

A preliminary base-case model in SS3.30 for the 2023 North Pacific Swordfish Assessment

Michelle L. Sculley,* Yi-Jay Chang,+ Hirotaka Ijima^

*NOAA NMFS Pacific Islands Fisheries Science Center, Honolulu, HI

michelle.sculley@noaa.gov

+National Taiwan University, Taipei, Taiwan

^Fisheries Research Agency, Yokohama, Japan

Abstract

A preliminary base-case model in Stock Synthesis 3.30 for North Pacific (NP) swordfish (*Xiphias gladius*) is described for consideration as the 2023 base-case model. The base-case model covers the Western and Central North Pacific north of the Equator and the Eastern Pacific Ocean north of 10°N from 1975 to 2021. It includes data from three International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) countries and other countries in aggregate from the Western and Central Pacific Fisheries Commission (WCPFC) and the Inter-American Tropical Tuna Commission (IATTC). This paper describes the data available for inclusion and the proposed base-case model. The model converges and appears to fit the data well. Initial diagnostics do not indicate major problems. Preliminary results suggest the North Pacific swordfish stock is being not fished above F_{MSY} and spawning stock biomass is above SSB_{MSY} .

Introduction

The International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) Billfish Working Group (BILLWG) has proposed to run a benchmark assessment on North Pacific swordfish (*Xiphias gladius*, SWO). Data were compiled from the International Scientific Committee for North Pacific Tuna and Tuna-like Species (ISC) member countries and other Western and Central Pacific Fisheries Commission (WCPFC) and Inter-American Tropical Tuna Commission (IATTC) countries. Countries were asked to contribute catch, CPUE, and size-frequency data. It was decided to run the assessment using a sex-specific model in Stock Synthesis version 3.30.20 (Methot and Wetzel, 2013) with fleets-as-areas separating the Eastern Pacific Ocean (EPO) region from the Western and Central North Pacific (WCNPO) region, as defined in the December 2022 BILLWG meeting. The available data and the preliminary model results and diagnostics for a base-case model will be presented in this document for consideration at the ISC BILLWG SWO stock assessment meeting.

Methods

Spatial Temporal Structure

Data were compiled by region assuming a two region model of the North Pacific Ocean with boundaries based upon those detailed in Ichinokawa and Brodziak (2008) with the modification that the Eastern Pacific Ocean (EPO) region ends at 10°N (Figure 1). Countries were asked to contribute catch, CPUE, and length frequency data partitioned by these two regions so that the North Pacific could be modeled with two implicit areas using fleets as areas. The working group agreed to start model in 1975.

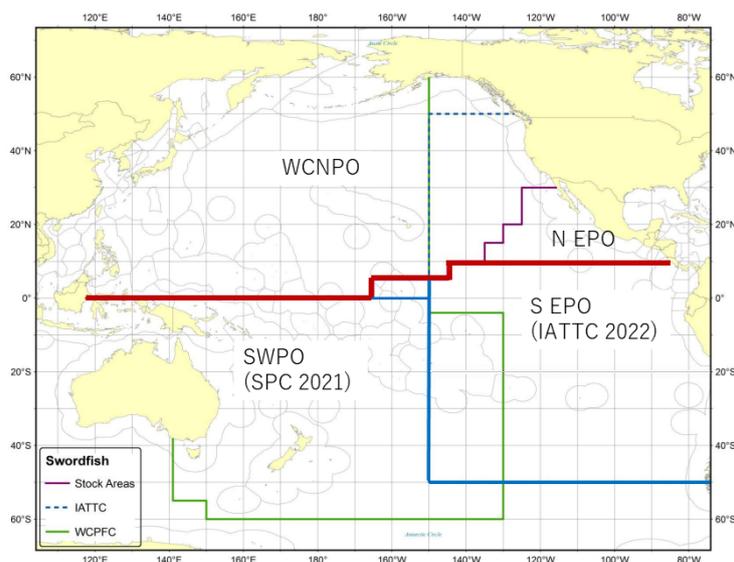


Figure 1: Western and Central North Pacific Ocean and North Eastern Pacific Ocean swordfish stock boundaries for the 2023 North Pacific swordfish assessment. Spatial structure is treated implicitly using fleets as areas.

Definition of Fisheries

Data are available for fourteen different fleets in the WCNPO: 14 catch time series, with some fleets split into early and late periods and 8 CPUE indices of which one is a recruitment index. The fleet names and numbers are detailed in Table 1. The data available for each fleet is in Figure 2. The acronyms in the fleet names are defined as follows: WCNPO is Western and Central North Pacific Ocean; EPO is Eastern Pacific Ocean; OSDWLL is offshore distant water longline; OSDWCOLL is offshore distant water and coastal longline; early is the early time period; late is the late time period, Area1 and 2 are the Japanese fishery areas in the WCNPO as defined in Ijima 2022; OSDF is offshore driftnet gear; CODF is coastal driftnet gear, JPN_WCNPO_Other is Japanese small-scale coastal longline vessels which are not under obligation to submit logbook data, bait, and net fishing gear; DWLL is distant water longline gear, TWN_WCNPO_Other is Taiwanese offshore longline, coastal longline, gillnet, harpoon and other gears; LL is longline gear; shallow is the Hawaii shallow-set sector; deep is the Hawaii deep-set sector; GN is gillnet gear; US_WCNPO_Other is harpoon and other

gears; WCPFC is other WCPFC and IATTC longline gear in the WCNPO; IATTC is longline gear in the EPO north of 10°N.

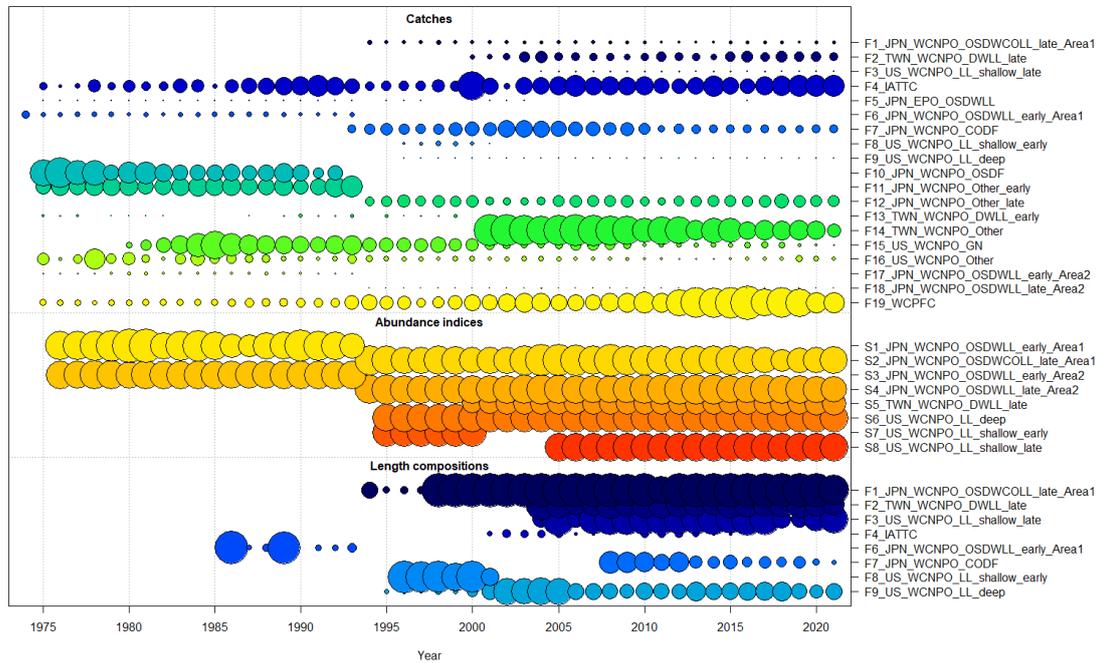


Figure 2: Catch, CPUE index, and size composition data included in the 2023 NP swordfish stock assessment. The size of the bubble indicates the relative number of observations available.

Catch

Some of the time series of catch were divided into early and late periods to coincide with divisions of the CPUE indices (Table 1, Figure 3). Three ISC countries contributed catch time series: Japan, Taiwan, and the US. In addition, catch from countries reporting to the WCPFC and IATTC were obtained from each RFMO, respectively. The CV for catch was set to 0.05 for all fleets. Catch for fleets with only annual data were divided equally into each quarter.

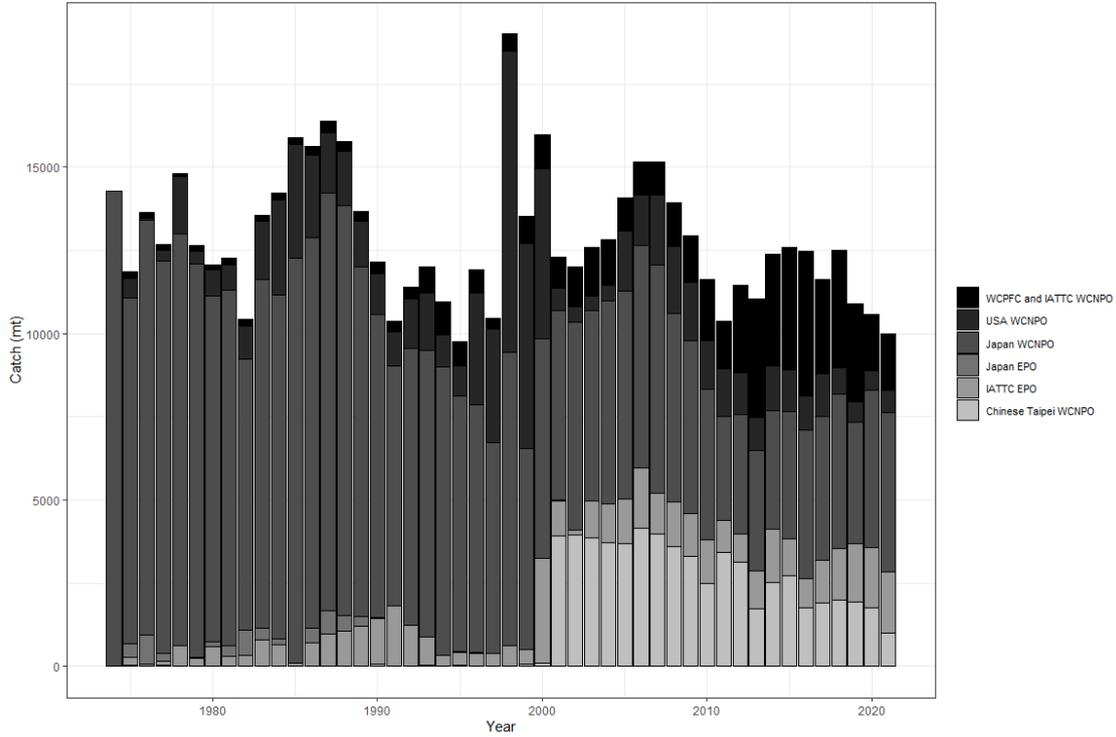


Figure 3: Annual catch of NP swordfish by country or commission and area.

Relative Abundance Indices

Each of the eight CPUE time series were assigned to a quarter based upon the recommendations of the country providing the index and are assumed to represent the quarter in which the highest catches take place for each fishery. Japanese longline fleets (S1-4) were all assigned to quarter 1; Taiwanese longline fleet (S5) was assigned to quarter 3; US longline deep-set (S7) was assigned to quarter 2, and US longline shallow-set (S8 and S9) were assigned to quarter 1. US longline deep-set fleet S6 was included as an index of recruitment because the fishery catches large numbers of young-of-the-year fish (Fleet type 31, Bohaboy and Sculley, 2023). The CPUE indices were assumed to be linearly proportional to biomass where catchability (q) was assumed to be constant and occur in the first month of the quarter assigned.

The CVs for each CPUE index were assumed to be equal to their respective calculated SEs on the log scale. The minimum CV was scaled to a minimum of 0.20 or the root-mean-square error (RMSE) (i.e., square root of the residual variance) of what we would expect the assessment model to fit the CPUE index best by adding a constant to each CV value. This was calculated as the square root of the residual variance of a loess smoother fit to each index (Francis 2011, Lee et al., 2014).

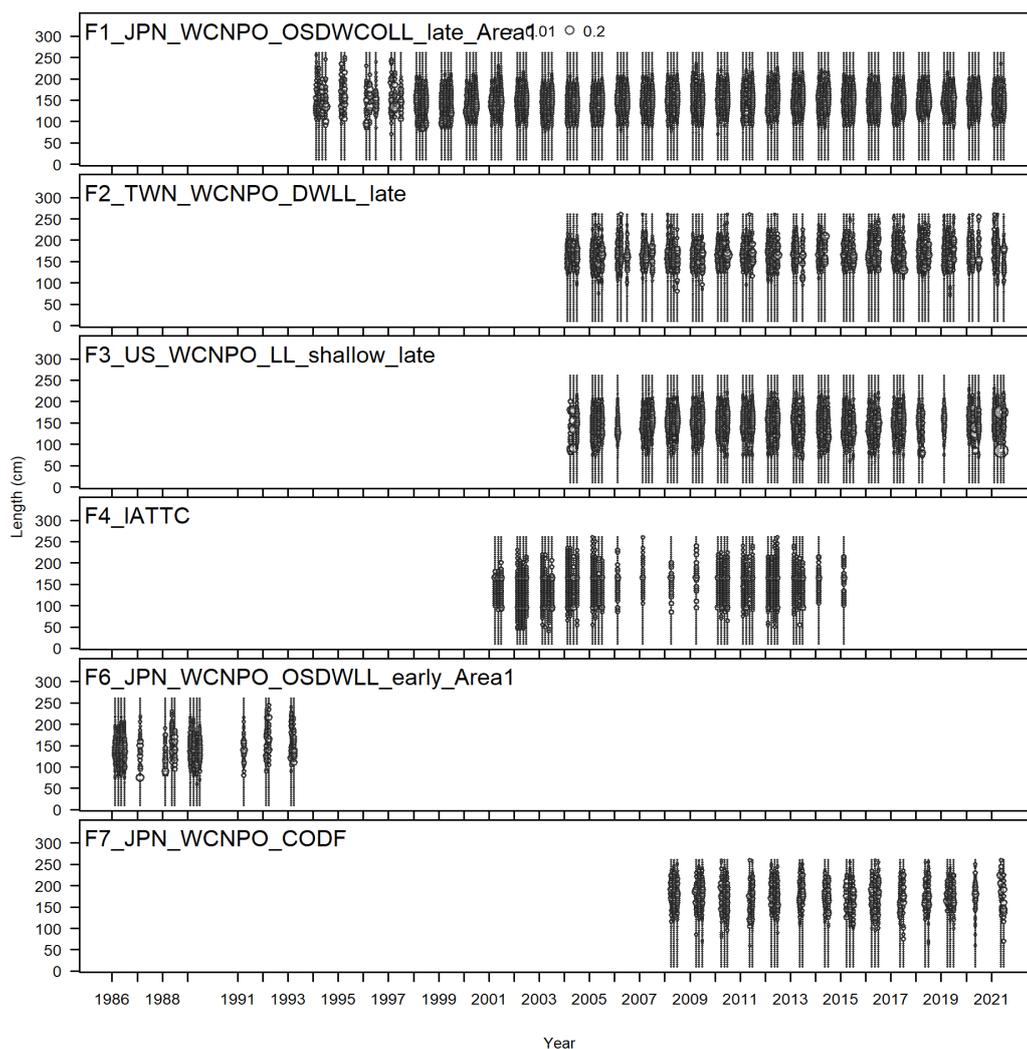
$$RSMSE_{smoother} = \sqrt{(1/N) \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}$$

where Y_t is the observed CPUE in year t on the log scale, \hat{Y}_t is the predicted CPUE in year t from the smoother fit to the data on the log scale, and N is the number of CPUE

observations. RMSE values for each index are listed in Table 2. If the input SE was greater than these values, it was left unchanged.

Length Composition

Length composition data were available for six WCNPO fleets and two EPO fleets (Figure 4). Length composition data were available in quarterly time steps. Quarters with fewer than 15 total samples were removed from the time series due to limited sample size, as agreed upon by the modeling sub-group. In addition, the length composition data for F5 were excluded as they were sparse. Data were fit using a multinomial error structure. Length composition data were weighted based upon the Francis (2011) method. Input effective sample size was estimated as the total number of fish measured in each year-quarter divided by 10, with any year-quarter over fifty set to fifty to reduce the influence of large sampling events. Then, weighting was attempted based upon the T.A1.8 equation (Francis 2011) as calculated by the model using r4ss, an R package for plotting SS results (R version 3.4.0, R Core Team, 2017, r4ss version 1.28.0, Taylor et al., 2017). Length composition data were only down-weighted if the model suggested it, otherwise the fleets were not adjusted. The only fleet to be down-weighted using the Francis methods was F9, the US deep-set LL data.



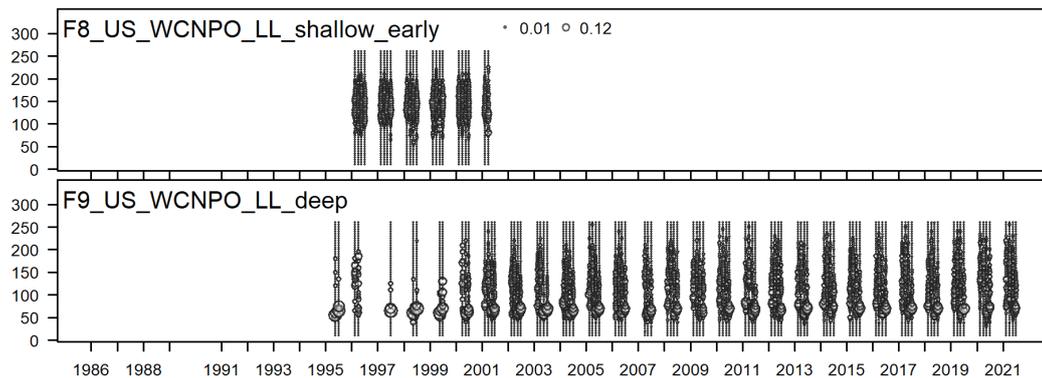


Figure 4: Length Composition data available in 5cm size bins for the 2023 NP swordfish stock assessment.

Initial Base-case Model Description

The assessment was conducted with Stock Synthesis (SS) version 3.30.20.00 released 09/30/2022 using Otter Research ADMB 13.0 (Methot and Wetzel 2013). The model was set up as a single area model with two sexes and four seasons (quarters) with fleets as areas. Spawning was assumed to occur in May (month 5) while recruitment was assumed to occur in July (month 7). Age at recruitment was calculated based upon the model estimated average selectivity at age based upon the quarterly selectivity at length. The maximum age of swordfish was set to 15 years. Sex specific biological parameters were used, with sex- and age-specific natural mortality (Table 3) as agreed upon in the BILLWG data preparatory meeting (ISC BILLWG, 2023). In addition, the CV of the growth curve was set to 0.1 for small males and females and 0.15 for large males and females, and the sex ratio at birth was assumed to be 1:1. The model used a Beverton-Holt spawner-recruit relationship with steepness (h) fixed at 0.9 and σ_r initially fixed at 0.6 and rescaled to 0.42 based upon the model suggestion.

Twenty-seven fleets were included in the model: 19 catch fleets and 8 survey fleets. Initial equilibrium catch was estimated for 1974.

Main recruitment deviations were estimated from 1985-2021, as this is the period during which there are size composition data. The recruitment deviations were bias-adjusted based upon the estimates from Methot and Taylor (2011) provided from the model results. No bias adjustment was applied to recruitment deviations from 1965-1967. 1967-1995 was the “ramp-up” period where the bias adjustment of σ_r was 0 at the beginning of the period and increased linearly to the maximum bias adjustment 0.94 in 1995. Full bias adjustment was from 1995-2021. The early period of recruitment deviations represents a data-poor period where there is little information to drive recruitment. The main recruitment period represents a data-rich period where there is enough data to drive the bias-adjustment of the recruitments. The ramp up period allows for a gradual ramp up of the bias-adjustment between the data-poor and data-rich periods.

The population model and the fishery length data had 51 five cm length bins from 10-260+ cm. The population had 16 annual ages from age 0 to 15+. There were no age data.

Fishery length data were used to estimate selectivity patterns which controlled the size distribution of the fishery removals. All fleets with length data were estimated as six parameter double normal (dome-shaped) selectivity patterns. Survey selectivity patterns mirrored their respective catch fleets (Table 4). Estimating parameters six of the double normal selectivity pattern for F9 resulted in an improved fit to the size composition data for that fleet.

Model estimated time series of total biomass (B in metric tons, $mt = 1000 \text{ kg}$), age 1+ total biomass (B_{1+} mt), female spawning biomass (SSB mt) and recruitment (R in 1000s of fish) were tabulated on an annual basis. Annual exploitation rate (F) was calculated as $Catch/B_{1+}$. Stock status indicators were calculated based upon MSY -based reference points as proxies, given that the WCPFC has not set biological or other reference points for swordfish.

Convergence Criteria and Diagnostics

The model was assumed to have converged if the standard error of the estimated parameters could be derived from the inverse of the negative hessian matrix. Various convergence diagnostics were also evaluated. Excessive CVs ($>50\%$) on estimated parameters would suggest uncertainty in the parameter estimates or model structure. A gradient of >0.001 would suggest poorly fit parameter estimates. The correlation matrix was also evaluated to identify highly correlated ($>95\%$) and non-informative (<0.01) parameters. Parameter estimates hitting bounds of the prior was also indicative of poor model fit.

Several diagnostics were run to evaluate the fit of the model to the data. An Age-Structure Population Model (ASPM) was used to evaluate the influence of the length composition data on the population trends (Carvalho *et al.*, 2017). The ASPM was also used to explore how each CPUE index informed the population trends by running one-off ASPMs for each index. Profiling the likelihood on R_0 , where the R_0 is fixed at a range of values around the maximum likelihood estimate and then the likelihood is estimated, was used to identify influential data components (Lee *et al.*, 2014). A runs test was used to evaluate randomness in the residuals of the CPUE data (Carvalho *et al.*, 2021). Residual plots and plots of the observed vs expected data were examined to evaluate goodness-of-fit. Finally, a retrospective analysis and hindcast cross-validation were used to evaluate the predictive ability of the model (Carvalho *et al.*, 2021).

Results

Model fit

The base-case model ran in about 10 minutes, estimated 89 parameters, and had a total likelihood of 1338.75 The inverse Hessian was positive definite, which allowed for the estimation of parameter standard deviations and suggests that the model converged, and the maximum gradient component was less than 0.001. None of the parameter estimates hit a bound, two selectivity parameters had a correlation above 0.95 and no parameters had correlations below 0.01. All twenty early recruitment deviations (1965-1984) and 30 of the 37 main recruitment deviations had CVs $> 50\%$. Five of the 34

selectivity parameters had CVs >50%, most of which were for the estimated width of the peak (i.e. a very small number).

Fits to the abundance indices were relatively good (Figure 5 - Figure 13). The expected CPUE trend for S5 TWN LL was relatively flat and did not fit the data well, and the observed CPUE in the last 3-5 years for S4 JPN LL area 1 declined but the expected CPUE increased. Three indices failed the runs test, S5, S6, and S8 (Figure 11). S6 is the US deep set LL fleet which is a recruitment index and not fitted as a survey index, so its fit can be ignored. S8 is the US shallow-set LL fleet in the late time period, and indicates a long term decline. Additional investigation into the CPUE indices indicate that there does appear to be some conflict between the trends for each fleet, which could be driving some of the mis-fit within the model.

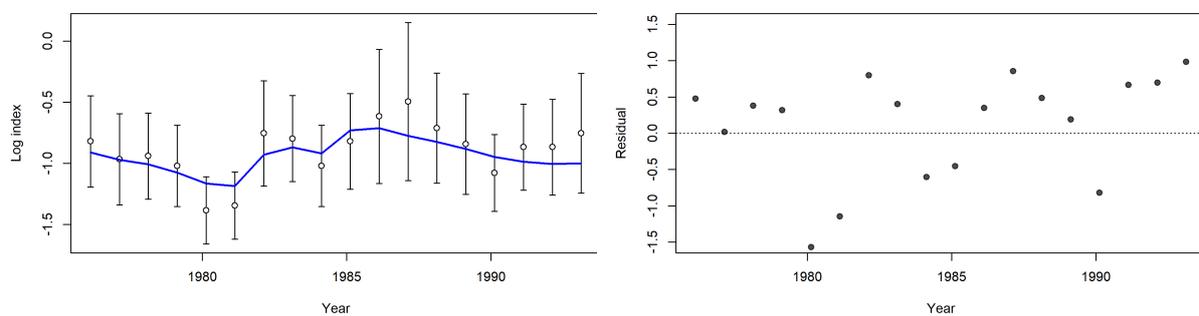


Figure 5: Fit to the S1 Japanese early area 1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

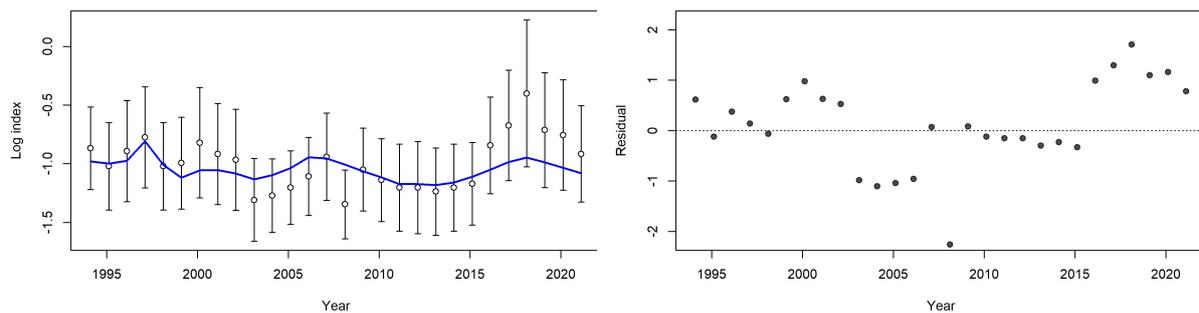


Figure 6: Fit to the S2 Japanese late area 1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

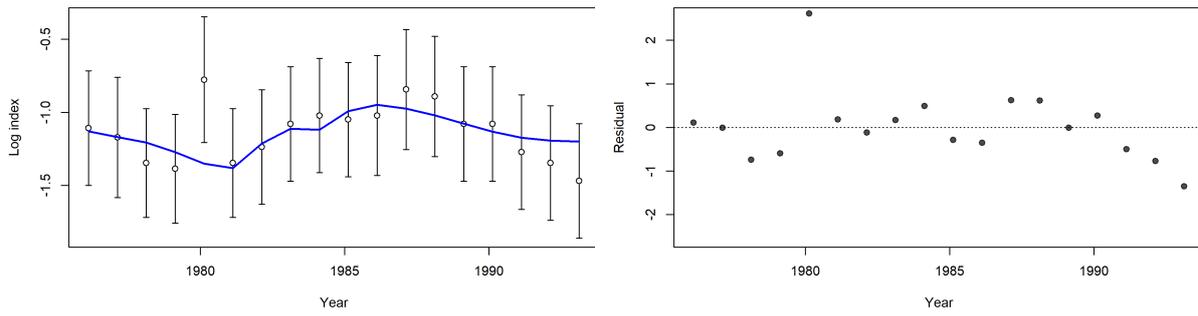


Figure 7: Fit to the S3 Japanese early area 2 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

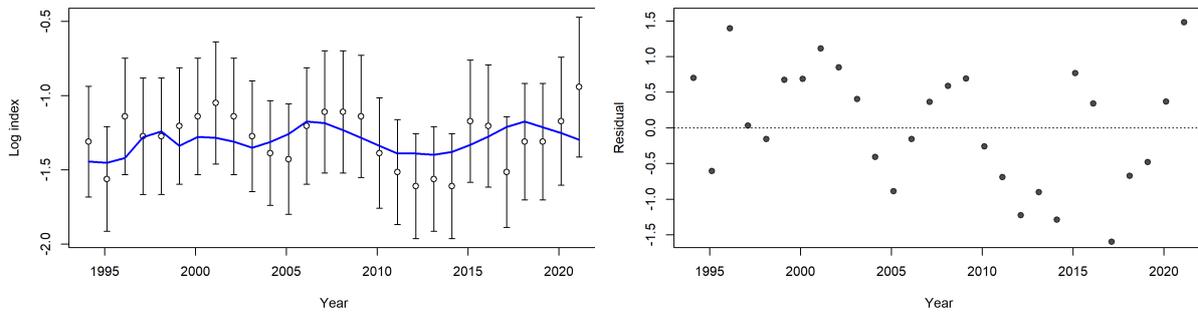


Figure 8: Fit to the S4 Japanese late area 2 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

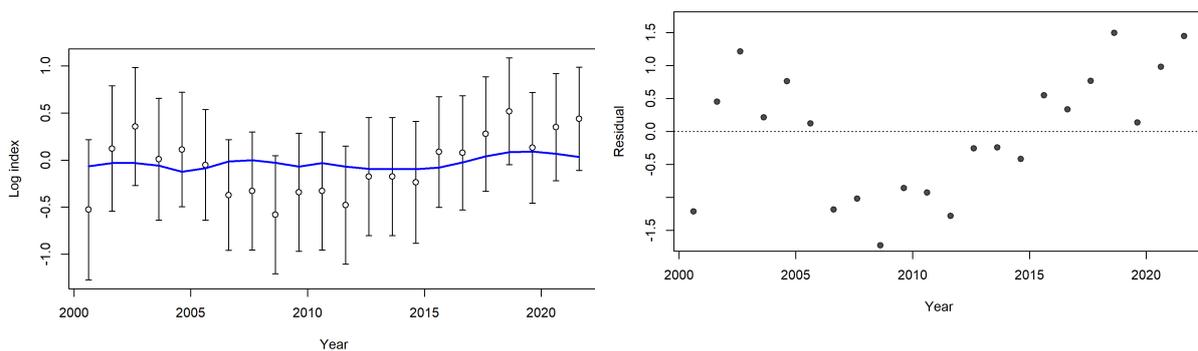


Figure 9: Fit to the S5 Chinese Taipei late LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

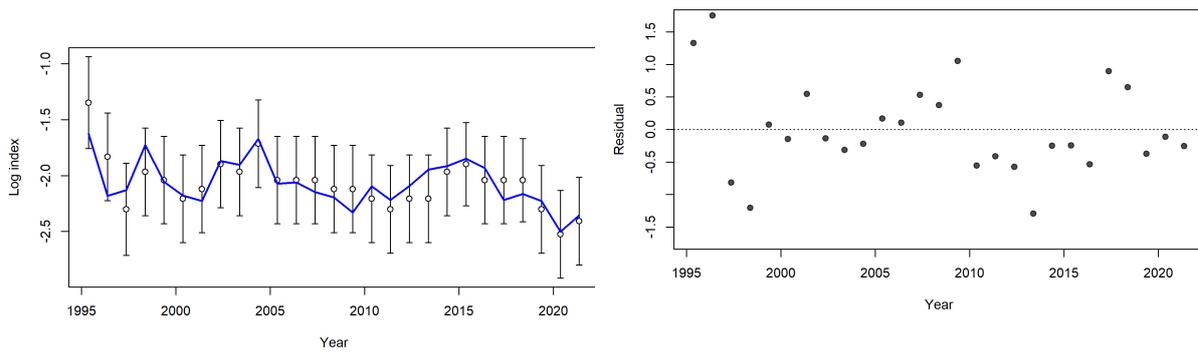


Figure 10: Fit to the S6 US deep-set LL CPUE recruitment index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

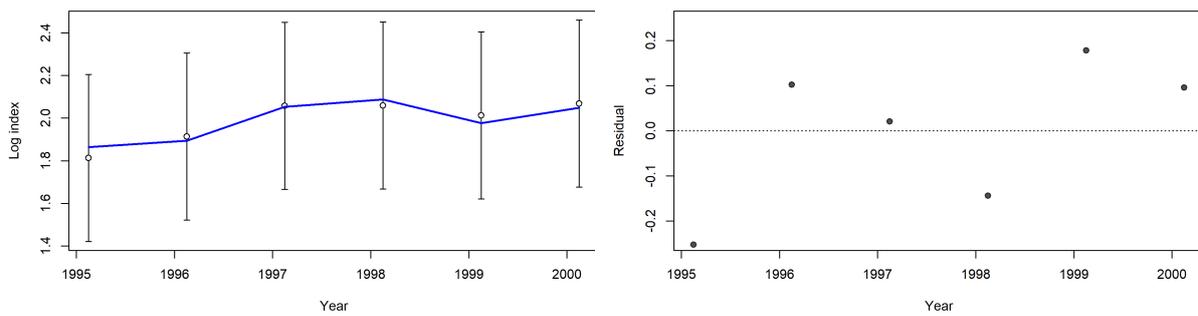


Figure 11: Fit to the S7 US shallow-set LL early CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

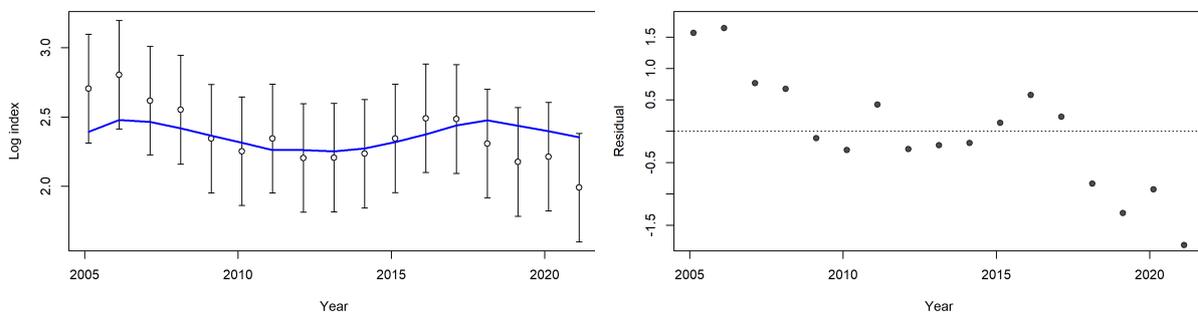


Figure 12: Fit to the S8 US shallow-set LL early CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

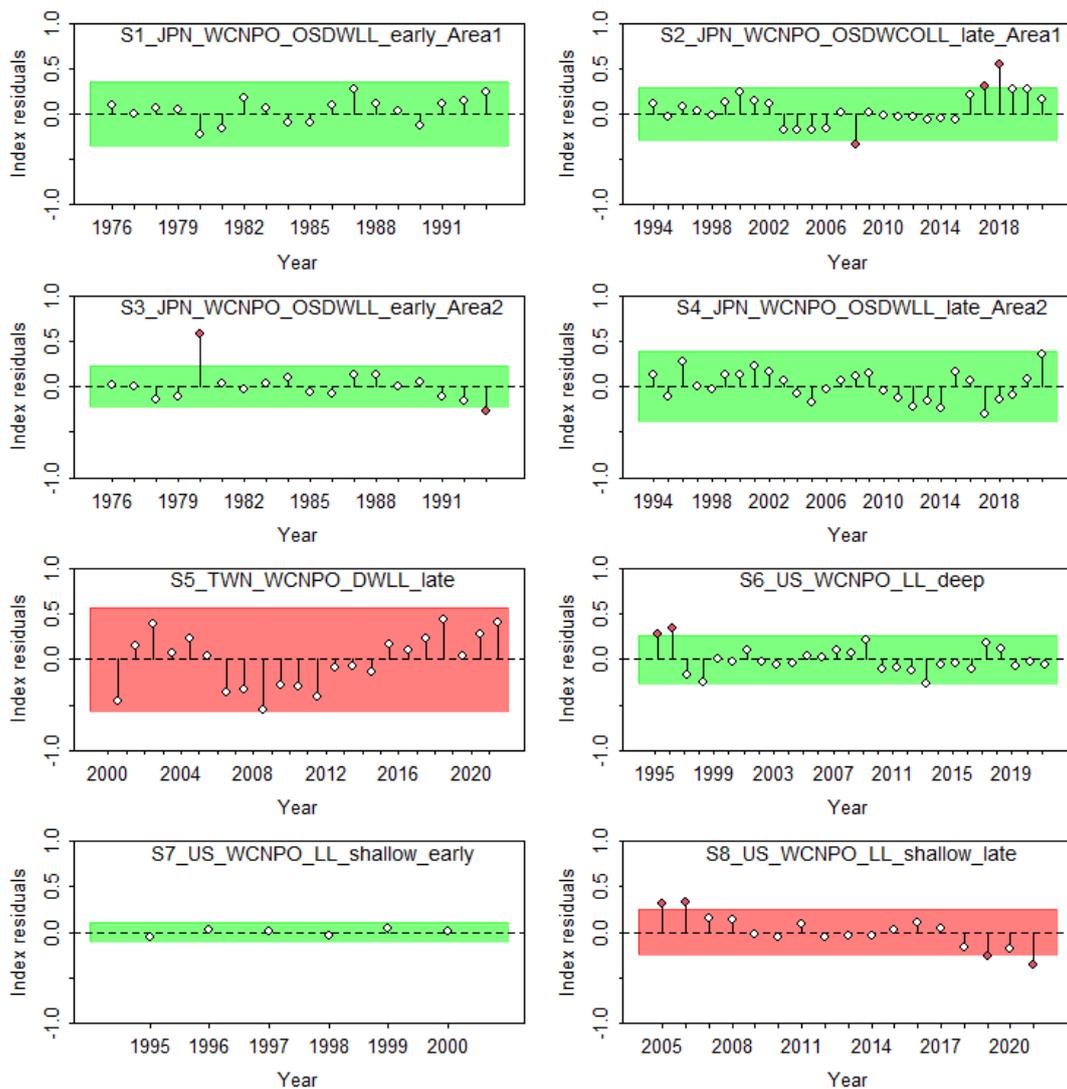


Figure 13: Results from a runs test for each CPUE index. Red indicates the index failed the test (residuals are not random), green indicates the index passed the test.

Estimated selectivity for each fleet are in (Figure 14 - Figure 17). Fits to the length composition data were also relatively good (Figure 18 - Figure 21), although there are still problems fitting the US deep-set longline data (F9). The fit to the US size data is challenging because size distribution changes substantially seasonally, with a sharp peak of small fish entering the fishery in quarters 3 and 4 which are not observed in such large numbers in quarters 1 and 2. Attempts to separate the fleet into a quarterly fleets to estimate selectivity are ongoing but not yet successful. In addition, F1 JPN LL area 1 late and F4 IATTC both fail the runs test (Figure 22).

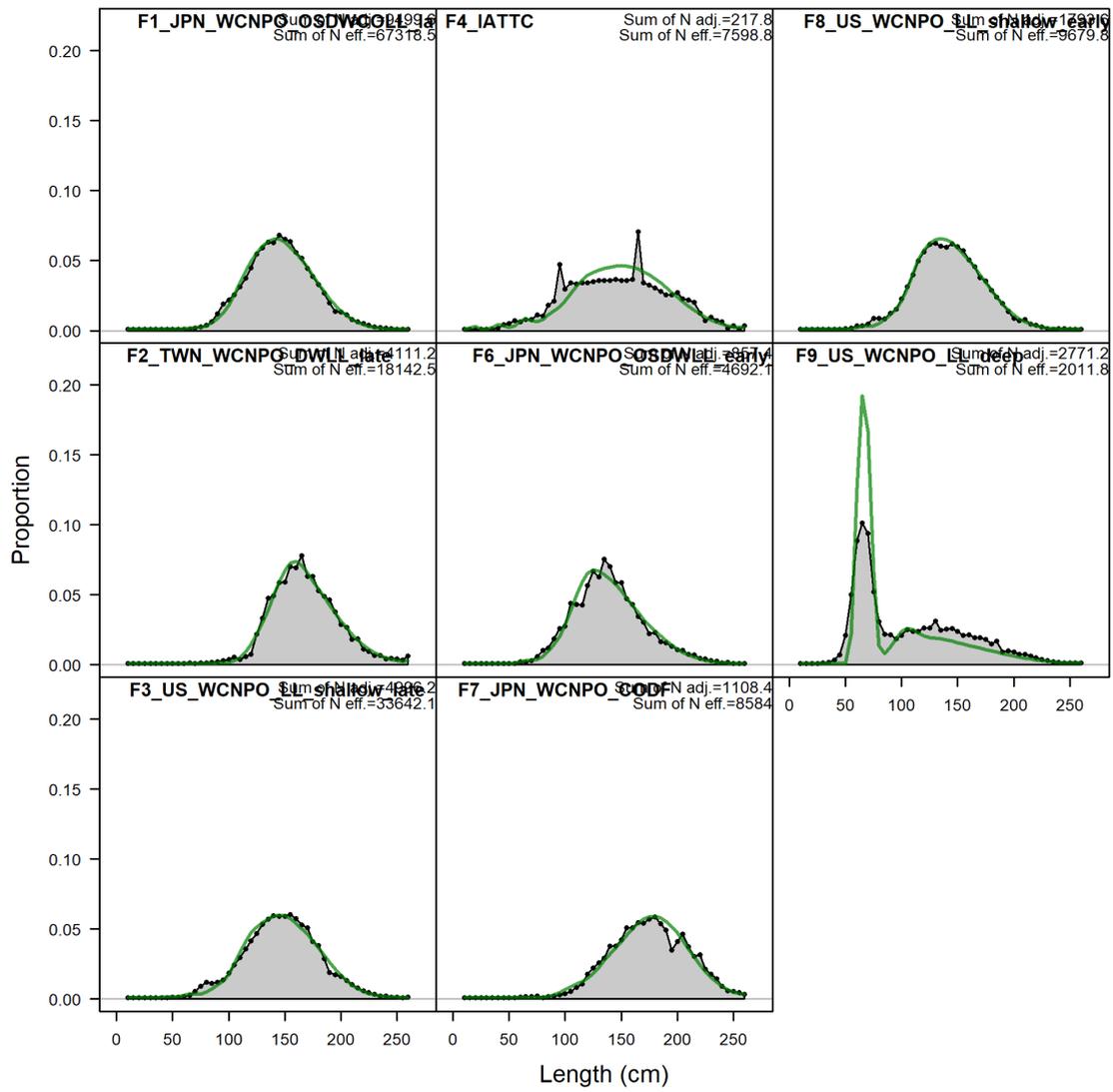


Figure 14: Aggregated Size comp data (grey) and model fit (green)

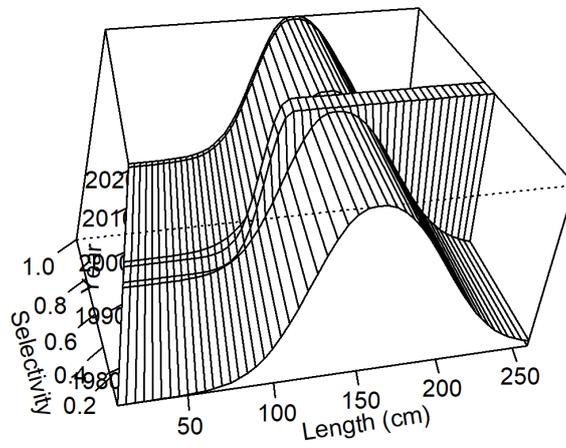


Figure 15: Time-varying selectivity estimated for F01 Japan LL area 1 late.

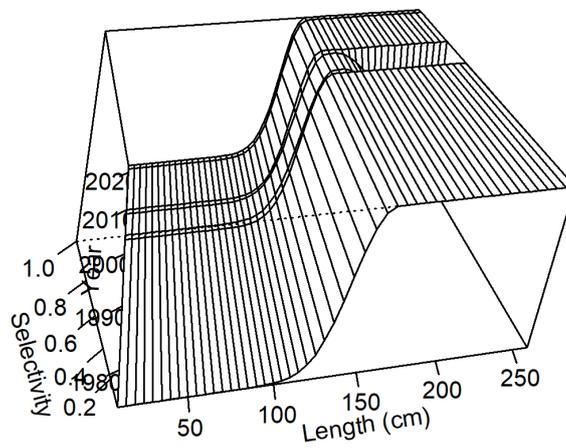
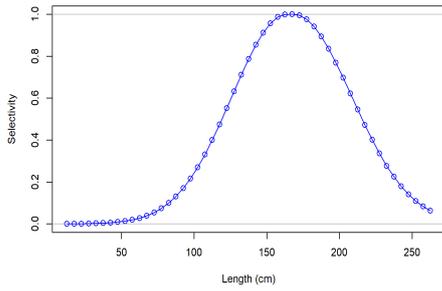
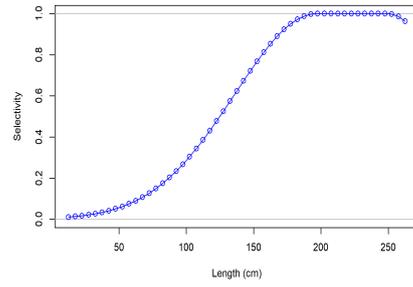


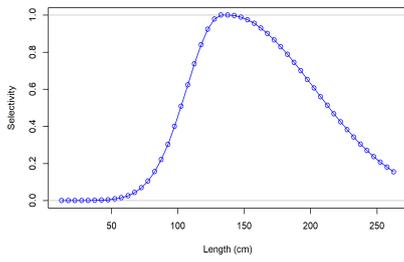
Figure 16: Time-varying selectivity estimated for F02 Chinese Taipei LL late.



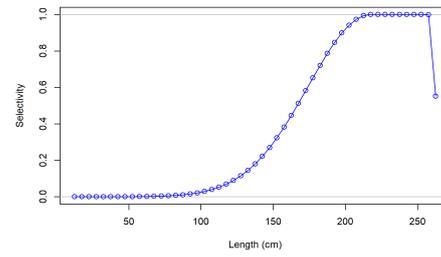
F03_US_WCNPO_LL_shallow_late



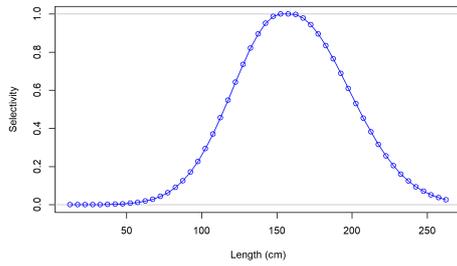
F04_IATTC



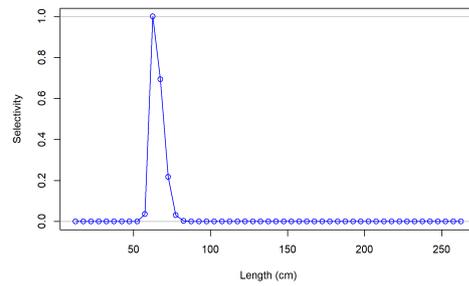
F06_JPN_WCNPO OSDWLL_early_Area1



F07_JPN_WCNPO_CODF



F08_US_WCNPO_LL_shallow_early



F09_US_WCNPO_LL_deep

Figure 17: Selectivity estimates for each of the 6 fleets without time-varying parameters.

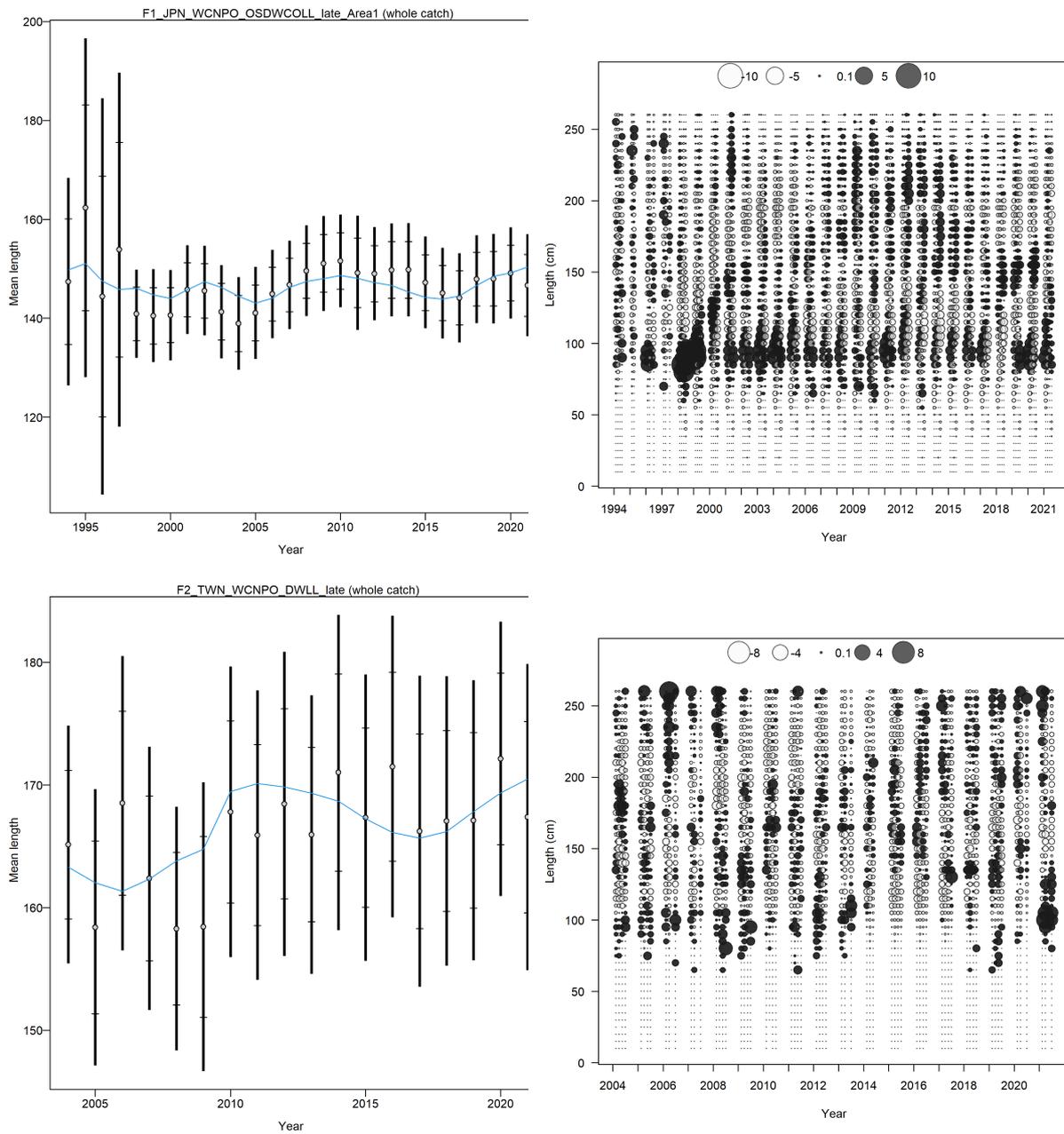


Figure 18: Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan LL Area 1 late (top) and Chinese Taipei LL late (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

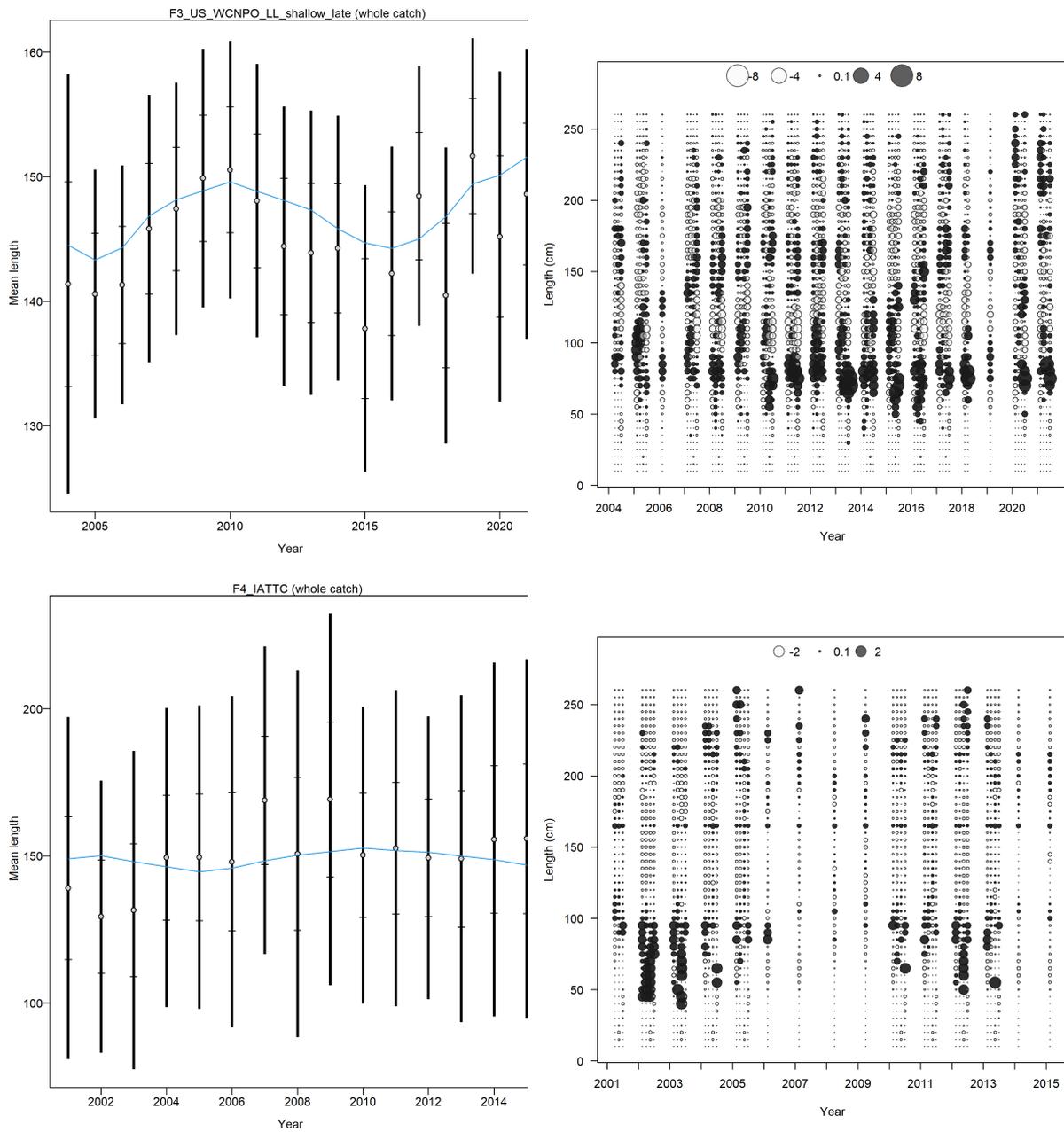


Figure 19: Fits to the annual mean length (left panels) and quarterly residuals (right panels) for US shallow-set LL late (top) and IATTC EPO (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

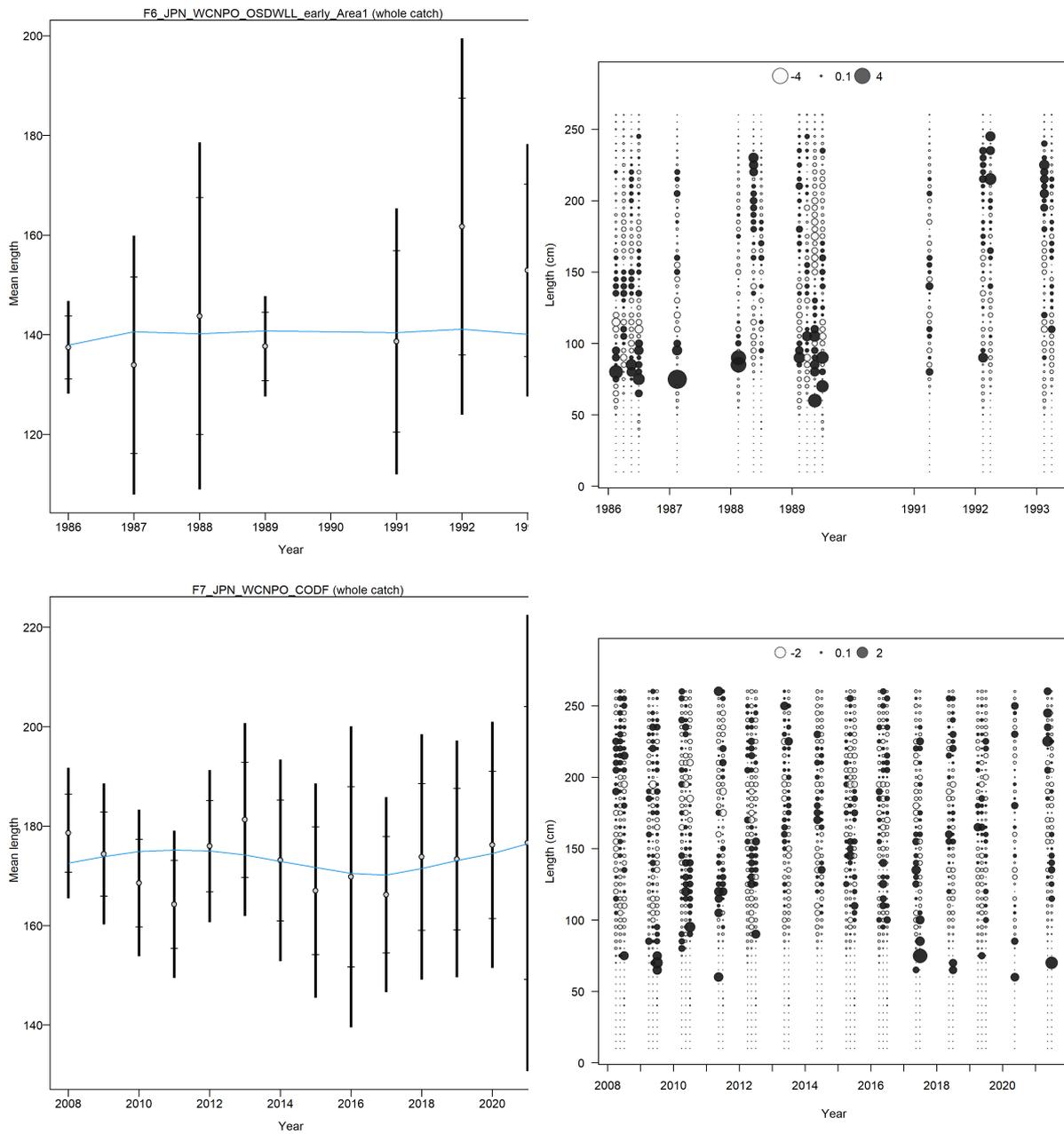


Figure 20: Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan LL Area 1 early (top) and Japan coastal driftnet (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

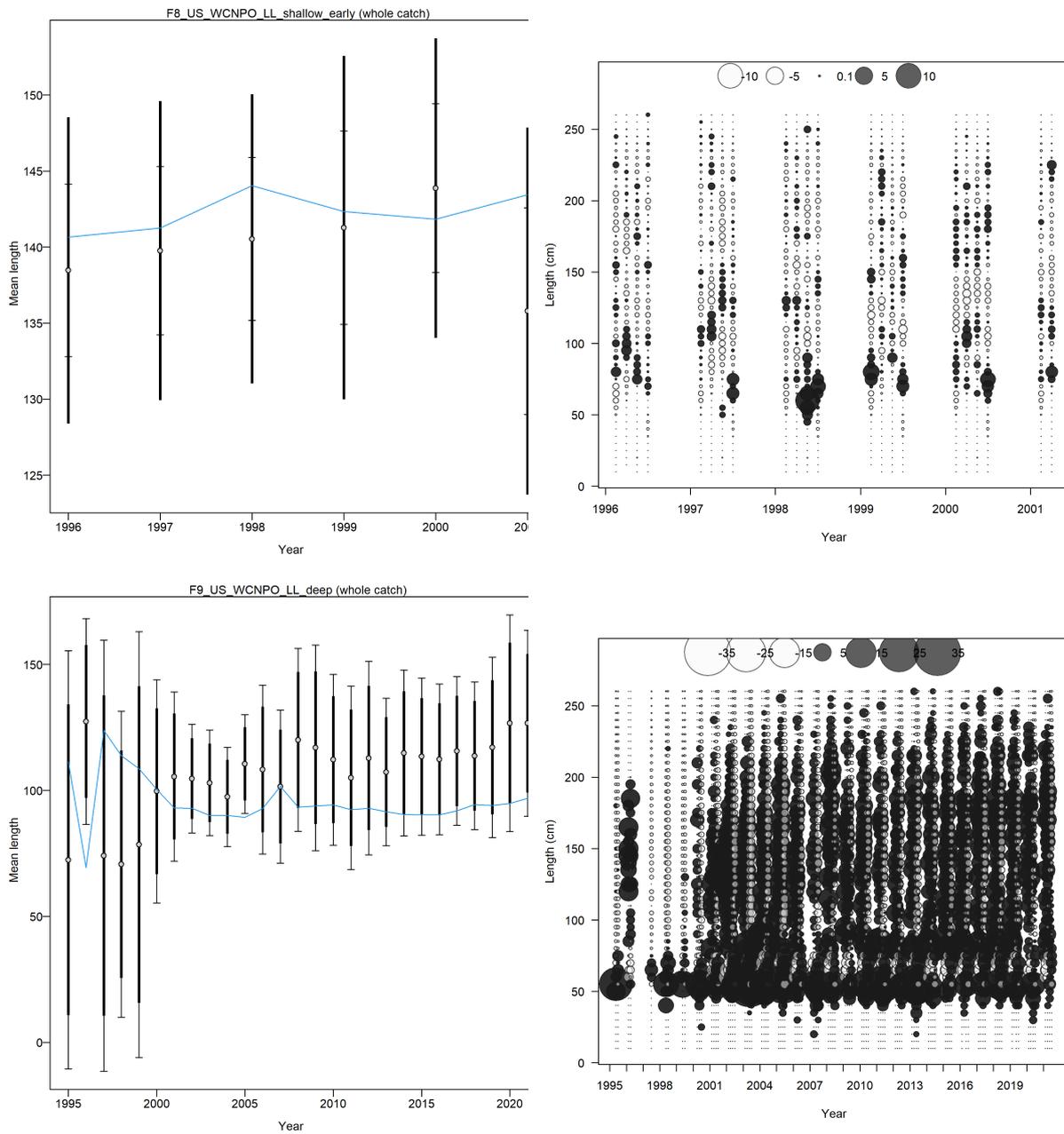


Figure 21: Fits to the annual mean length (left panels) and quarterly residuals (right panels) for US shallow-set LL early (top) and US deep-set LL (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

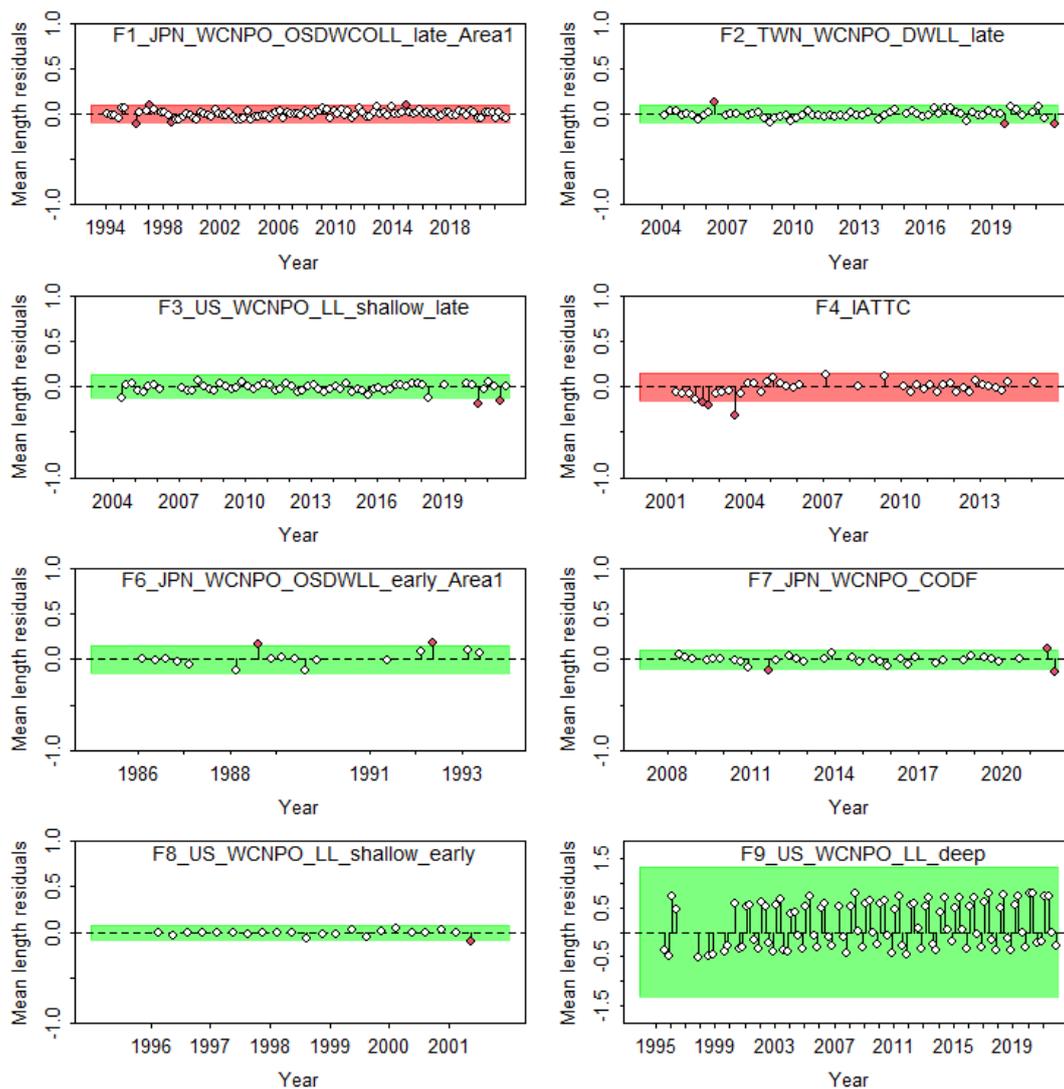


Figure 22: Results from a runs test for each size composition time series. Red indicates the time series failed the test (residuals are not random), green indicates the time series passed the test.

Model estimates of age 1+ biomass show a general increase in biomass from through the entire assessment time horizon, with some periodic dips in abundance (Figure 23). Initial spawning stock biomass was estimated to be approximately 32,000 mt and virgin SSB was around 112,000 mt (Figure 24). Annual fishing mortality is reported as the average for fish ages 1-10 (Figure 25). Fishing mortality was below MSY for all years and has been decreasing generally since 1998. Recruitment deviations relatively random around the equilibrium value, with occasional strong year classes (Figure 26). Current depletion, as estimated as the age 1+ biomass in 2020 compared to the virgin age 1+ biomass was estimated to be 0.516213.

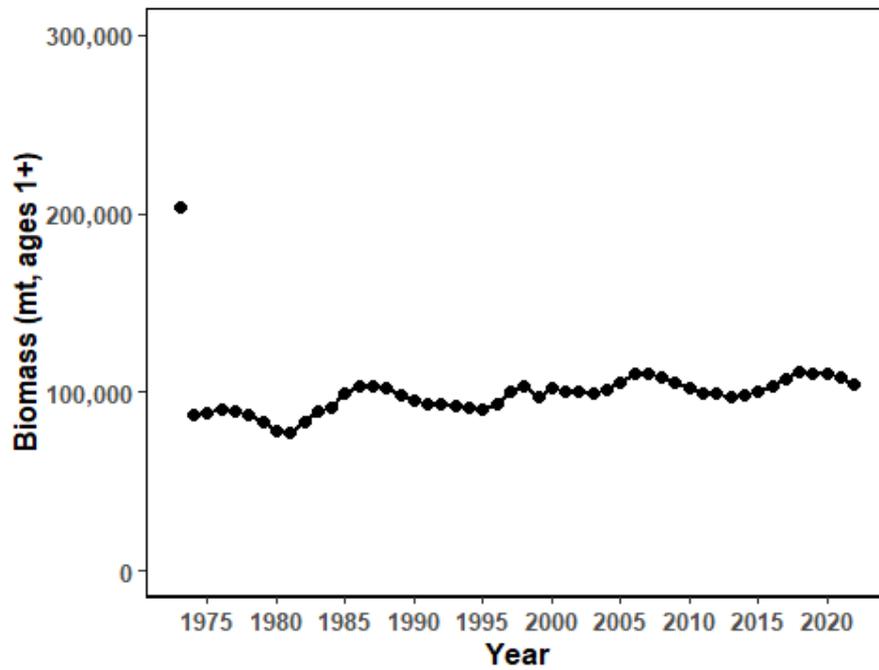


Figure 23: Estimated biomass (mt) of NP swordfish 1+ from the base-case model.

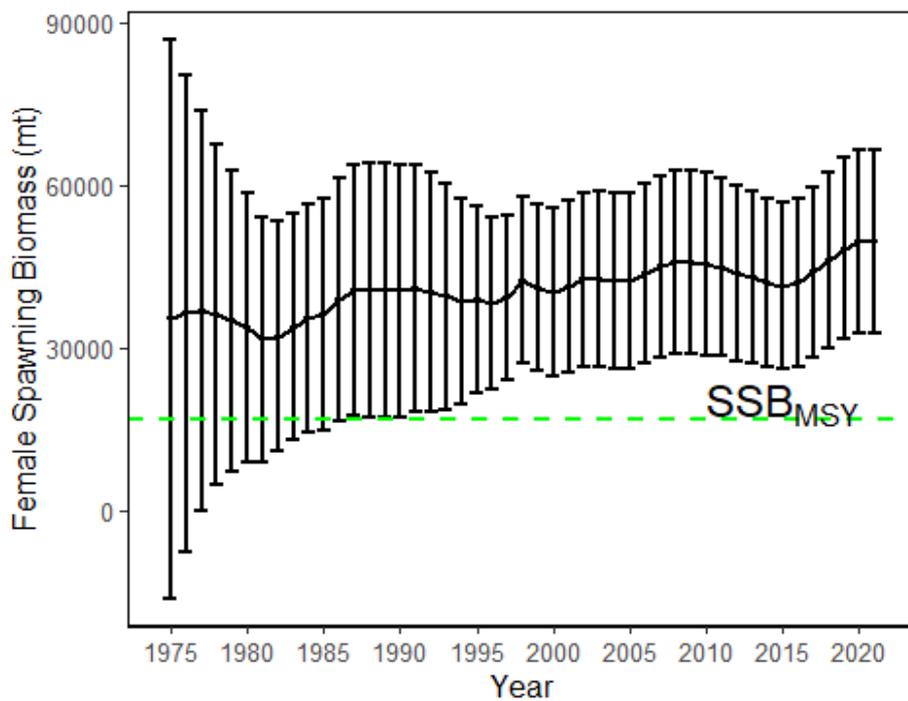


Figure 24: Estimated NP swordfish Spawning Stock Biomass (SSB) from the with 95% confidence intervals. SSB_{MSY} is indicated by the dashed green line.

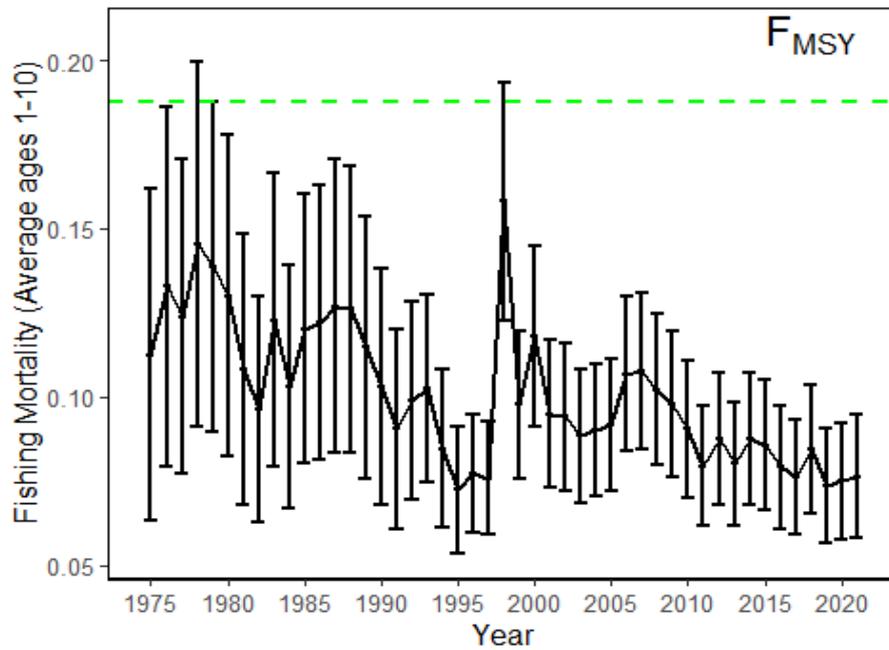


Figure 25: Estimated annual fishing mortality (average ages 1-10) the with 95% confidence intervals. F_{MSY} is indicated by the dashed green line.

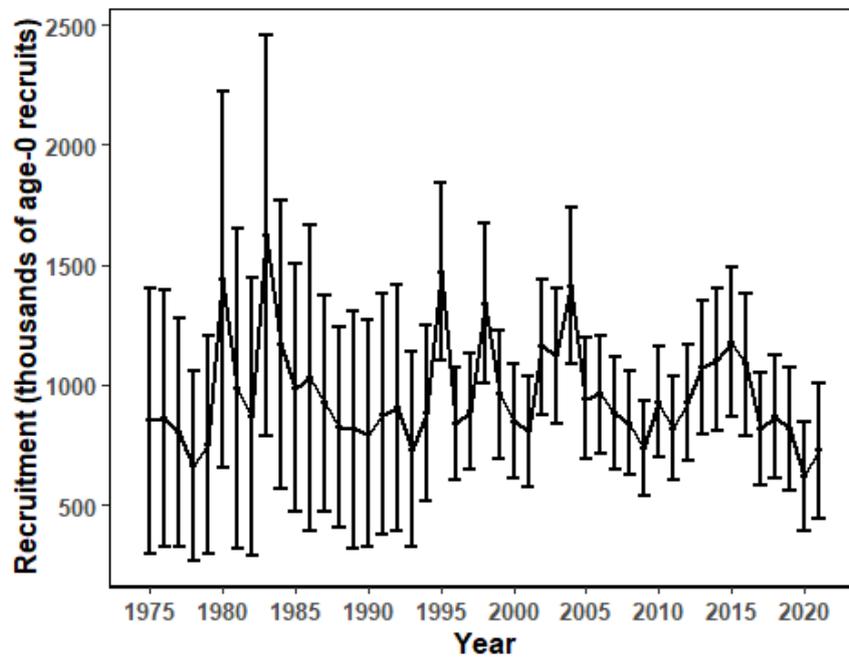


Figure 26: Estimated annual recruitment (thousands of age-0 fish) with 95% confidence intervals.

Diagnostics

Profiling on R_0 showed that the recruitment estimates were highly influential in the model results, but there was relative agreement between the CPUE indices and the length composition data on the lower bound of R_0 (Figure 27 - Figure 29). The US data (CPUE and length comp) drive the model dynamics suggesting an $\ln(R_0)$ below 6.0, and Chinese Taipei data and Japanese size composition data suggesting an $\ln(R_0)$ around 7. Japanese CPUE data suggest an $\ln(R_0)$ around 6.1 (Tables 4-5).

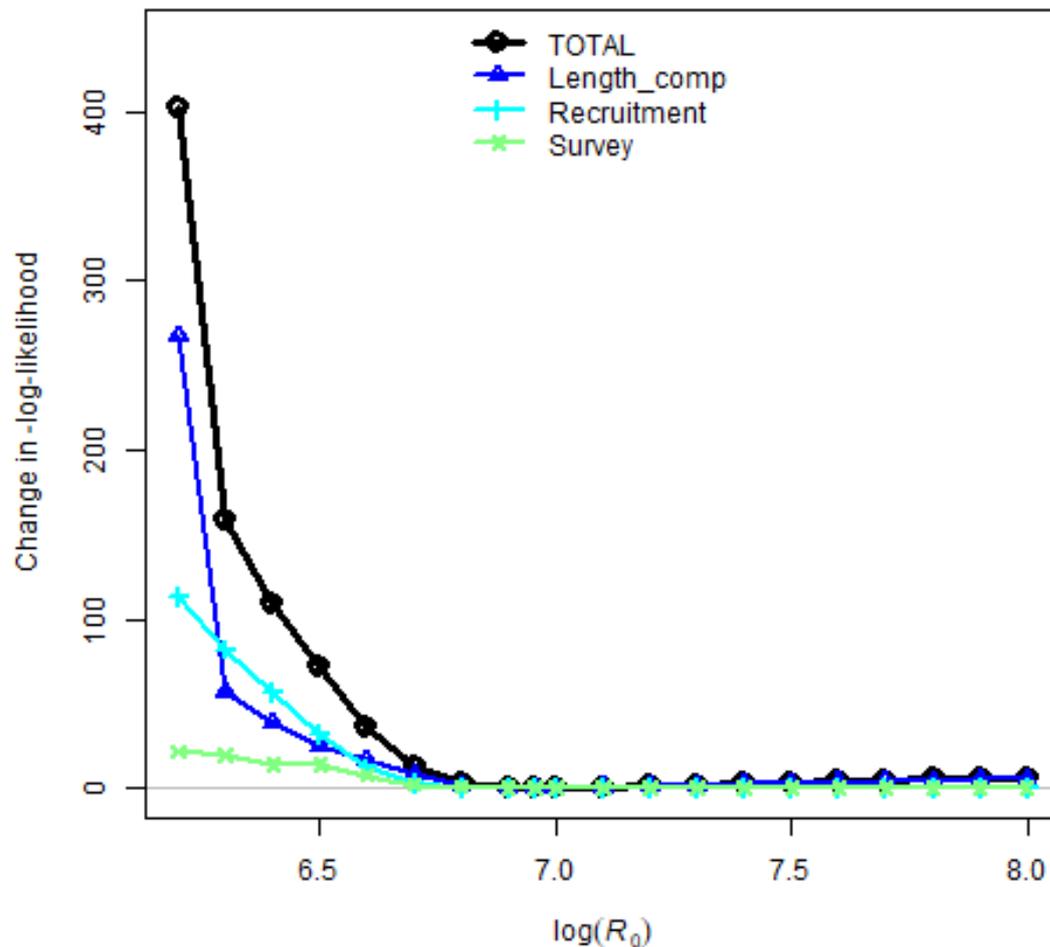


Figure 27: Likelihood profile over R_0 for the base-case model: total likelihood (black circles), recruitment (blue triangles), length composition data (light blue crosses), and survey/CPUE indices (yellow x).

Changes in index likelihood by fleet

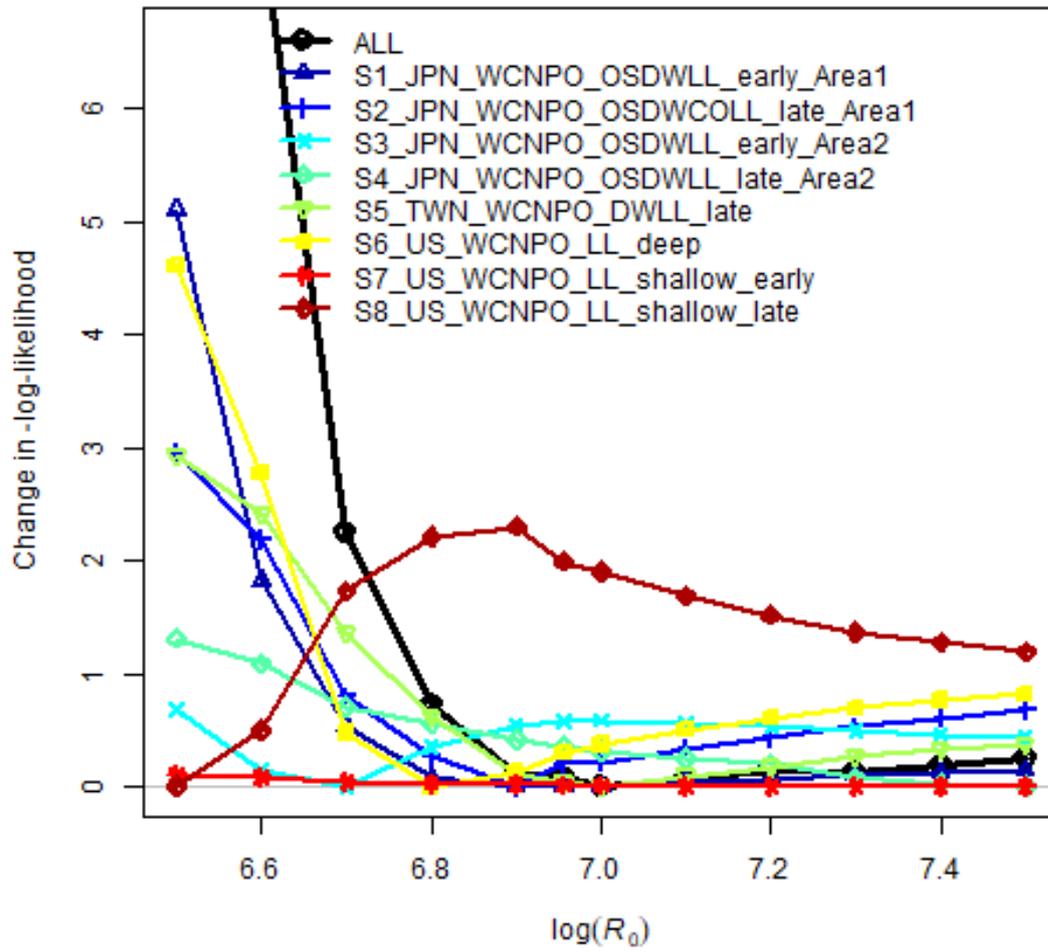


Figure 28: Likelihood profile over R_0 by CPUE index for the base-case model.

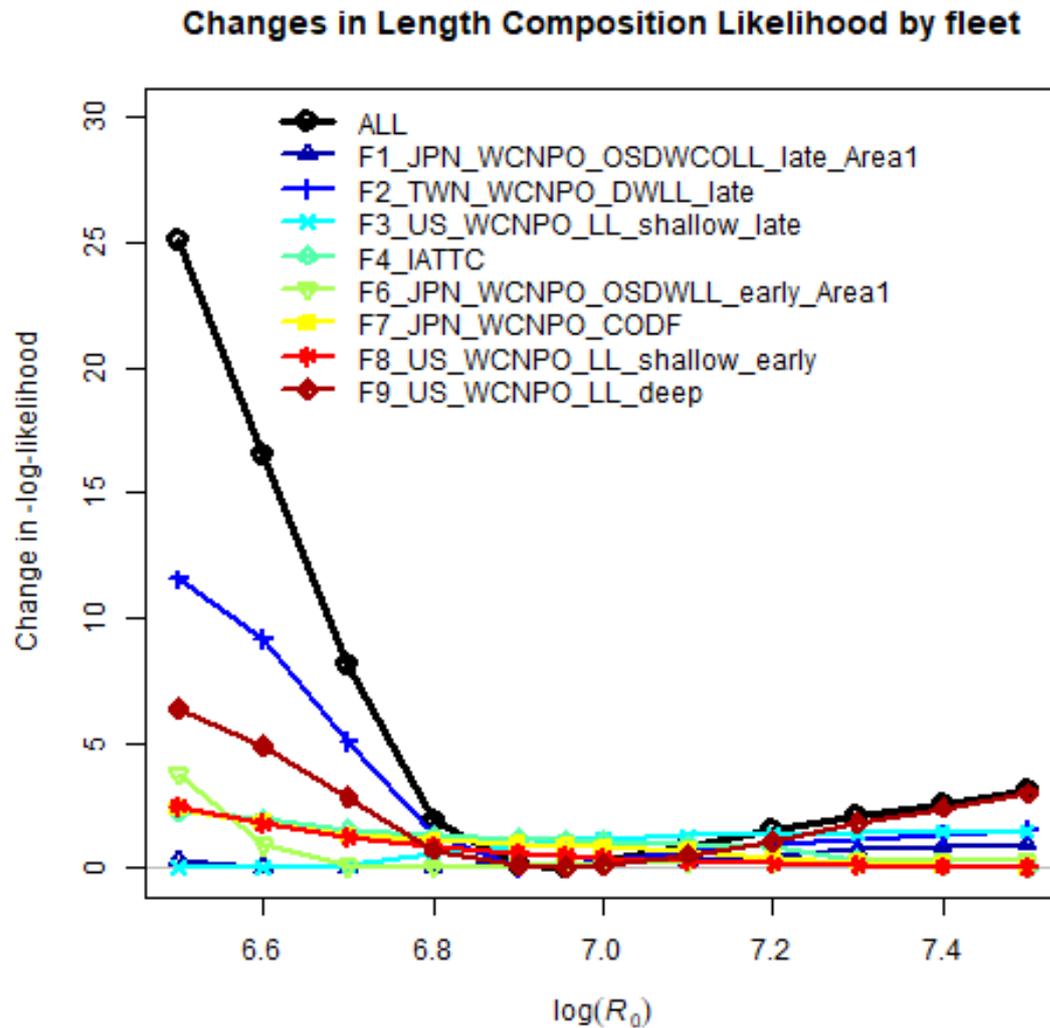


Figure 29: Likelihood profile over R_0 for each length composition time series for the base-case model.

Results from the ASPM model showed a similar population trend as the full model although the scale of the ASPM is larger than the base-case model (Figure 30). This suggests that while the Catch and CPUE data do provide information for the production function, the size composition data provide information about the overall scale of the population.

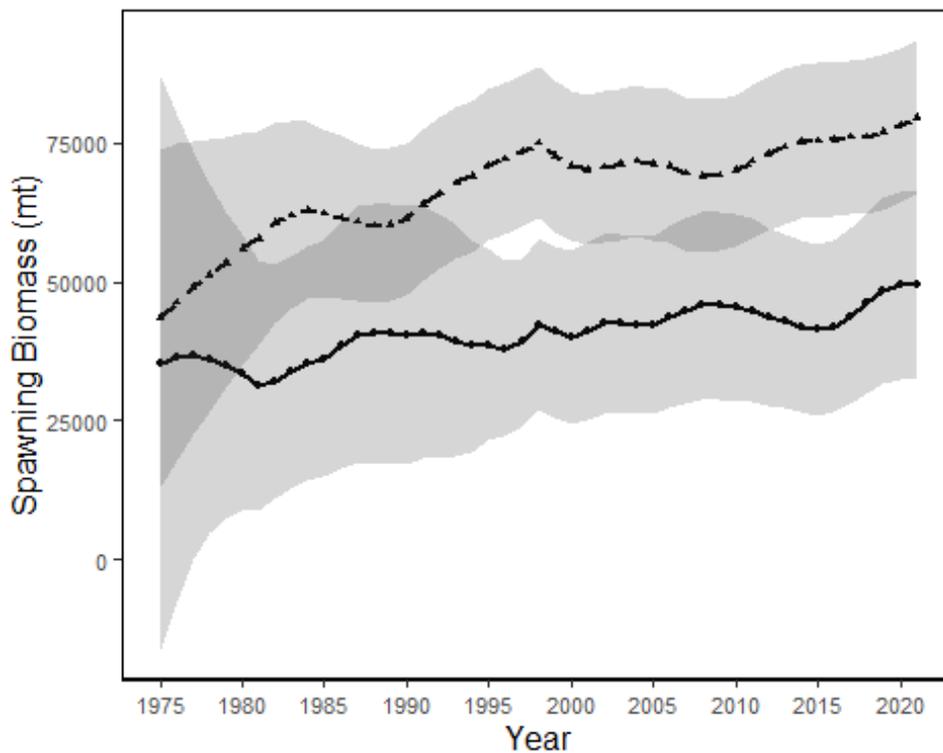


Figure 30: Spawning stock biomass trend for the ASPM model run (dashed line, triangles) and the base-case model (solid line, circles). Grey shading indicates 95% confidence intervals for each model.

Results of the hindcast with cross-validation indicate that of the five CPUE indices at the end of the assessment horizon, only the Japanese LL area 2 fleet had reasonable predictive ability (MASE = 0.85), with all other fleets MASE > 1 (Figure 31). Comparing the predictive ability of the size composition data, two fleets had good predictive ability (MASE < 0.5, F3 and F9), one had good predictive ability (MASE < 1 and > 0.5, F2) and two had poor predictive ability (MASE < 1, F1 and F7, Figure 32).

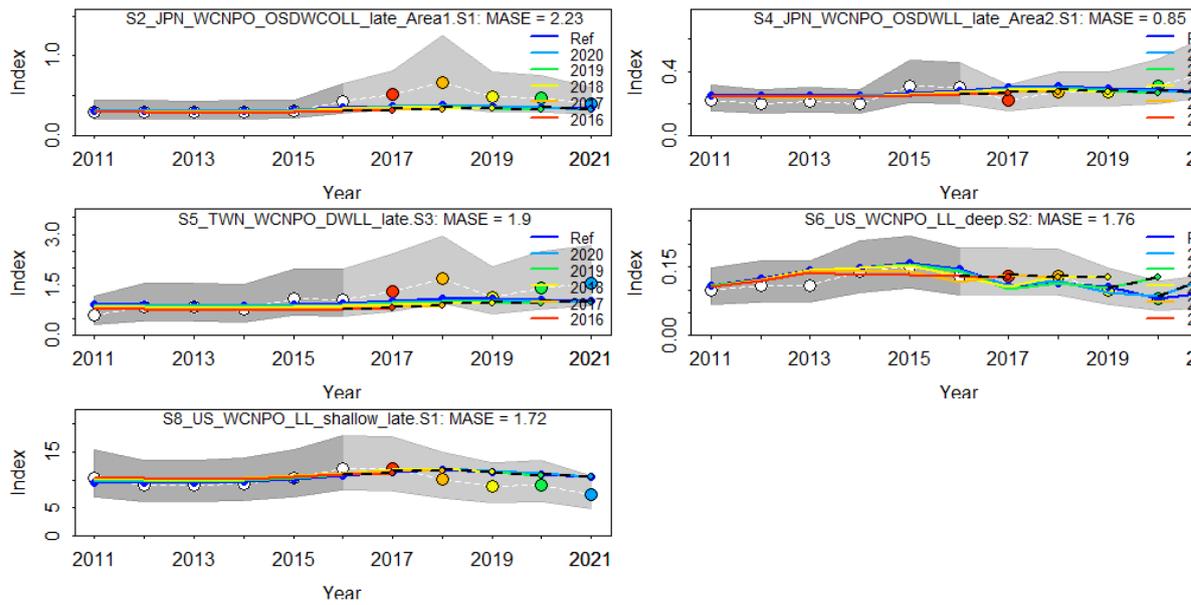


Figure 31: Hind casting cross-validation (HCxval) results for Japanese longline late area 1 (top right), Japanese LL late area 2 (top left), Chinese Taipei deep water longline late (center right), US Hawaii deep-set longline (center left) CPUE, and US Hawaii deep-set longline (bottom left) fits, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the old growth model. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95% confidence intervals (light-grey shading). The model reference year refers to the endpoint of each one-year-ahead forecast and the corresponding observation. The mean absolute scaled error (MASE) score associated with each CPUE time series is denoted in each panel.

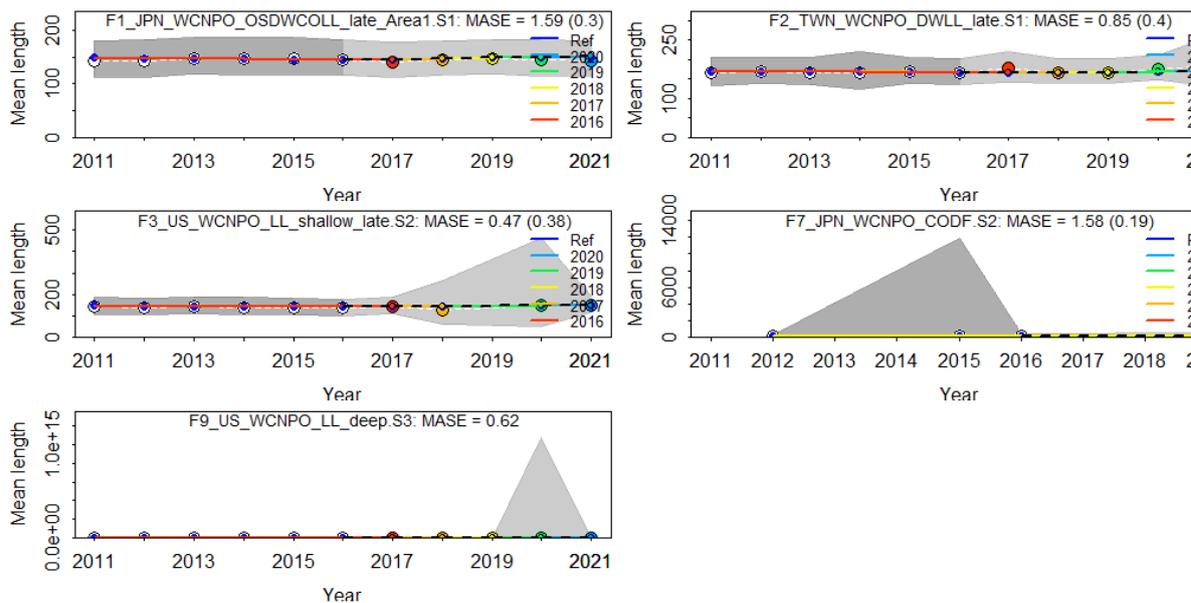


Figure 32: Hind casting cross-validation (HCxval) results for size composition mean lengths, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the old growth model. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95% confidence intervals (light-grey shading). The model reference year refers to the endpoint of each one-year-ahead forecast and the corresponding observation. The mean absolute scaled error (MASE) score associated with each size composition time series is denoted in each panel.

Conclusions

Overall, the model diagnostics do not indicate substantial misfit of the model to the data. While there remains some room for improvement in the fits to the US deep-set LL size data, this fleet is a very minor contributor to the total catch and likely would not be influential to the overall stock biomass trends. The results of the preliminary model indicate that the stock is not overfished and that overfishing is not occurring. Inclusion of the North EPO area to the model results in an increase in total biomass in the North Pacific compared to the 2018 WCNPO SWO assessment, but the results are generally consistent.

References

- Bohaboy, E. and Sculley M. (2023). Standardization of Western and Central North Pacific swordfish (*Xiphias gladius*) Catch Per Unit Effort in the Hawai'i longline fishery from 1995-2021. ISC/22/BILLWG-02/02.
- Brodziak, J. (2020). On the probably distribution of stock-recruitment steepness for Western and Central North Pacific swordfish. ISC/20/BILLWG-01/06.
- Carvalho, F., A. E. Punt, Y.-J. Chang, M. N. Maunder and K. R. Piner (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., Wetzel C.R., Doering, K. Johnson, K.F. and Methot, R.D. (2021). A cookbook for using model diagnostics in integrated stock assessments. Doi: 10.1016/j.fishres.2021.105959
- DeMartini, E.E., Uchiyama, J.H., Humphreys Jr., R.L., Sampaga, J.D. and Williams, H.A. (2007). Age and growth of swordfish (*Xiphias gladius*) caught by the Hawaii-based pelagic longline fishery. Fish. Bull. 105:356-367.
- Francis, R. I. C. C. (2011). Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68(6): 1124-1138.
- Ichinokawa, M., & Brodziak, J. (2008). Stock boundary between possible swordfish stocks in the northwest and southeast Pacific judged from fisheries data of Japanese longliners. ISC/08/Special Session on Billfish Stock Structure, 4, 14.

- ISC BILLWG. (2023). Report of the Billfish Working Group Workshop. 13, 15-18 December 2023.
- Kapur, M., Brodziak, J. Fletcher, E. and Yau, A. (2017). Summary of Life History and Stock Assessment Results for Pacific Blue Marlin, Western and Central North Pacific Striped Marlin, and North Pacific Swordfish. ISC/17/BILLWG-01/02.
- Lee, H.-H., K. R. Piner, R. D. Methot Jr and M. N. Maunder (2014). Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. *Fisheries Research* 158: 138-146.
- Methot, R. D. and I. G. Taylor (2011). Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* 68(10): 1744-1760.
- Methot Jr, R. D. and C. R. Wetzel (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142: 86-99.
- Nishikawa, Y., M. Honma , S. Ueyanagi , S. Kikkawa. 1985. Average distribution of larvae of oceanic species of scombroid fishes, 1956-1981. *Far Seas Fish. Res. Lab., S Series* 12, 99p.
- Taylor, I.G.; Stewart, I.J.; Hicks, A.C.; Garrison, T.M.; Punt, A.E.; Wallace, J.R.; Wetzel, C.R.; Thorson, J.T.; Takeuchi, Y.; Ono, K.; Monnahan, C.C.; Stawitz, C.C.; A'mar, Z.T.; Whitten, A.T.; Johnson, K.F.; Emmet, R.L.; Anderson, S.C.; Lambert, G.I.; Stachura, M.M; Cooper, A.B.; Stephens, A.; and Klaer, N. (2021). r4ss package: R Code for Stock Synthesis. Version 1.28.0. URL <https://github.com/r4ss>

Tables

Table 1. Descriptions of fisheries catch and abundance indices included in the base case model for the stock assessment.

Fleet No	Fleet name	Catch units	Size data	CPUE
F1	F1_JPN_WCNPO_OSDWCOLL_late_Area1	N	Y	S2_JPN_WCNPO_OSDWLL_late_Area1
F2	F2_TWN_WCNPO_DWLL_late	B	Y	S5_TWN_WCNPO_DWLL_late
F3	F3_US_WCNPO_LL_shallow_late	N	Y	S8_US_WCNPO_LL_shallow_late
F4	F4_IATC	B	Y	N
			Y	
F5	F5_JPN_EPO_OSDWLL	N	(not used)	N
F6	F6_JPN_WCNPO_OSDWLL_early_Area1	N	Y	S1_JPN_WCNPO_OSDWLL_early_Area1
F7	F7_JPN_WCNPO_CODF	B	Y	N
F8	F8_US_WCNPO_LL_shallow_early	N	Y	S7_US_WCNPO_LL_shallow_early
F9	F9_US_WCNPO_LL_deep	N	Y	S6_US_WCNPO_LL_deep
			N	
			(mirror	
F10	F10_JPN_WCNPO_OSDF	B	F6)	N
			N	
			(mirror	
F11	F11_JPN_WCNPO_Other_early	B	F6)	N
			N	
			(mirror	
F12	F12_JPN_WCNPO_Other_late	B	F1)	N
F13	F13_TWN_WCNPO_DWLL_early	B	N	N

			(mirror F2) N	
F14	F14_TWN_WCNPO_Other	B	(mirror F1) N	N
F15	F15_US_WCNPO_GN	B	(mirror F2) N	N
F16	F16_US_WCNPO_Other	B	(mirror F2) N	N
F17	F17_JPN_WCNPO_OSDWLL_early_Area2	N	(mirror F3) N	S3_JPN_WCNPO_OSDWLL_early_Area2
F18	F18_JPN_WCNPO_OSDWLL_late_Area2	N	(mirror F3) N	S4_JPN_WCNPO_OSDWLL_late_Area2
F19	F19_WCPFC	B	(mirror F2)	N

Table 2. Mean input standard error (SE) in log-space (i.e., $\log(\text{SE})$) of lognormal error and root-mean-square-errors (RMSE), and additional variance added for the relative abundance indices for North Pacific swordfish used in the base-case model.

Fleet	RMSE	Mean input SE	Input+Additional Variance	Additional Variance	Fleet Name
20	0.143	0.201	0.201	0	S1_JPN_WCNPO_OSDWLL_early_Area1
21	0.190	0.203	0.203	0	S2_JPN_WCNPO_OSDWCOLL_late_Area1
22	0.170	0.202	0.202	0	S3_JPN_WCNPO_OSDWLL_early_Area2
23	0.161	0.198	0.198	0	S4_JPN_WCNPO_OSDWLL_late_Area2
24	0.302	0.205	0.315	0.11	S5_TWN_WCNPO_DWLL_late
25	0.148	0.2	0.2	0	S6_US_WCNPO_LL_deep
26	0.024	0.2	0.2	0	S7_US_WCNPO_LL_shallow_early
27	0.169	0.2	0.2	0	S8_US_WCNPO_LL_shallow_late

Table 3. Key life history, recruitment, and selectivity parameters for the NP swordfish model. From Table 2 in the ISC BILLWG Data Preparatory report (2023).

Parameter	Female	Male	Reference
Growth age for L1	1	1	-
Growth age for L2	15	15	-
Natural mortality	0.42 (0)	0.4 (0)	Kapur et al. 2017
	0.37 (1)	0.38 (1)	
	0.32 (2)	0.37 (2)	
	0.27 (3)	0.37 (3)	
	0.22 (4+)	0.37 (4)	
		0.37 (5)	
		0.36 (6+)	
L at Amin GP 1	97.7	99	DeMartini et al. 2007
L at Amax GP 1	226.3	206.4	DeMartini et al. 2007
VonBert K GP 1	0.246	0.271	DeMartini et al. 2007
CV young GP 1	0.1	0.1	
CV old GP 1	0.1	0.1	
Weight – length par 1	1.30E-05	1.30E-05	DeMartini et al. 2007
Weight – length par 2	3.07	3.07	DeMartini et al. 2007
50% maturity length	143.68	-	Kapur et al. 2017
Mat slope	-0.1034	-	
Fecundity	Proportional to spawning biomass	-	
Spawning season	July	-	Nishikawa 1985
R ₀	0.42	-	
Steepness	0.9	-	Brodziak 2020

Table 4. Relative negative log-likelihoods of abundance index data components in the base case model over a range of fixed levels of virgin recruitment in log-scale ($\log(R_0)$). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of $\log(R_0)$ was 7.25. See Table 1 for a description of the abundance indices.

$\ln(R_0)$	S01	S02	S03	S04	S05	S06	S07	S08
6.2	4.54	3.43	1.46	2.42	3.27	8.97	0.26	0
6.3	2.36	2.57	2.01	1.74	2.69	5.77	0.21	0.52
6.4	2.38	2.21	1.40	1.48	2.39	4.52	0.18	0.79
6.5	1.82	1.80	0.67	1.39	2.10	3.32	0.16	1.07
6.6	0.84	1.00	0	1.03	1.46	1.20	0.11	1.64
6.7	0.11	0.18	0.35	0.71	0.26	0	0.06	2.42
6.8	0	0	0.62	0.63	0	0.09	0.03	2.51
6.9	0.03	0.06	0.67	0.57	0.14	0.30	0.02	2.32
7	0.08	0.15	0.67	0.50	0.27	0.47	0.01	2.12
7.1	0.12	0.25	0.65	0.44	0.38	0.61	0.004	1.94
7.2	0.15	0.37	0.60	0.35	0.47	0.72	0.002	1.79
7.26	0.17	0.44	0.57	0.23	0.55	0.79	0.0002	1.69
7.3	0.19	0.48	0.56	0.20	0.58	0.82	0.0001	1.65
7.4	0.22	0.57	0.53	0.15	0.64	0.89	0	1.55
7.5	0.25	0.65	0.50	0.12	0.69	0.95	0.0002	1.47
7.6	0.28	0.73	0.48	0.09	0.73	1.00	0.001	1.39
7.7	0.30	0.80	0.46	0.06	0.78	1.04	0.001	1.33
7.8	0.32	0.86	0.44	0.03	0.82	1.07	0.002	1.27
7.9	0.34	0.91	0.43	0.02	0.85	1.10	0.002	1.22
8	0.36	0.96	0.41	0	0.88	1.13	0.003	1.18

Table 5. Relative negative log-likelihoods of length composition data components in the base case model over a range of fixed levels of virgin recruitment in log-scale ($\log(R_0)$). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of $\log(R_0)$ was 7.26. See Table 1 for a description of the composition data.

$\ln(R_0)$	F01	F02	F04	F06	F08	F09
6.2	4.66	5.92	2.90	19.79	5.14	35.40
6.3	1.49	5.80	2.72	16.59	2.95	9.55
6.4	1.37	5.29	2.51	10.27	4.24	9.25
6.5	1.07	4.43	2.24	4.36	4.00	8.58
6.6	0.73	2.33	1.70	0.35	1.76	6.66
6.7	0.36	0.50	1.30	0	1.31	3.78
6.8	0.13	0.19	1.18	0.12	0.97	1.61
6.9	0	0.03	1.12	0.22	0.74	0.50
7	0.002	0	1.06	0.29	0.56	0.07
7.1	0.05	0.03	1.00	0.34	0.43	0
7.2	0.09	0.07	0.94	0.36	0.35	0.16
7.26	0.21	0.08	0.88	0.38	0.31	0.51
7.3	0.21	0.10	0.77	0.38	0.28	0.62
7.4	0.21	0.15	0.57	0.39	0.22	0.91
7.5	0.21	0.19	0.42	0.39	0.17	1.22
7.6	0.21	0.22	0.30	0.39	0.13	1.55
7.7	0.21	0.24	0.20	0.39	0.09	1.88
7.8	0.20	0.27	0.12	0.39	0.06	2.18
7.9	0.20	0.29	0.05	0.39	0.03	2.48
8	0.19	0.30	0	0.39	0	2.75