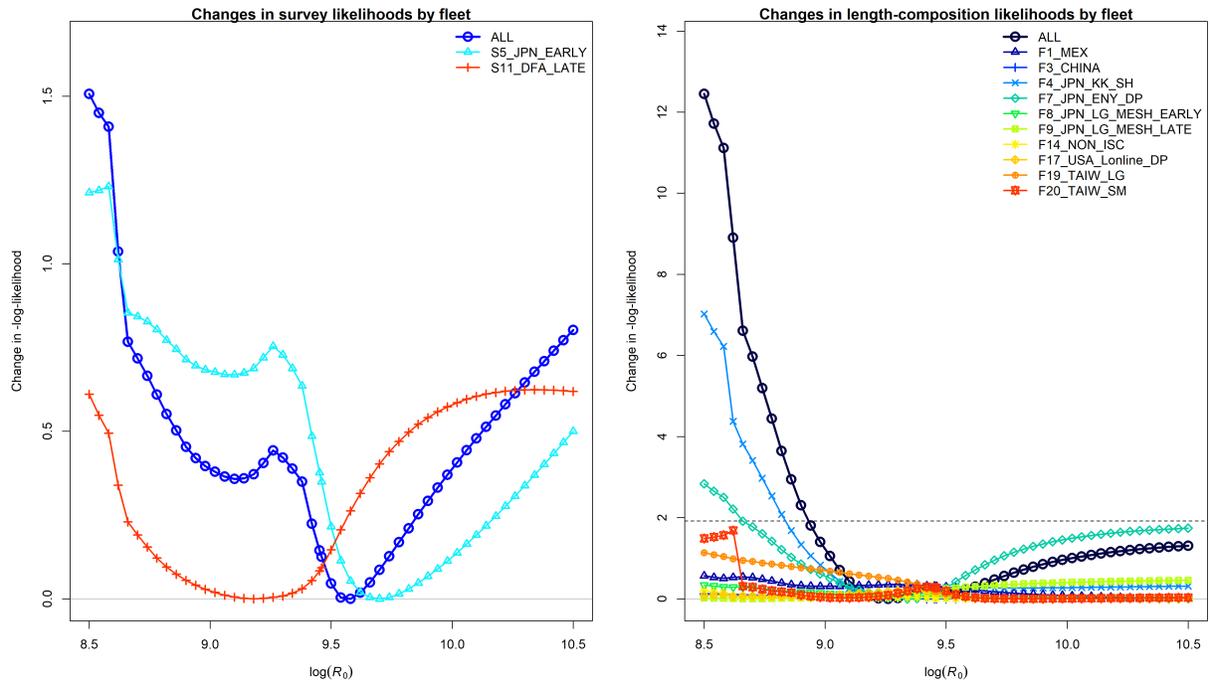


Figure 50: Model name: S11_ess_0.2: R0 profile for two main likelihood components by fleet: survey index and length composition.

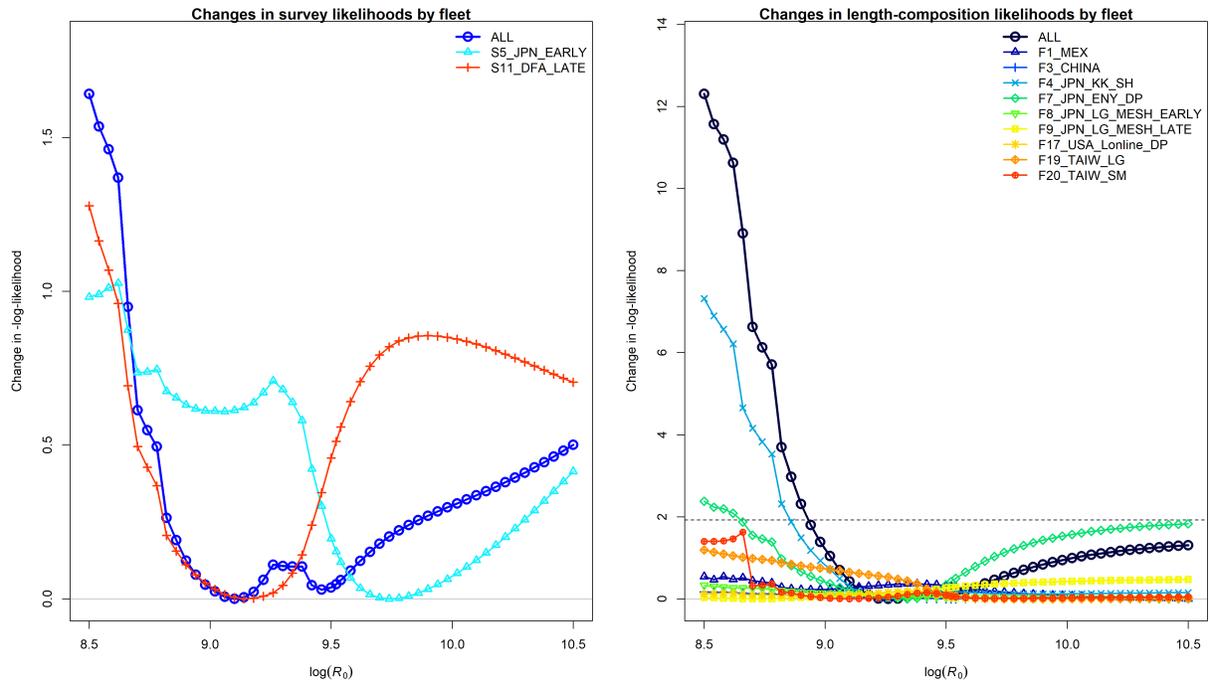


Figure 51: Model name: S11_ess_0.4: R0 profile for two main likelihood components by fleet: survey index and length composition.

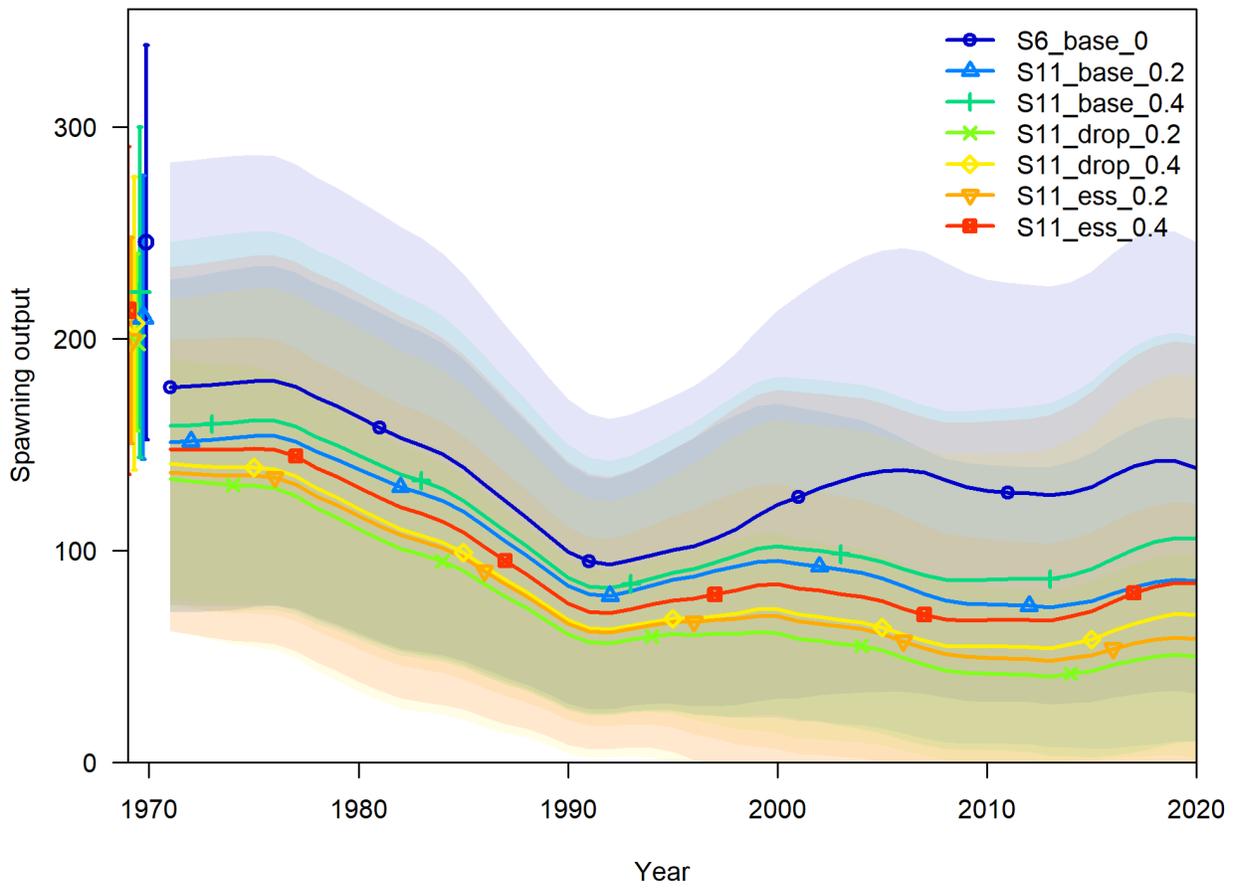


Figure 52: Spawning biomass (with uncertainty) for each of the six S11 candidate models with the S6 model shown for comparison.

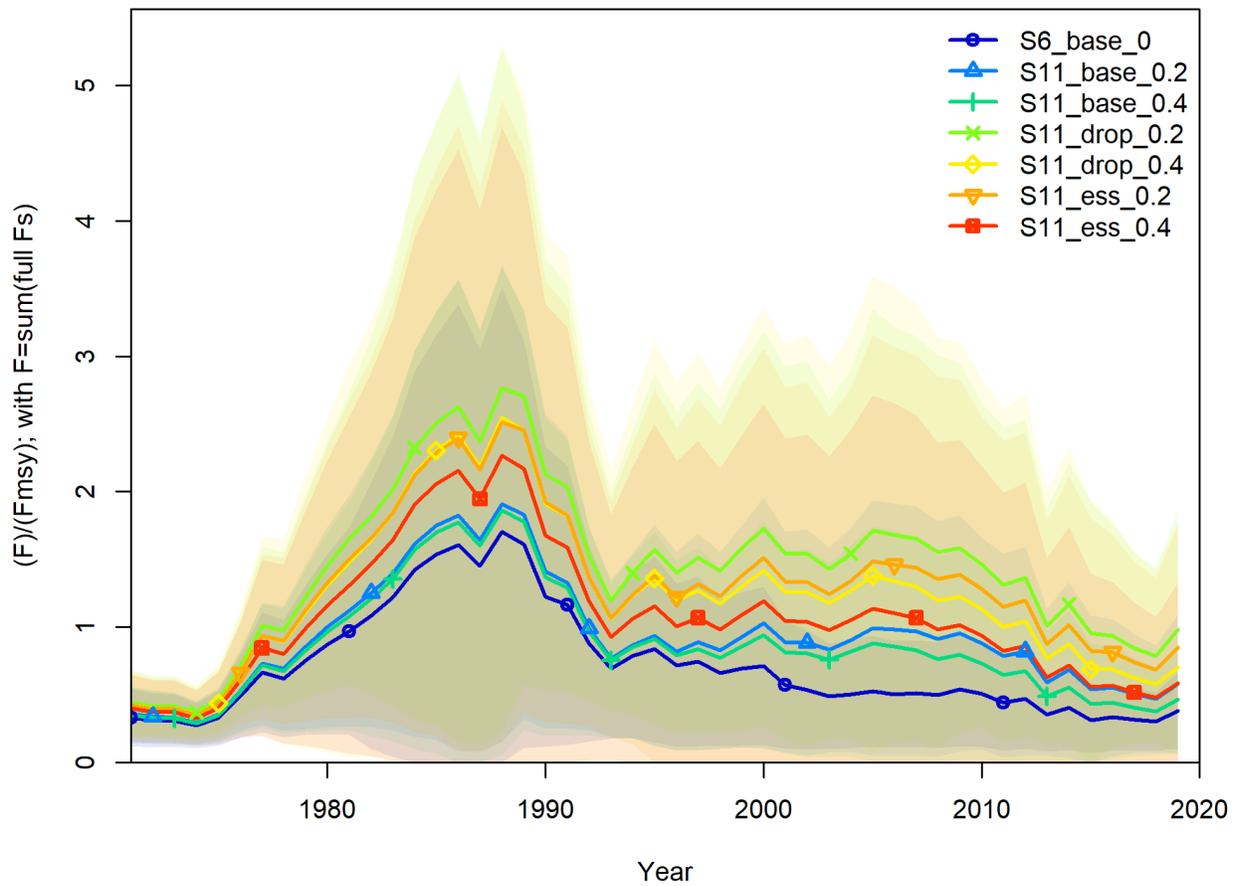


Figure 53: Fishing mortality (with uncertainty) for each of the six S11 candidate models with the S6 model shown for comparison.

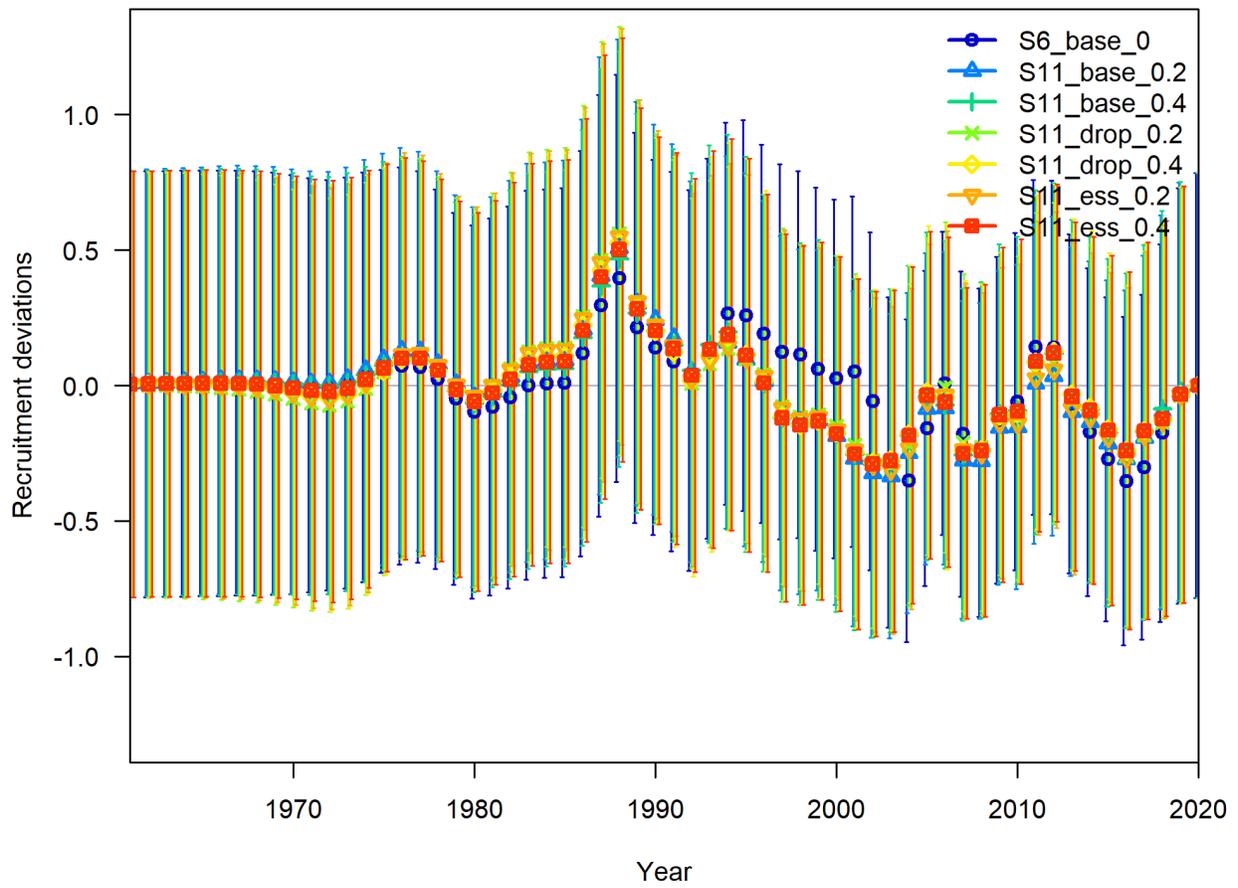


Figure 54: Recruitment deviates (with uncertainty) for each of the six S11 candidate models with the S6 model shown for comparison.

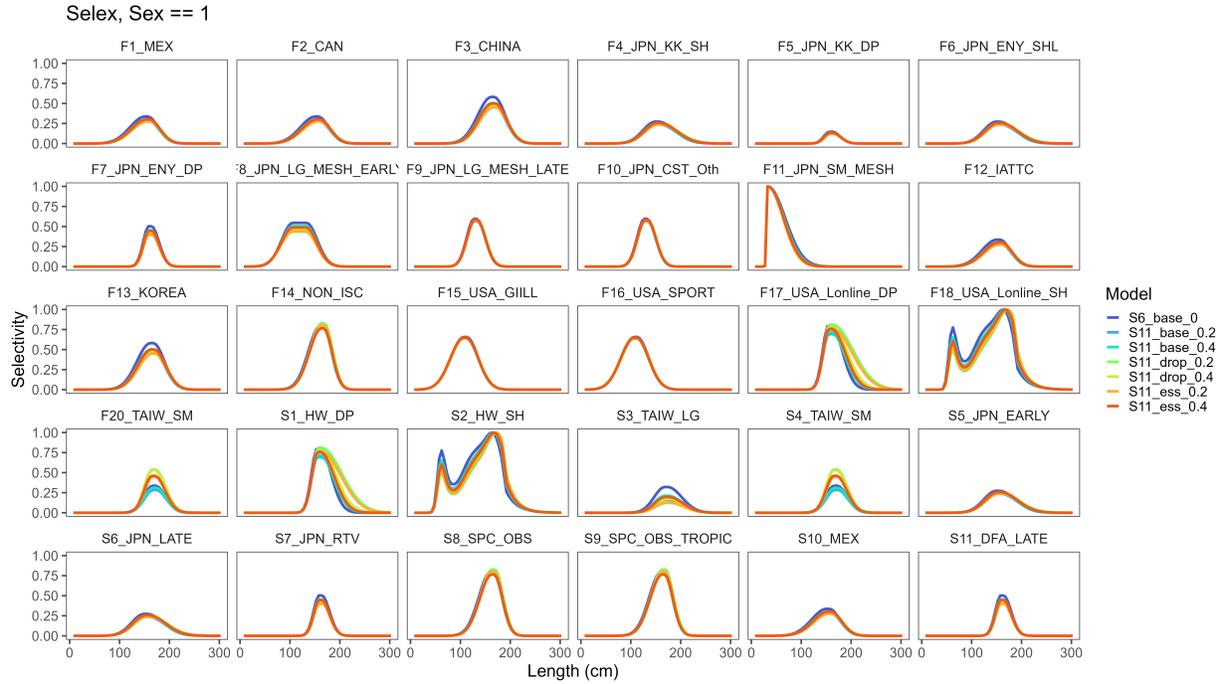


Figure 55: Time-varying selectivity for sex 1 (females) for non-*F19_TAIW_LG* fleets by candidate model.

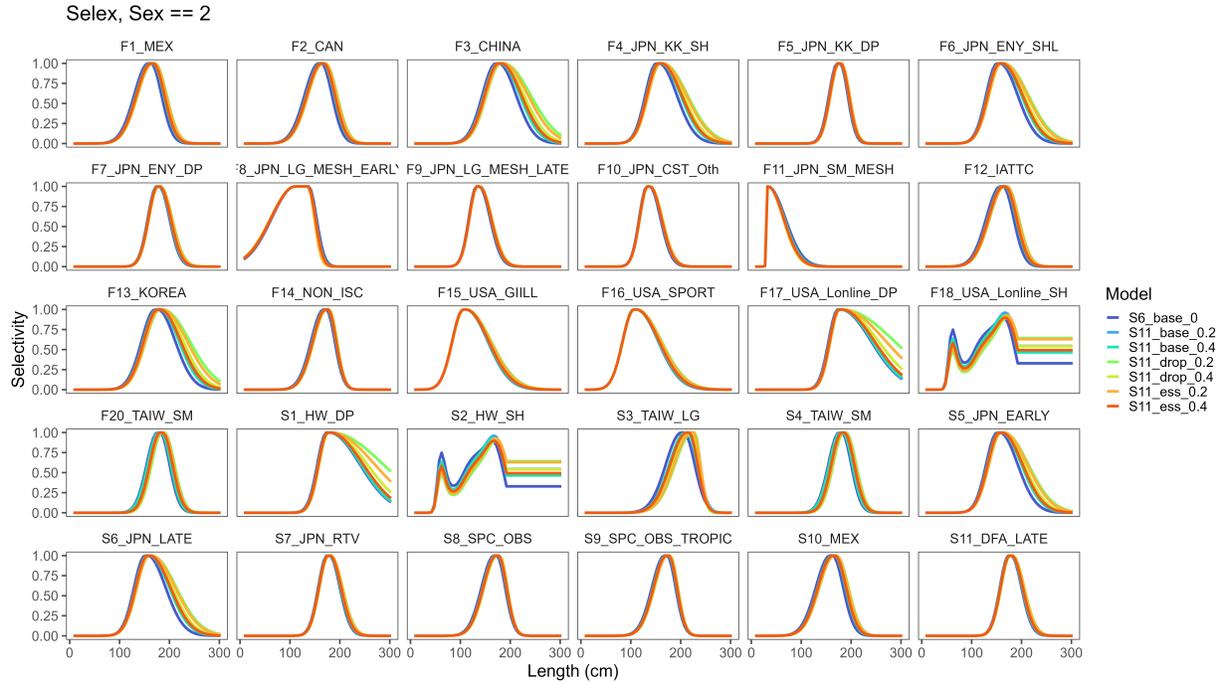


Figure 56: Time-varying selectivity for sex 2 (males) for non-*F19_TAIW_LG* fleets by candidate model.