A Base Case Model in Stock Synthesis 3.30 for the 2018 North Pacific Swordfish (*Xiphias gladius*) Stock Assessment

Michelle Sculley[a] Hirotaka Ijima[b] and Yi-Jay Chang[c]

a Joint Institute for Marine and Atmospheric Research, University of Hawaii
c/o National Marine Fisheries Service
1845 Wasp Boulevard
Honolulu, HI 96818

b National Research Institute of Far Seas Fisheries, Fisheries Research Agency
5-7-1 Orido, Shimizu-ku
Shizuoka-shi, Shizuoka, 424-8633 Japan

c Institute of Oceanography, National Taiwan University
No.1, Sec. 4, Roosevelt Road
Taipei 106, Taiwan

Working document submitted to the Billfish Working Group Swordfish Stock Assessment Meeting, International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, April 17-24, 2018, Shimizu, Japan. Document not to be cited with author's permission.

Abstract

A base case model in Stock Synthesis 3.30 for North Pacific Swordfish (*Xiphias gladius*) is described. The base case model covers 1975-2016 for the Western Central North Pacific Ocean (WCNPO) region as determined by the Billfish Working Group at the January 2018 working group meeting. It includes all the data available for the WCNPO region as of the January Billfish WG data preparatory meeting with the exception of two WCNPO indices which are not included in the likelihood estimation, and includes data from three International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) countries and from other countries in aggregate from the Western Central Pacific Fisheries Commission (WCPFC) and Inter-American Tropical Tuna Commission (IATTC). Two alternative models are also described. Alternative model one was the base case model with the inclusion of the remaining two WCNPO indices in the likelihood estimation, to evaluate how they may have impacted model results if included. Alternative model two was the base case model plus two environmental indices for recruitment. These indices are the Southern Oscillation Index from 1952-2016 which has been shown to correlate with swordfish recruitment deviations, and an index of estimated phytoplankton biomass from 2002-2016 which has been shown to correlate with bigeye tuna recruitment. The final base case model has converged, but additional work is required to improve the fit to the CPUE and length composition data. Initial results suggests the WCNPO swordfish stock is being fished below 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 in 2018. The ISC BILLWG data preparatory meeting was held in January 2018 to evaluate new life history, catch, length, and CPUE data and strategize for the assessment (ISC Billfish WG, 2018). It was decided to run the assessment using a two stock model in Stock Synthesis version 3.30 (Methot and Wetzel, 2013) using fleets as areas but prioritizing an assessment of the Western Central North Pacific Ocean (WCNPO) over an assessment of the entire North Pacific stock. This document details a preliminary base case model of the WCNPO region, and several alternative models for consideration by the working group. The preliminary base case model was a product of collaboration of a modeling sub-group of the ISC Billfish WG including a representative from each country present at the data preparatory meeting: Michelle Sculley (USA), Hirotaka Ijima (Japan) and Yi-Jay Chang (Taiwan). A series of teleconferences were held with the subgroup members to develop the preliminary base case model. A detailed document on the data available for this assessment will be presented separately at this meeting. The final base case model is a result of the ISC BILLWG Stock Assessment meeting held in Shimizu, Japan in April 2018.

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 the equator (Figure 1). Countries were asked to contribute catch, CPUE, and length frequency data partitioned by these two regions so that two SS models could be developed: one of just the Western Central North Pacific Ocean (WCNPO) and one of the entire North Pacific Ocean with fleets as areas. The working group agreed to start model in 1952. The priority was to develop the WCNPO model and address the North Pacific model time permitting.

Definition of Fisheries

Data are available for thirty different fleets in the WCNPO: 18 catch time series, 12 CPUE indices of which one is a recruitment index, and two environmental indices. 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 2018; 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; Mex_LL_EPO is Mexican longline gear in the EPO; WCPFC_LL is longline gear in the WCNPO; IATTC_LL is longline gear in the EPO north of the equator; IATTC_LL_Overlap is longline gear in the overlap area of the IATTC convention area and the WCNPO areas.

Catch

The 18 time series of catch for the WCNPO model were divided into early and late periods to coincide with divisions of the CPUE indices (Table 1, Figure 2). 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.

Relative Abundance Indices

The ten CPUE indices available for inclusion in the WCNPO model are detailed in the input data working paper by Sculley and Yau (WP01) submitted to this meeting. The CPUE 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 fleets (S5 and S6) were assigned to quarter 3; US longline deep-set (S7) was assigned to quarter 2, US longline shallow-set (S8 and S9) were assigned to quarter 2, and US gillnet (S10) was assigned to quarter 4. Of these, fleets S5 and S10 were excluded from the base case model. In the base case model, Taiwanese fleet S5 (longline early) was excluded from the likelihood estimation (but included in the model along with a selectivity) because of poor data quality (Chang, pers. comm.). US gillnet fleet S10 was similarly excluded from the likelihood estimation but included in the model along with a selectivity because the area covered was very small compared to the WCNPO region and it was suggested that it may not represent dynamics of the entire population. US longline deep-set fleet S7 was included as an index of recruitment because the fishery catches large numbers of young-of-the-year fish (Fleet type 33, Sculley et al. 2018). 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 0.25 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).

$$RMSE_{smoother} = \sqrt{\left(\frac{1}{N}\right)\sum_{t=1}^{N}(Y_t - \hat{Y}_t)^2}$$

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 seven WCNPO fleets; length composition data were detailed in the input data working paper (WP01) submitted for this meeting (Figure 2). 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 only represented two time periods and were sparse. Data were fit using a multinomial error structure. Length composition data were weighted using the 2-stage process based upon the Francis (2011) method. In the first stage, the effective sample size was scaled to a mean of 25 by multiplying each number of samples by a constant. The second stage 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). However, because the model was sensitive to reweighting of the length composition data, input sample sizes were not iteratively re-weighted in stage 2.

Initial Base case Model Description

The assessment was conducted with Stock Synthesis (SS) version 3.30.08.03-SAFE released 09/29/2017 using Otter Research ADMB 11.6 (Methot and Wetzel 2013). The WCNPO model was set up as a single area model with two sexes and four seasons (quarters). 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 Billfish WG 2018). In addition, the CV of the growth curve was set to 0.1 for 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 sigmaR (σ_r) fixed at 0.6.

Twenty-eight fleets were included in the model: 18 catch fleets and 8 survey fleets. The population was assumed to be in equilibrium prior to 1951, with an estimated equilibrium exploitation catch of 20 mt per quarter (80 mt annual total). This estimated catch was based upon a linear regression fit to the annual catch of the F1 data from 1952-1960 and extrapolated to 1951.

Main recruitment deviations were estimated from 1975-2016. 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 1952-1963. 1964-1982 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 0.95 in 1982. Full bias adjustment was from 1983-2016. 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 except for the IATTC Overlap length data which was estimated as a two parameter asymptotic logistic selectivity pattern. Survey selectivity patterns mirrored their respective catch fleets (Table 4

Table 4). Including dome-shaped selectivity on fleets F1-2, F6, F10, and F12-14 resulted in better fits to the length frequency data. An asymptotic lognormal selectivity was used for IATTC Overlap, F18, because the fleet was comprised of multiple countries' length composition data. Selectivity parameter priors were assumed to be diffuse lognormal for the asymptotic lognormal model and diffuse symmetric beta for the double normal model.

Model estimated time series of total biomass (B in metric tons, mt = 1000 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 (APSM) was used to evaluate the influence of the length composition data on the population trends (Carvalho *et al.*, 2017). 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). Finally, residual plots and plots of the observed vs expected data were examined to evaluate goodness-of-fit.

Alternative Model Descriptions

In addition to the base case model, two alternative models are summarized below. Alternative model one included all the WCNPO CPUE indices provided to the working group at the Data Preparation meeting, adding back in S5 and S10 in the likelihood estimation to the preliminary base case model. Alternative model two included two environmental indices as an index of recruitment to the final base case model. Survey 11 was an index of the median phytoplankton cell size in biomass (units of pg carbon where pg = 10^{-12} grams) from July through September in a box bounded by 30°N, 10°N 175°W, and 140°W. This area was chosen as it reflects the approximate fishing area of the US HI longline deep-set sector, which primarily catches young-of-the-year recruits. The median phytoplankton cell size has been shown to be strongly correlated with the strength of a year class in bigeye tuna in the North Pacific Ocean and may be a good predictor of strong year classes for North Pacific swordfish (P. Woodworth-Jefcoats, pers. comm.). These estimates were calculated from sea surface temperature and chlorophyll a measurements from satellite imagery. This index was included as survey type 31, an environmental index, which allows the index to be proportional to $e^{recruit\ dev}$ (Methot *et al.,* 2017). This index was assigned to month 7 which is the first month of recruitment to the fishery in the assessment model.

Survey 12 was the mean Southern Oscillation Index (SOI) for 1952-2016 in the spawning season April – July (NOAA NCDC, 2017). This index has been shown to be strongly correlated with the recruits per spawning biomass ($\rho = -0.55$, $p<0.001$) for North Pacific Swordfish (Brodziak *et al.,* 2010). This index was input as the log(SOI). Using survey type 33 which is an index of age-0 fish, this allowed for the model to approximate the same environmental relationship of $e^{recruit\ dev}$as survey type 31 and have lognormal error distribution, as recommended (Methot et al., 2017).

Final Base case Model Description

During the working group meeting an iterative approach was taken to produce a base case model which converged and had a maximum gradient component which was close to zero, which would indicate stable parameter estimation. The final model to be used for management advice was based upon the preliminary base case model described above with the following changes:

- The model start year was changed from 1952 to 1975. This change was made after discussion about the very large catches reported by Japan in the 1950s. Japanese scientists clarified that the reporting of catch during this period had high uncertainty due to the method of reporting catches from the fishermen. It was agreed that these very high catches were driving the initial population size and the population dynamics during this early period because there were no CPUE indices or length composition data to inform the model. Removing this data improved the convergence of the model. Estimation of early recruitment deviations began in 1960 and main recruitment deviations began in 1975. SS will estimate early recruitment deviations for each age class in the model if main recruitment deviations begin in the first year of the model.

- Four length composition time series were removed: Japan longline area 1 early (F1); Japan Coastal Driftnet (F6); US longline deep-set (F12); and US longline shallow-set early (F13). Fleet 12 was removed because it was a significant component in the log-likelihood however it had a very different selectivity pattern catching primarily age 0-1 fish and representing only ~0.5% of the total catch. Fleets F1, F6, and F13 were removed because they were shown to be in conflict with the trend in the CPUE indices and other length composition data from the profiling on $\ln(R_0)$ (Figure 3 and Figure 4).

- The phase for initial F was changed from 1 to 2. This allowed the model to estimate R_0 in the first phase and initial F in the second phase which makes the parameters less likely to be confounded.

- The selectivity patterns were changed for F2, F10, and F14. F10 (Taiwan longline) was changed from double normal to asymptotic lognormal. Selectivity for F2 and F14 were changed from a 6-parameter double normal pattern to a 4-parameter double normal pattern. In the 4-parameter double normal pattern parameters 5 and 6, which are the initial and final selectivity parameters, were decayed to small and large fish, respectively. This reduced the number of parameters to be estimated in the model and improved fitting and convergence.

- The selectivity patterns for F1 was mirrored to F2, F6 was mirrored to F18, and F12 and F13 were mirrored to F14.

- The CV of growth for old fish was changed from 0.1 to 0.15. SS is sensitive to the value of this parameter. A larger CV for growth of old fish allowed the model more flexibility to fit the large fish caught. This allowed to the model to fit the fish caught which were larger than L_{amax} which otherwise may have caused problems with fitting the length composition data and convergence of the model.

- Adjusted variance for the length composition data was changed from 0.5 to 1, which changed the average effective sample size from 12.5 to 25. Additional reweighting was not attempted as this would result in up-weighting the length composition data, which would not improve the model fit and cause problems with convergence.

- The model was found to be robust in the estimation of R0 and the selectivity parameters, but estimates of recruitment deviations changed significantly depending on the initial values provided. The initial recruitment deviation values also caused the maximum gradient component to change. The model was run iteratively until a maximum gradient component was close to zero and the par file from that model run was used for all addition model runs and diagnostics.

Initial Base case Model Results

The base case model ran in about 12 minutes, estimated 115 parameters, and had a total likelihood of 1614.05. The inverse Hessian was positive definite, which allowed for the estimation of parameter standard deviations and suggests that the model converged; however the maximum gradient component was 2.13 which is greater than 0.001 and suggests poor parameter estimation. None of the parameter estimates hit a bound but two selectivity parameters had correlation values of about 0.95 and 12 parameters had correlation values below 0.01: 7 early recruitment deviations (1952-1958) and 5 selectivity parameters. All of the early recruitment deviations (1951-1974) and 35 of 42 (83%) of the main recruitment deviations had CVs > 50%. 17 of 42 selectivity parameters had CVs >50%. These parameters were from the dome shaped selectivity functions and were either parameters 2 (the width of the plateau), 4 (descending width of the distribution), 5 (selectivity in the first length bin), or 6 (selectivity in the final length bin). All of the parameters below the threshold for uncorrelated parameters also had CVs > 50%.

Fits to the abundance indices were relatively good, with no substantial divergences between the expected and estimated CPUEs (Figure 5). However, all the indices in the last 4-5 years of the model showed increasing or stable CPUEs but the model estimated decreasing CPUEs. This pattern was likely driven by the decrease in mean length seen in the Japanese longline length composition data (Fleet 2) as this pattern was no longer present in the CPUE model estimates when the length composition data were excluded from the model.

Fits to the length composition data were also relatively good, although several problems are evident in the fitting to the Japanese length composition data (F1 and F2) and the US Hawaii longline deep-set length composition data (F12). The residual patterns for F1 and F2 show more small fish caught than expected from 1984-1993 and, to a lesser extent, 1994-1998 which suggest that the selectivity of the fleet changed during that time period (Figure 6-Figure 8). The residual patterns from F12 show systematic positive residuals for very small fish which are caught in large numbers for this fleet. This fishery does not target swordfish, it catches them as bycatch, and the CPUE index was an index of recruitment rather than relative abundance. Also, based upon these data's contribution to the likelihood, these data were relatively influential in the model results, therefore further discussion should be had on whether these data sets (F1, F2, F12) should be included in the assessment model. The model overall was highly sensitive to the inclusion and weighting of the length composition data. These data were highly influential in the likelihood and changing the relative weights often resulted in a failure to converge. Further investigation is necessary to better evaluate the use of the length composition data for North Pacific swordfish.

Final Base case Model Results

Model fits

The final base case model had 75 estimated parameters, took 6 minutes to run, had a final maximum gradient component of 0.05, and a total likelihood of 224.13. None of the parameter estimates hit a bound. No parameters had correlation values over 0.95 but 8 parameters had correlation values below 0.01: all were early recruitment values. All of the early recruitment deviations (1960-1974) and 37 of 42 (83%) of the main recruitment deviations had CVs > 50%. Only 2 of the 12 selectivity parameters had CVs >50%. These parameters were both parameter 2 (the width of the plateau) from the dome shaped selectivity functions for F2 Japan LL late area 1 and F14 US HI LL Shallow Late length composition data. All of the parameters below the threshold for uncorrelated parameters also had CVs > 50%.

Model estimates of the CPUE data were generally acceptable, although some patterning in the residuals was present in the S2 Japan LL Area 1 Late time series (Figure 9). Japanese scientists indicated that this patterning was likely due to remaining uncertainty in the standardization of the CPUE data as well as misfitting with the length composition data from this fleet. Patterns in selectivity suggest on average, F14 catches slightly larger fish than F2 and F18 catches slightly larger fish than F10 (Figure 10). The length composition data from F2 shows residual patterns which suggest the periodic presence of strong year classes (Figure 11). Also, length composition data from 1994-1998 are converted from weight data, therefore there is likely some uncertainty around the data in this time period. There were some large residuals for F14 US HI LL Shallow Late length composition which were primarily in quarters one and two (Figure 12). This suggests some seasonal changes in the selectivity of the fleet. It may be useful to explore seasonal selectivity patterns in future work. There were minimal residual patterns in the length composition data from F10 and F18 (Figure 13Figure 14).

A pattern of large positive recruitment deviations followed by 5-7 years of negative recruitment deviations suggests either some problem in the estimation of recruitment or strong periodic recruitment pluses (Figure 15). These pulses are supported by the evidence of strong year classes in the length composition data which are offset from these pulses by a few years. These recruitment deviations are also correlated with the Southern Oscillation Index (Figure 16).

Model estimates of age 1+ SSB show a relatively flat trend with a slight decrease form 1975-1999 and a slight increase from 2000 to2016. (Figure 17). Initial female spawning stock biomass was estimated to be approximately 97,000 mt. Early and main recruitment deviation bias adjustment was >2 times the ratio of RMSE to σ_r. Current depletion, as estimated as the age 1+ biomass in 2016 compared to the virgin age 1+ biomass was estimated to be 0.3.

Diagnostics

Profiling on R_0 showed that the length composition data and CPUE indices showed the same minimum likelihood solution (Figure 18-Figure 20). Results from the ASPM model showed the similar population trend as the base case model with the ASPM model is within the 95% confidence interval for the base case model, but estimated a slightly smaller initial spawning stock biomass, 97,000 mt compared to 97,000 mt in the full model (Figure 21). This suggested that the length composition data did not have substantial conflict with the abundance indices but did scale the population size slightly.

Alternative Models

Two alternative models were run in addition to the base case model. Alternative one was the preliminary base case model plus two additional CPUE indices, Taiwan LL early (S5) and US Gillnet (S10). Both of these indices were excluded from the base case model a priori for reasons described above, but an investigation was done to see if these indices would substantially alter the model results. Both indices exhibit overall CPUE trends that are flat (Figure 22), and are unlikely to provide additional information to the model. Based upon additional profiling, the Taiwan LL early CPUE showed some conflicting patterns in the likelihood profile over R_0 (not shown) and suggested that including the index in the base case model would have caused some model misspecification issues.

Alternative model 2 added two environmental indices to the final base case model to help estimate recruitment. Inclusion of these indices reduced the uncertainty around the recruitments and changed the number of recruits in some years, especially in the 1980s and 1990s but did not change the recruitment significantly in the 2000s (Figure 23) and did not result in any notable concerns in the diagnostics (not shown) or the female SSB estimates (Figure 24). Inclusion of the SOI index reduced the uncertainty around main recruitment deviations and some early recruitment deviations as the index was available as early as 1960 (Figure 25). Including this index would require assuming a steady state for projections as it is difficult to predict future values. The phytoplankton index was currently only available from 2002-2016, as it was calculated from sea surface temperature and chlorophyll *a* estimates from satellite observations. However, additional historical data are available from modeled estimates which could be used to extend the series back in time. Predictions of sea surface temperature around the Hawaiian Islands for the next 10 years have been shown to be relatively unbiased (Tommasi, *et al.,* 2017) and some preliminary work has shown chlorophyll *a* predictions to be unbiased 1-2 years in the future (Charles Stock, personal comm.) which would allow for the phytoplankton biomass to be calculated in future years. Therefore including this environmental index in the projections of this assessment model may be possible. Additional work is necessary to further explore this potential.

Conclusions

The base case model suggests that the 2016 fishing levels are below F_{MSY} and 2016 spawning stock biomass is above SSB_{MSY} (Figure 26). The base case model estimated initial population scale (R0) well, however the estimated current trends of stock includes some uncertainty. The addition of environmental indices reduced the uncertainty around the recruitment deviations but did not substantially change the stock status nor biomass estimates. The length composition data in the early period of the model needs further investigation to improve the fit within the model.

Acknowledgements

Dr. Jon Brodziak was involved in setting up the base case model, troubleshooting convergence problems, and answering questions on Stock Synthesis data input. Dr. Phoebe Woodworth-Jefcoats provided the phytoplankton recruitment index. Dr. Charles Stock provided information and preliminary results describing the potential for future predictions of sea surface temperature and chlorophyll *a* to estimate future values of phytoplankton biomass.

Literature Cited

Brodziak, J., Courtney, D., Piner, K., Lee, H.H., and DiNardo, G. (2010). Modeling recruitment responses of striped marlin (*Tetrapturus audax*) and swordfish (*Xiphias gladius*) to environmental variability in the North Pacific. PICES International Symposium: Climate Change Effects on Fish and Fisheries, April 25-29, 2010. Presentation.

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.

Chang, Y., Sun, C., Hsu, J., and Yeh, S. (2018). Standardized catch-rates of swordfish (*Xiphias gladius*) for the Taiwanese distant-water tuna longline fishery in the North Pacific Ocean for 1964-2016. ISC/18/BILLWG-01/06.

Courtney, D. and Fletcher, E. (2009). Input data for a North Pacific swordfish stock assessment using Stock Synthesis. ISC/09/BILLWG-02/04.

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., and 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.

Ijima, H. (2018). Brief information for Japanese fishery statistics of North Pacific swordfish (*Xiphias gladius*). ISC/18/BILLWG-01/03.

ISC Billfish WG. (2018). Report of the Billfish Working Group Workshop, 17-23 January 2018. Honolulu, HI, USA. ISC/18/BILLWG-01/REPORT.

Ito, R., Childers, J., and Yuhong, G. (2018). U.S. swordfish fisheries in the North Pacific Ocean. ISC/18/BILLWG-01/01.

Kanaiwa M. and Ijima H. (2018). Abundance indices of Swordfish (*Xiphias gladius*) by the Japanese offshore and distant-water longline fishery in the North-Western Central Pacific. ISC/18/BILLWG-01/07.

Lee, H.-H., Piner, K. R., Methot Jr., R. D., and Maunder, M. N. (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., A'mar, T., Wetzel, C., and Taylor, I. (2017). Stock Synthesis User Manual Version 3.30.05-3.30.08. November 14, 2017.

Methot, R. D. and Taylor, I. G. (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 Wetzel, C. R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142: 86-99.

NOAA NCDC. (2017). Spatially and temporally large-scale anomalies that influence the variability of the atmospheric circulation. Online Database https://www.ncdc.noaa.gov/teleconnections/ Accessed 31 March 2017.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sculley, M., Yau, A., and Kapur, M. (2018). Standardization of the swordfish *Xiphias gladius* *c*atch per unit effort data caught by the Hawaii-based longline fishery from 1994-2016 using generalized linear models. ISC/18/BILLWG-01/05.

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. (2017). r4ss package: R Code for Stock Synthesis. Version 1.28.0. URL https://github.com/r4ss

Tommasi, D., Stock, C. A., Alexander, M. A., Yang, X., Rosati, A., and Vecchi, G. A. (2017). Multi-Annual Climate Predictions for Fisheries: An Assessment of Skill of Sea Surface Temperature Forecasts for Large Marine Ecosystems. Frontiers in Marine Science 4:201.

Tables

Table 1. List of fleets with Catch and CPUE indices provided for the 2018 Western Central North Pacific Ocean Swordfish Stock Assessment and the source for more information about the standardization of the CPUE series and catch data.

Catch Index	Abundance Index	Fleet Name	Time Series	Source
F1	S1	JPN_WCNPO_OSDWLL_early_Area1	1975-1993	Kanaiwa and Ijima 2018
F2	S2	JPN_WCNPO_OSDWLL_late_Area1	1994-2016	Kanaiwa and Ijima 2018
F3	S3	JPN_WCNPO_OSDWLL_early_Area2	1975-1993	Kanaiwa and Ijima 2018
F4	S4	JPN_WCNPO_OSDWLL_late_Area2	1994-2016	Kanaiwa and Ijima 2018
F5	-	JPN_WCNPO_OSDF	1960-1992	Hirotaka Ijima, pers. comm.
F6	-	JPN_WCNPO_CODF	1993-2014	Hirotaka Ijima, pers. comm.
F7	-	JPN_WCNPO_Other_Early	1952-1993	Hirotaka Ijima, pers. comm.
F8	-	JPN_WCNPO_Other_Late	1994-2016	Hirotaka Ijima, pers. comm.
F9	S5	TWN_WCNPO_DWLL _early	1975-1999	Chang et al. 2018b
F10	S6	TWN_WCNPO_DWLL _late	2000-2016	Chang et al. 2018b
F11	-	TWN_WCNPO_Other	1959-2016	Yi-Jay Chang, pers. comm.
F12	S7	US_WCNPO_LL_deep	1995-2016	Sculley et al. 2018b
F13	S8	US_WCNPO_LL_shallow_early	1995-2000	Sculley et al. 2018b
F14	S9	US_WCNPO_LL_shallow_late	2005-2016	Sculley et al. 2018b
F15	S10	US_WCNPO_GN	1985-2006	Courtney et al. 2009
F16	-	US_WCNPO_Other	1970-2016	Ito *et al.,* 2018
F17	-	WCPFC_LL	1970-2016	Darryl Tagami pers. comm.
F18	-	IATTC_LL_Overlap	1975-2016	Shane Griffiths, pers. comm.

Table 2. Mean CV and calculated RMSE for the 10 CPUE Indices.

Fleet	Mean CV	RMSE
S1	0.009	0.127
S2	0.032	0.135
S3	0.018	0.166
S4	0.039	0.154
S5	0.213	1.147
S6	0.277	0.229
S7	0.492	0.152
S8	1.630	0.159
S9	0.371	0.188
S10	0.287	0.822

Table 3. Key life history, recruitment, and selectivity parameters used in the swordfish stock assessment model. The column labeled "Estimated ?" identifies if the parameters are expected to be estimated within the assessment model (Estimated), fixed at a specific value, i.e., not estimated (Fixed), or iteratively re-scaled to the match the predicted variance (Re-scaled). From Table 9.0 in the ISC BILLWG Data Preparatory report (2018).

Parameter (units)	**Value**	**Estimated?**
Natural mortality (M, age-specific yr)	Female: M_0 = 0.42, M_1 = 0.37, M_2 = 0.32, M_3 = 0.27, M_{4+} = 0.22	Fixed
	Male: M_0 = 0.40, M_{1-2} = 0.38, M_{3-5} = 0.37, M_{6+} = 0.36	
Length_at_min_age (EFL cm)	Female: $L(A_{min})$ = 97.7	Fixed
	Male: $L(A_{min})$ = 99.0	
Length_at_max_age (EFL cm)	Female: $L(A_{max})$ = 226.3	Fixed
	Male: $L(A_{max})$ = 206.4	
VonBert_K	Female: k = 0.246	Fixed
	Male: k = 0.271	
$W=aL^b$ (kg)	Both genders: a = 1.299 ×10-5	Fixed
	b = 3.0738	
Size at 50-percent maturity (EFL cm) and maturity ogive slope parameter	Female: L_{50} = 143.6, β = -0.103	Fixed
	Male: L_{50} = 102.0, β = -0.141	
Stock-recruitment steepness (h)	h= 0.9	Fixed
Unfished log-scale recruitment (Ln(R_0))	-	Estimated
Standard deviation of recruitment (σR)	σR = 0.6	Fixed
Initial age structure	-	Estimated
Recruitment deviations	-	Estimated
Selectivity	-	Estimated
Catchability		Estimated

Table 4. Table of selectivity functions for each catch and abundance time series.

Fleet	Selectivity Function Initial Base case	Selectivity Function Final Base case
F1	Double-normal	Mirror F2
F2	Double-normal	Double-normal
F3	Mirror F13	Mirror F14
F4	Mirror F14	Mirror F14
F5	Mirror F10	Mirror F10
F6	Double-normal	Mirror F18
F7	Mirror F1	Mirror F2
F8	Mirror F2	Mirror F2
F9	Mirror F10	Mirror F10
F10	Double-normal	Asymptotic lognormal
F11	Mirror F2	Mirror F2
F12	Double-normal	Mirror F14
F13	Double-normal	Mirror F14
F14	Double-normal	Double-normal
F15	Mirror F10	Mirror F10
F16	Mirror F10	Mirror F10
F17	Mirror F10	Mirror F10
F18	Asymptotic lognormal	Asymptotic lognormal
S1	Mirror F1	Mirror F2
S2	Mirror F2	Mirror F2
S3	Mirror F13	Mirror F14
S4	Mirror F14	Mirror F14
S5	Mirror F10	Mirror F10
S6	Mirror F10	Mirror F10
S7	Mirror F12	Mirror F14
S8	Mirror F13	Mirror F14
S9	Mirror F14	Mirror F14
S10	Mirror F10	Mirror F10

Figures

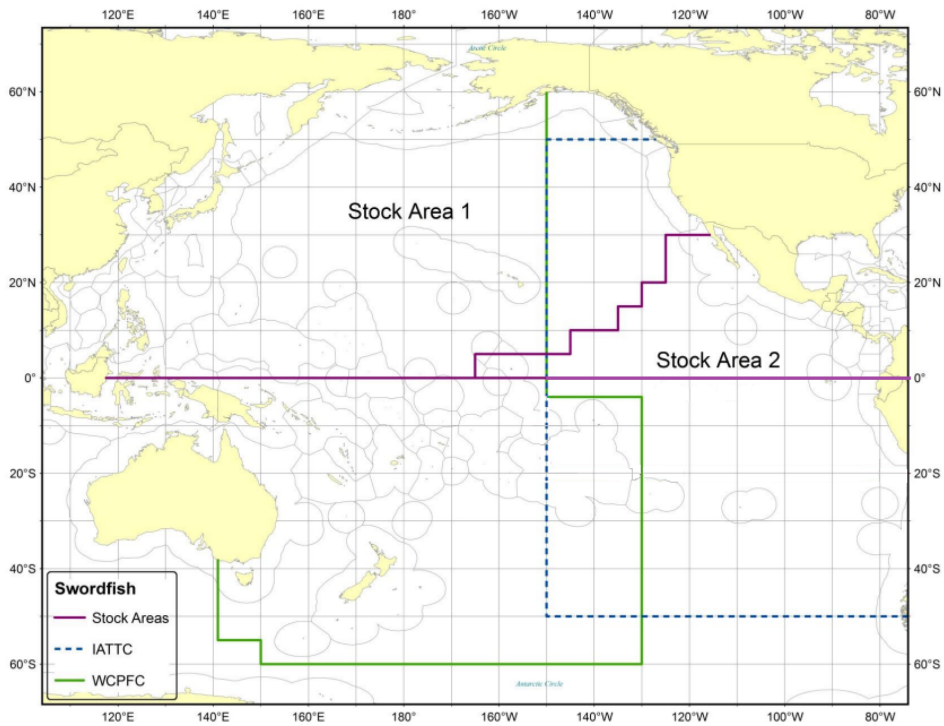

Figure 1. Stock boundaries for the 2018 North Pacific swordfish stock assessment, indicated by purple lines. Stock area 1 is the Western Central Pacific Ocean (WCNPO) and stock area 2 is the Eastern Pacific Ocean (EPO). The green line indicates the Western Central Pacific Fisheries Commission boundary and the blue dashed line indicates the Inter-American Tropical Tuna Commission boundary.

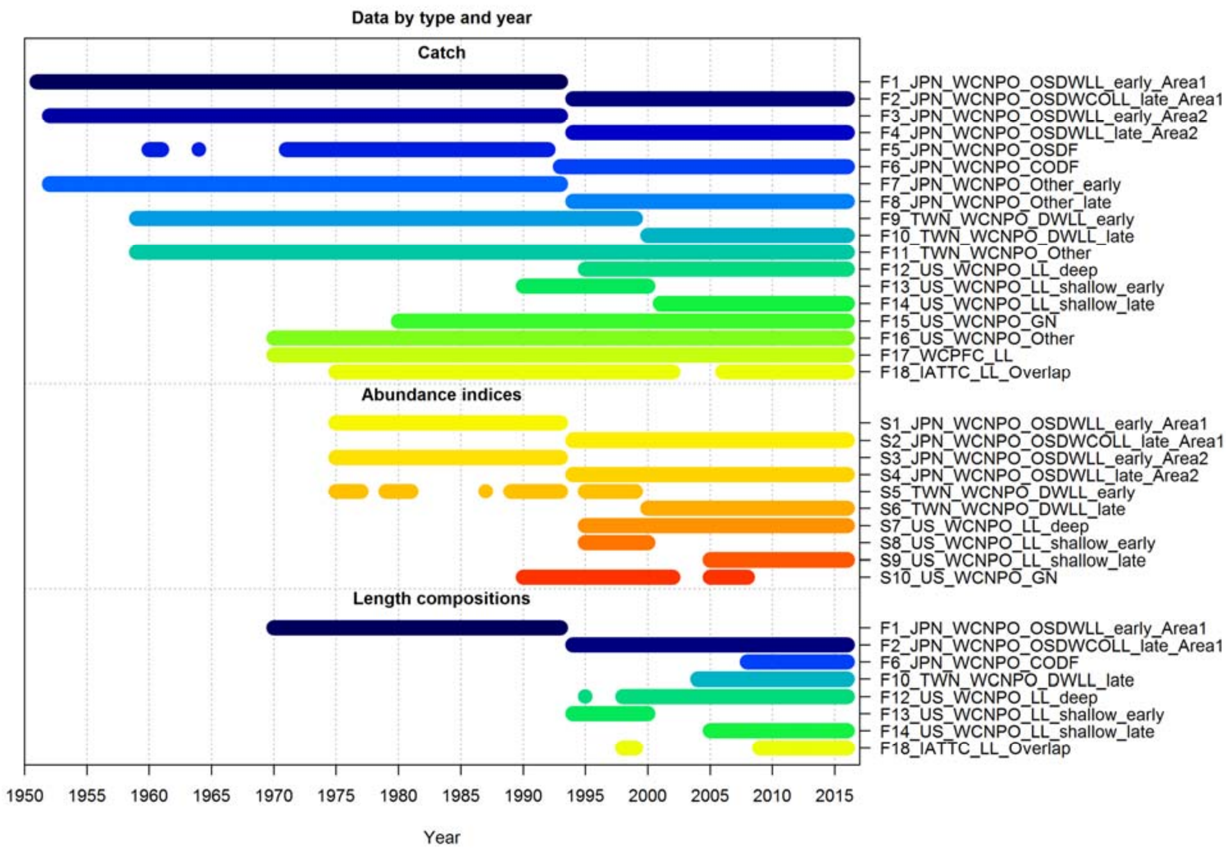

Figure 2. Catch, Abundance, and Length Composition data available for the WCNPO Stock Synthesis swordfish assessment model.

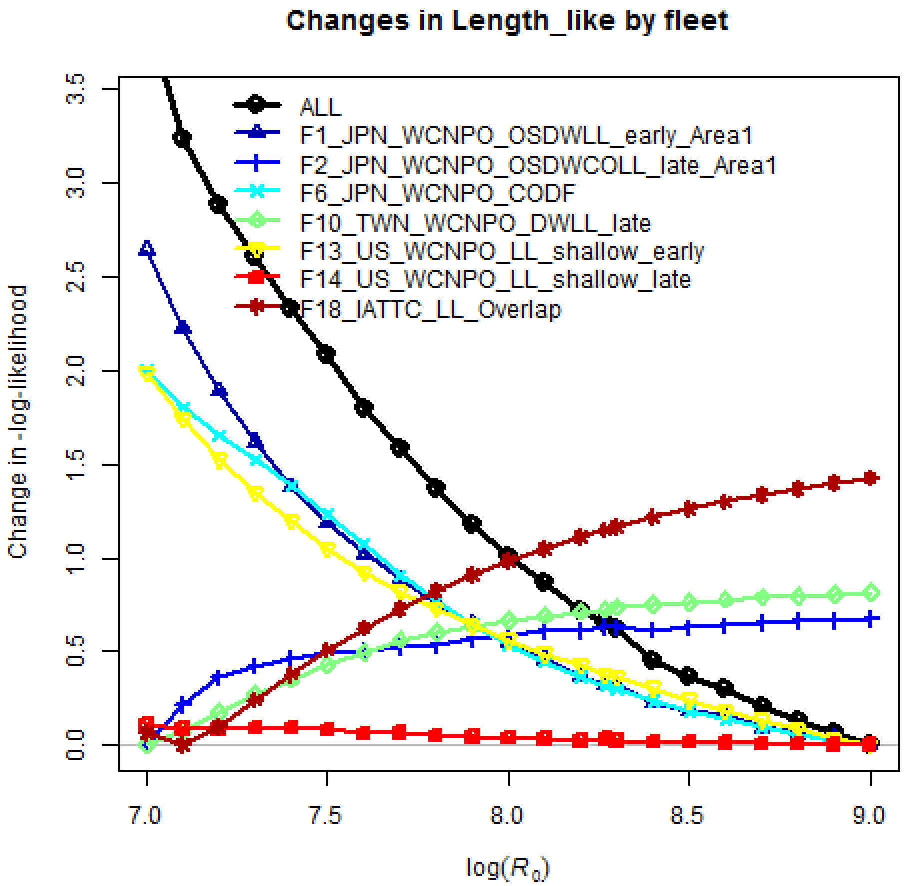

Figure 3. Likelihood profile on log(R_0) by fleet-specific length composition likelihood component in the initial base case model. F12 is excluded from the model due to poor fitting. F1 (dark blue triangles), F6 (light blue x), and F13 (yellow inverted triangles) trend shows a minum likelihood at a very large initial recruitment size while F2 (medium blue vertical lines), F10 (green diamonds), F14 (red squares), and F18 (dark red astericks) show a minimum likelihood around 7.1.

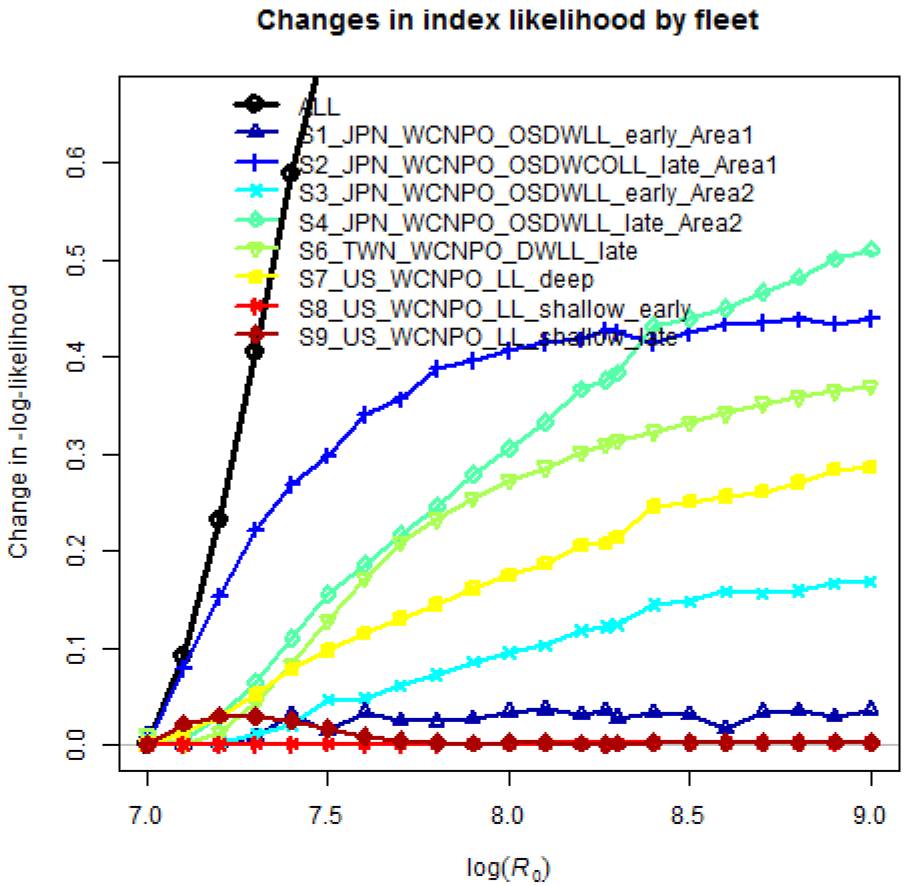

Figure 4. Likelihood profile on log(R0) by fleet-specific CPUE index likelihood component in the initial base case model. Fleets S1 (dark blue triangles), S2 (medium blue vertical lines), S3 (light blue x), S4 (green open diamonds), S6 (green inverted triangles), S7 (yellow squares), S8 (red astericks), and S9 (dark red closed diamonds) show a minimum likelihood around 7.0.

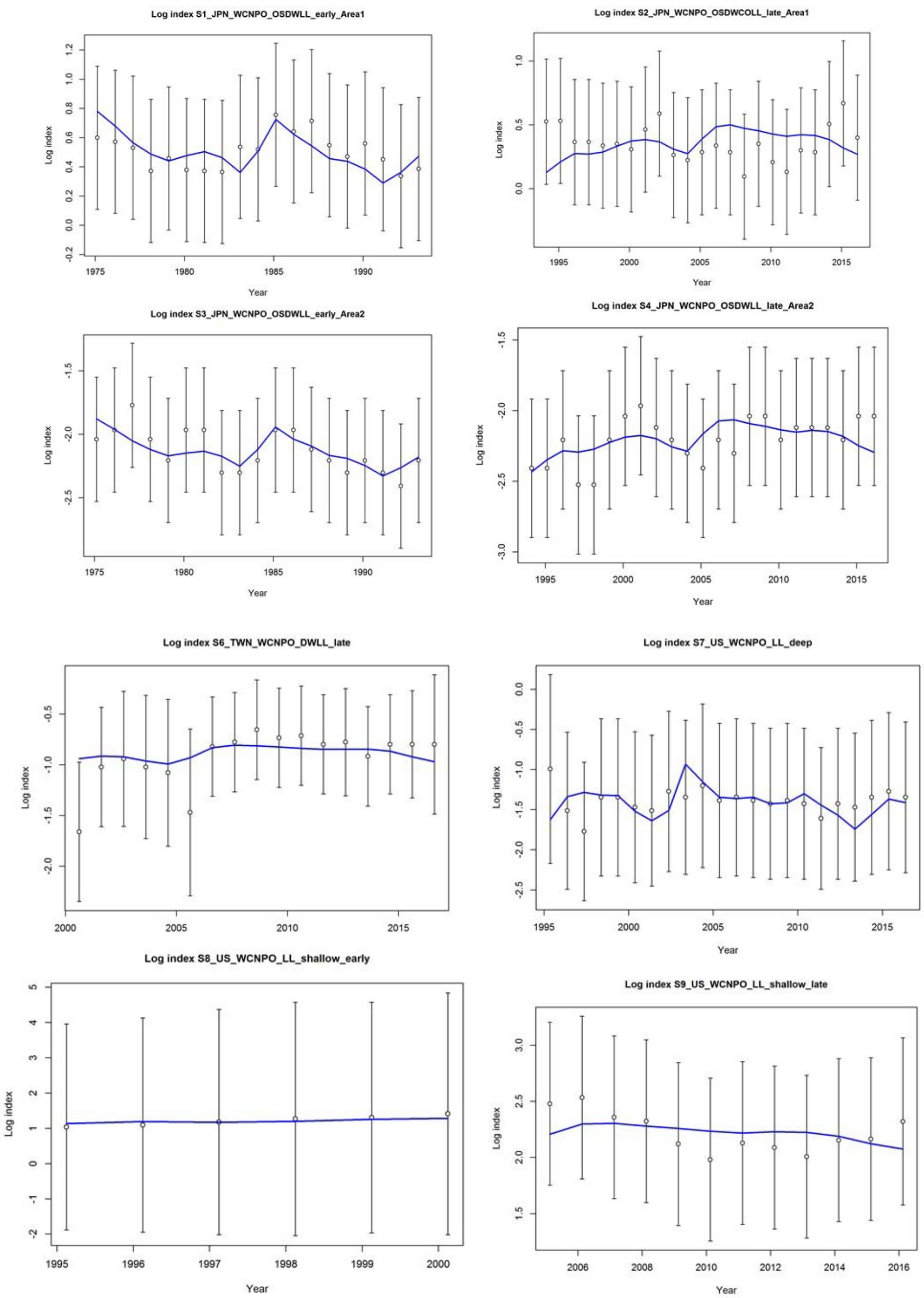

Figure 5. Model-estimated (blue line) versus observed (open circle) log(CPUE) for each index in the initial base case model. Error bars are input log(SE).

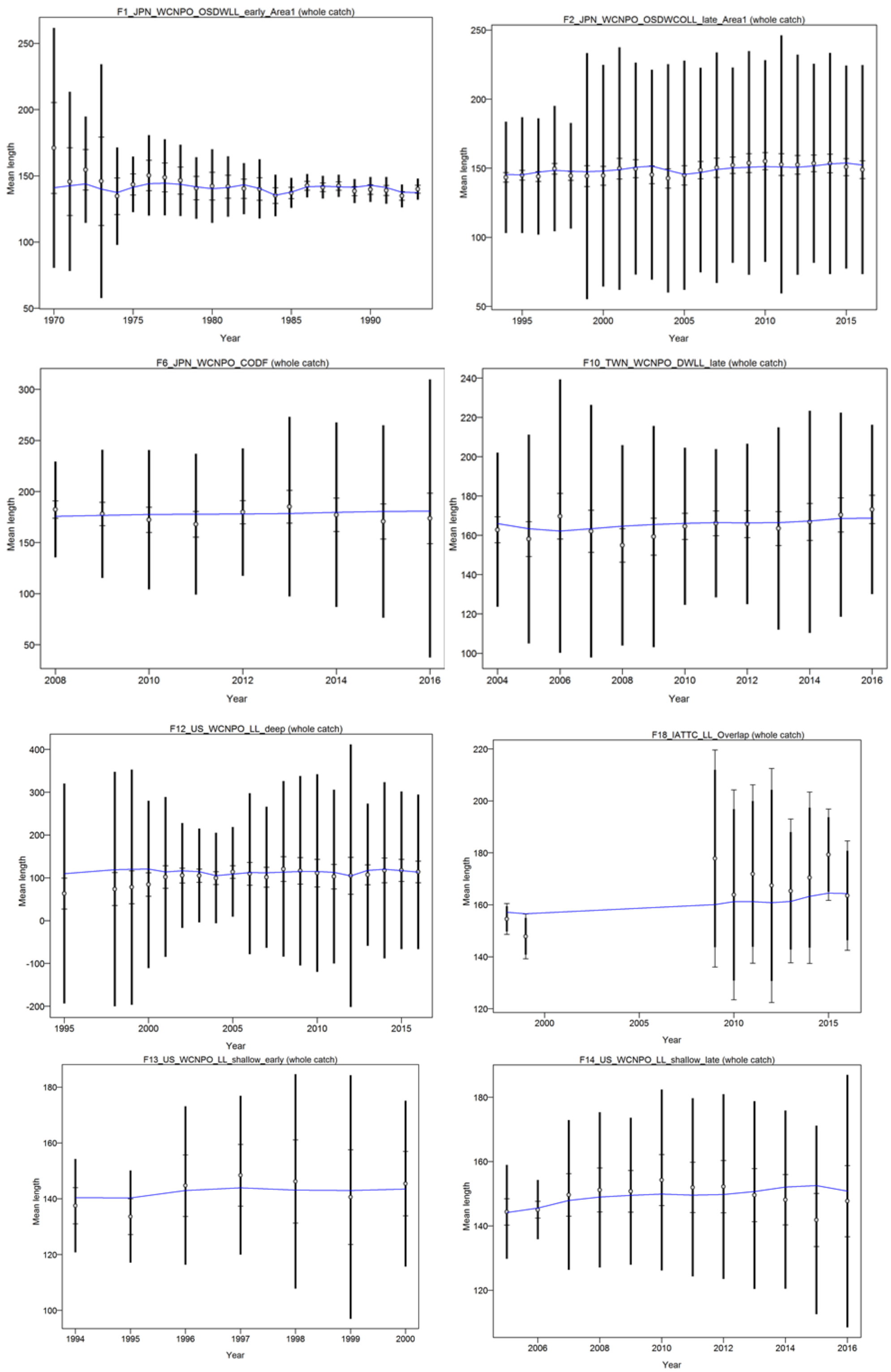

Figure 6. Model-estimated (blue line) and observed (open circle) annual mean length of the length composition data with 95% confidence intervals based upon input sample sizes (thick black lines) in the initial base case model. Thinner black lines (with capped ends) show result

after further adjusting sample sizes based on suggested multiplier (with 95% interval) for Francis (2011) TA1.8.

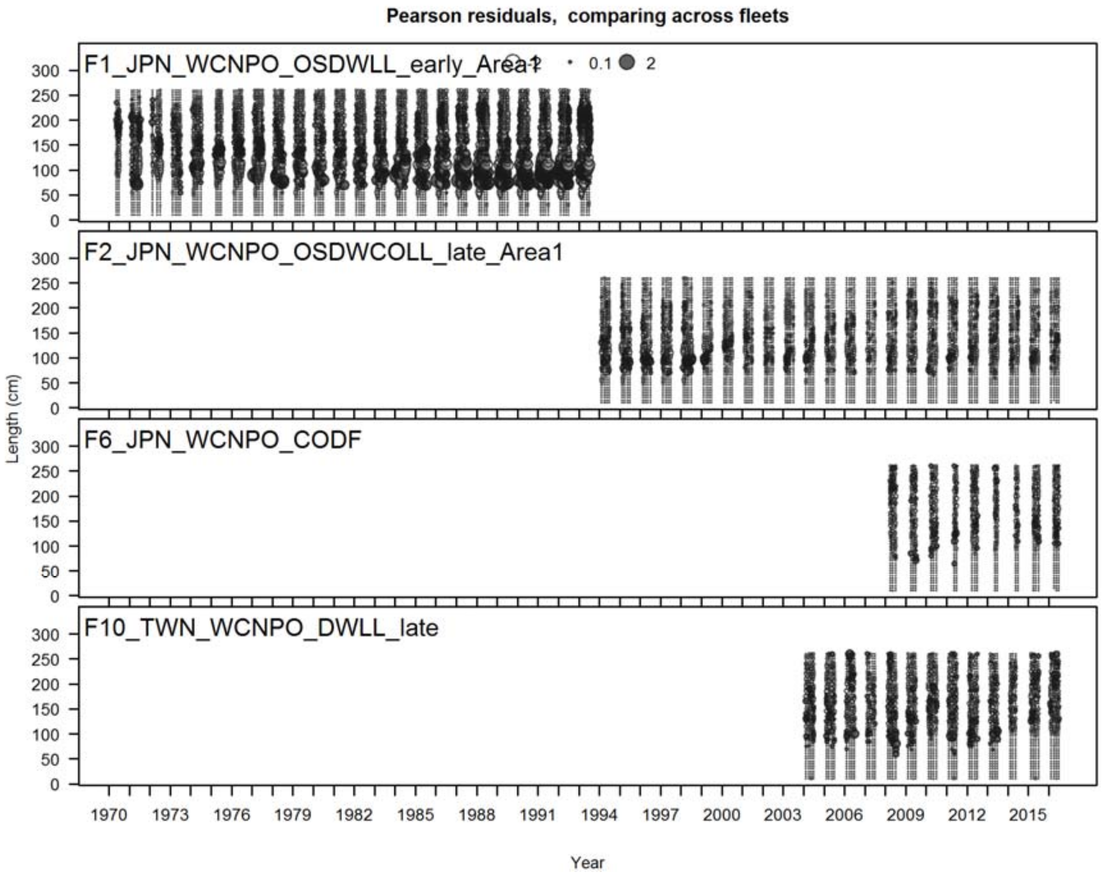

Figure 7. Pearson residuals for length composition fits for each year and quarter compared across fleets in the initial base case model. Closed bubbles are positive residuals (observed > model-estimated) and open bubbles are negative residuals (observed < model-estimated).

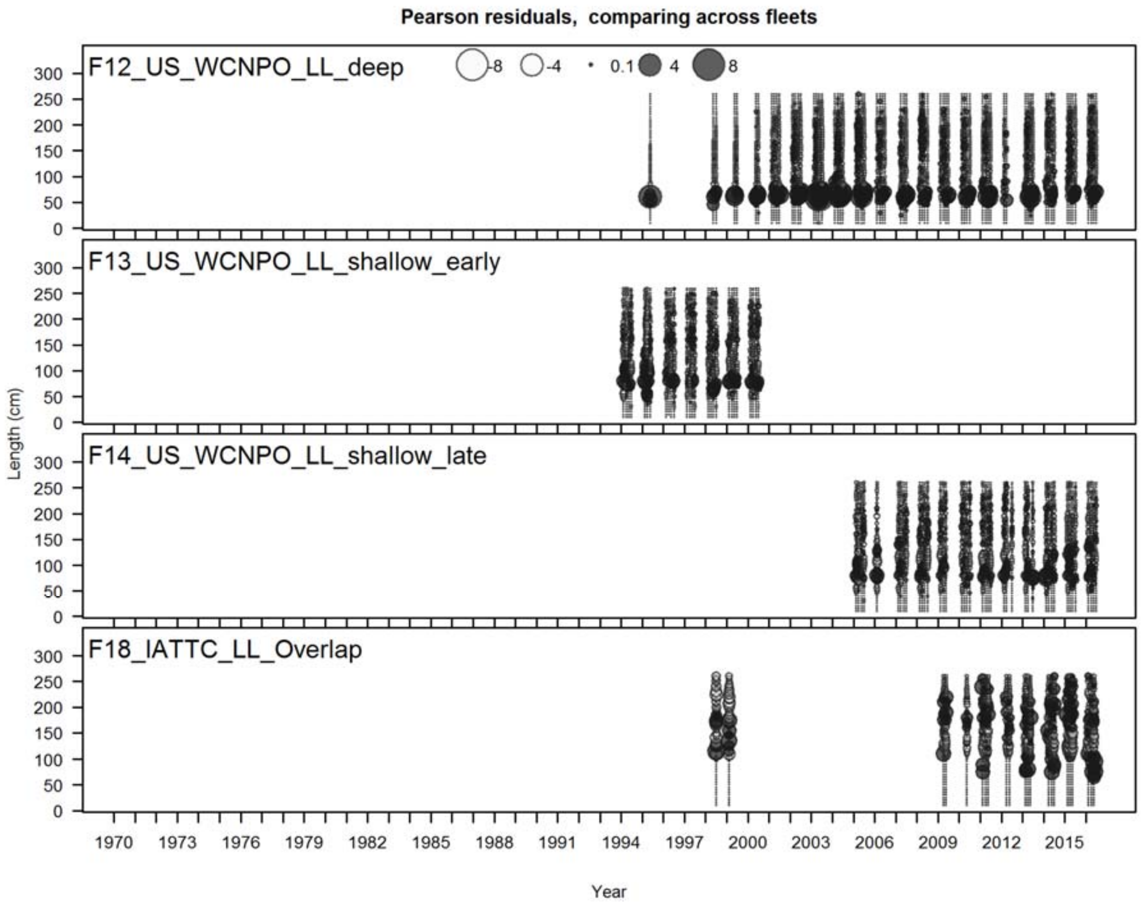

Figure 7. Continued

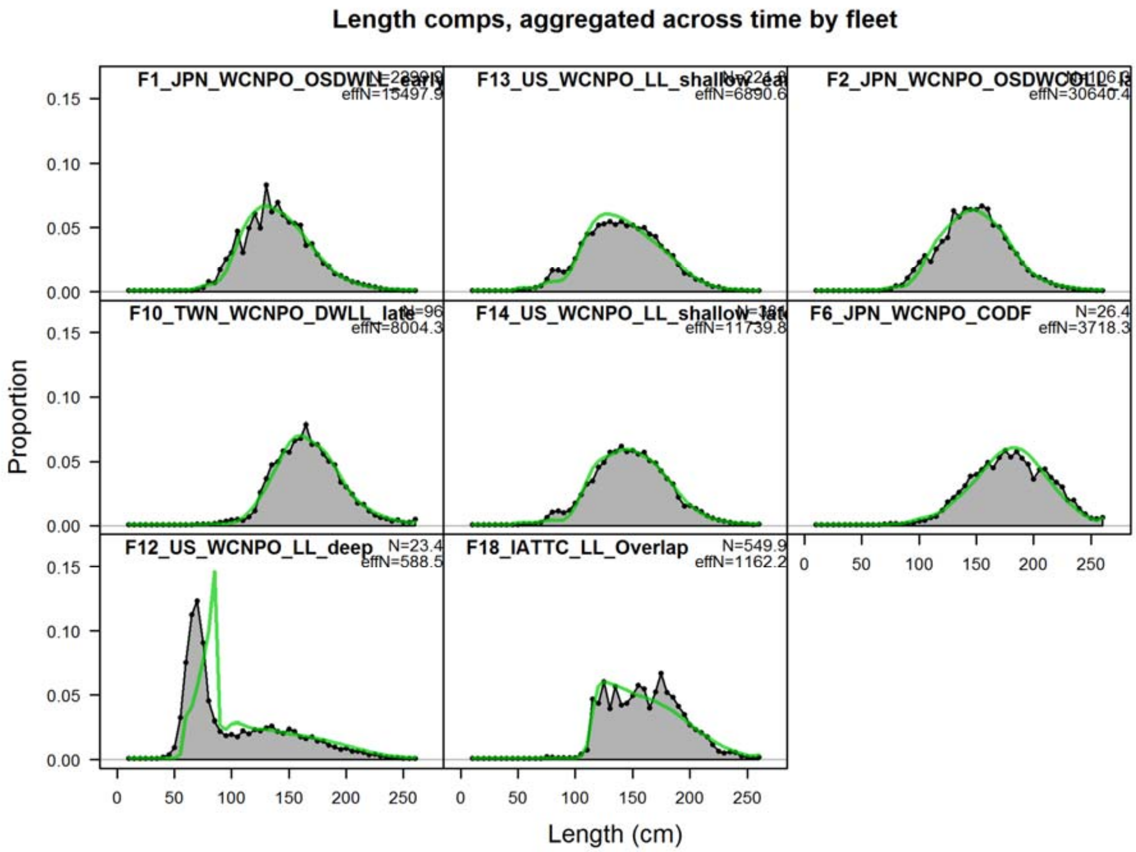

Figure 8. Length composition data observed (black line and grey shading) and initial base case model-estimated selectivity (green line), aggregated across time by fleet.

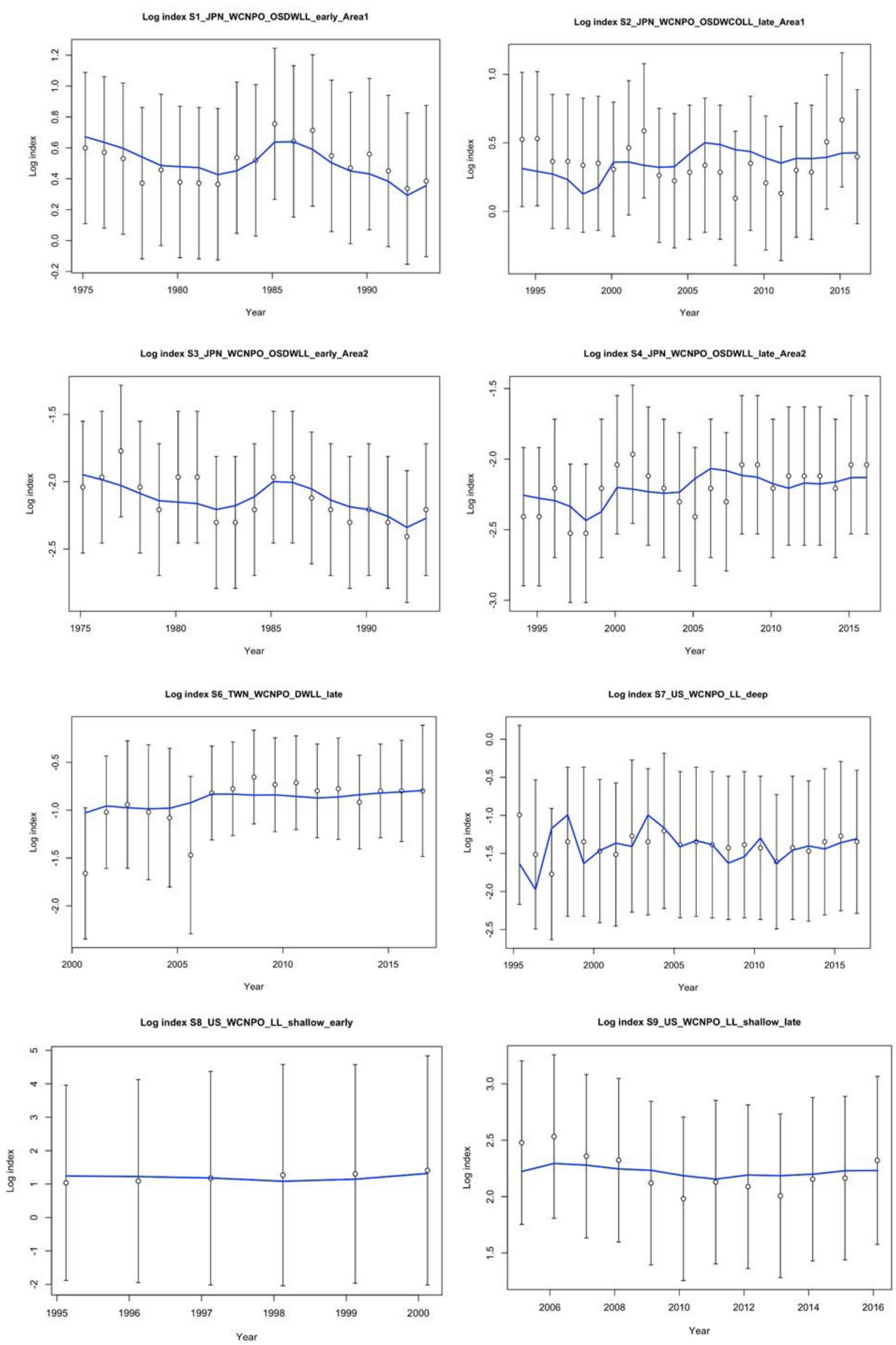

Figure 9. Model-estimated (blue line) versus observed (open circle) log(CPUE) for each index in the final base case model. Error bars are input log(SE).

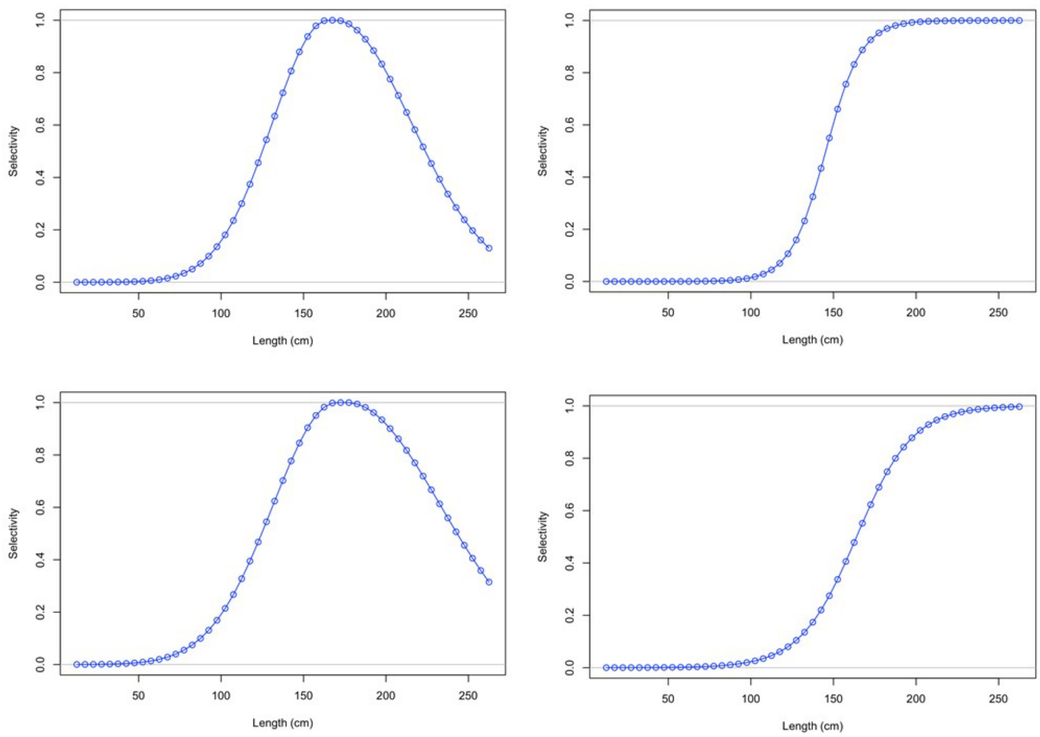

Figure 10. Estimated selectivity at length in the final base case model for F2 Japan LL Late Area 2 (top left), F10 Taiwan LL Late (top right), F14 US HI LL Shallow Late (bottom left), and F18 IATTC LL Overlap (bottom right).

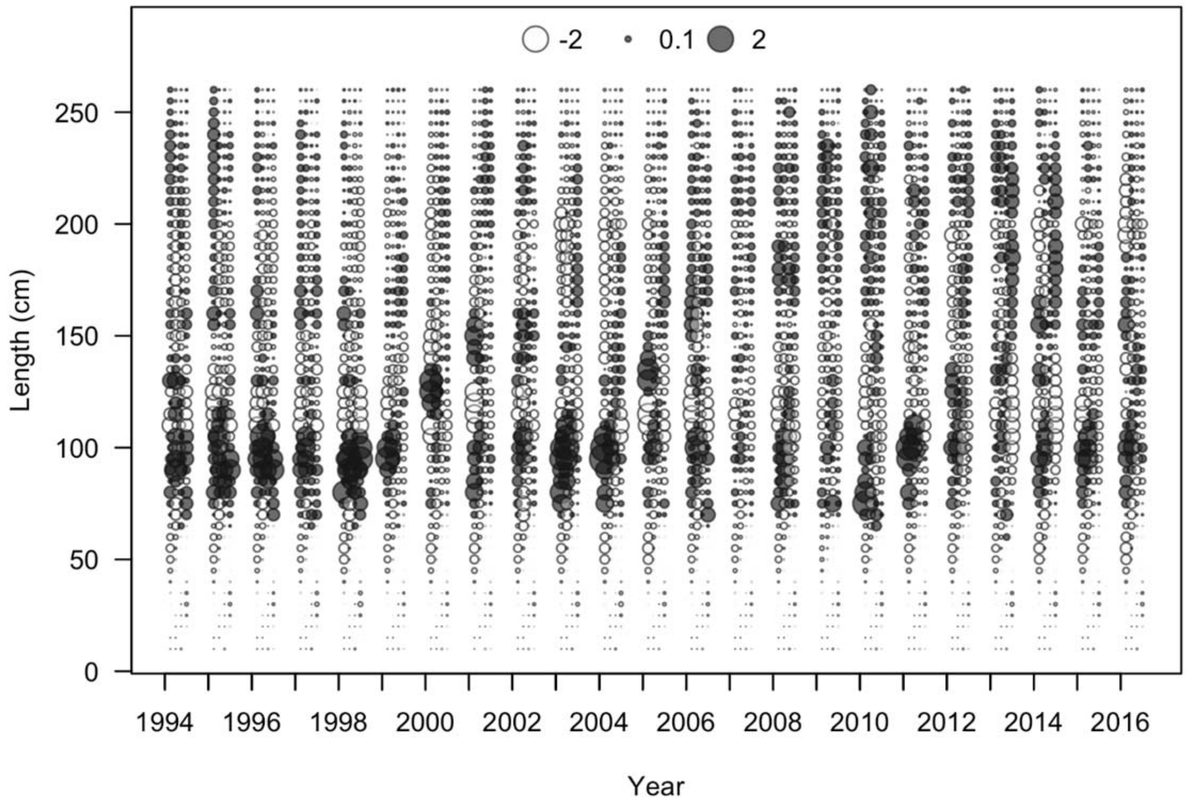

Figure 11. Pearson residuals for length composition fits for each year and quarter for F2 Japan LL Area 1 late in the final base case model. Closed bubbles are positive residuals (observed > model-estimated) and open bubbles are negative residuals (observed < model-estimated).

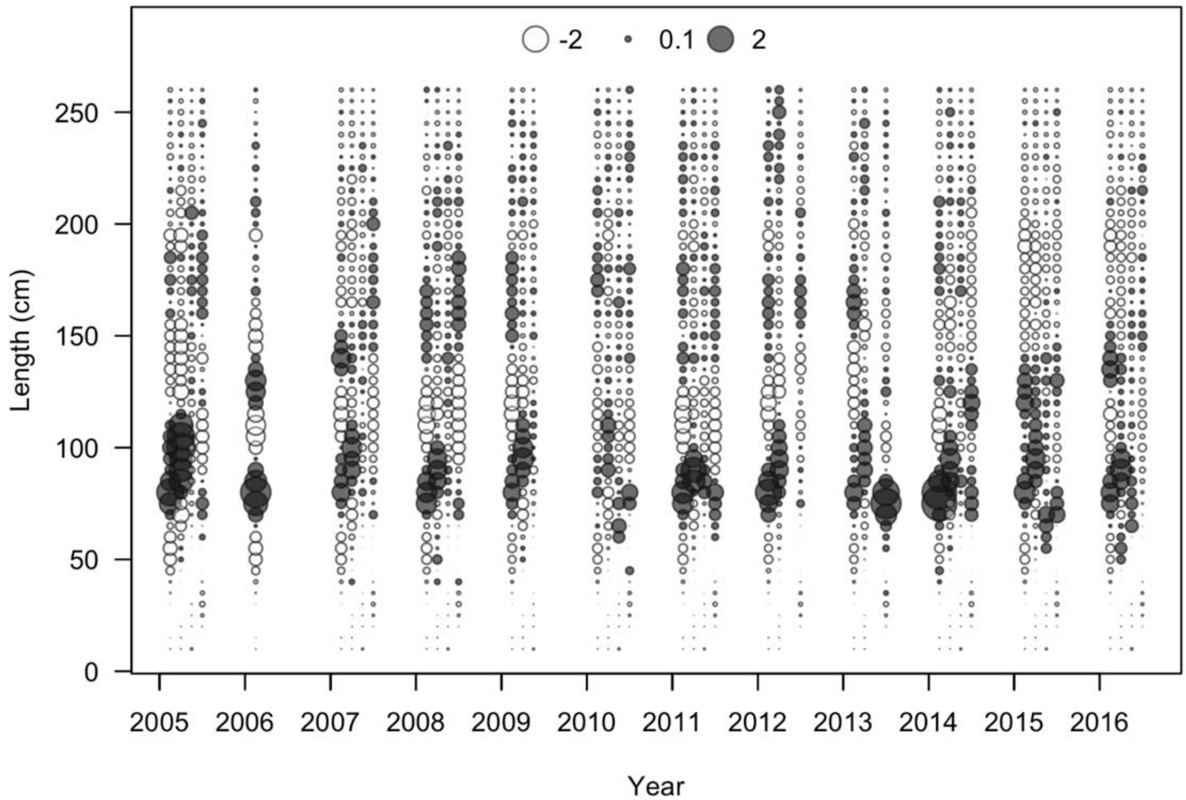

Figure 12. Pearson residuals for length composition fits for each year and quarter for F14 US HI LL Shallow late in the final base case model. Closed bubbles are positive residuals (observed > model-estimated) and open bubbles are negative residuals (observed < model-estimated).

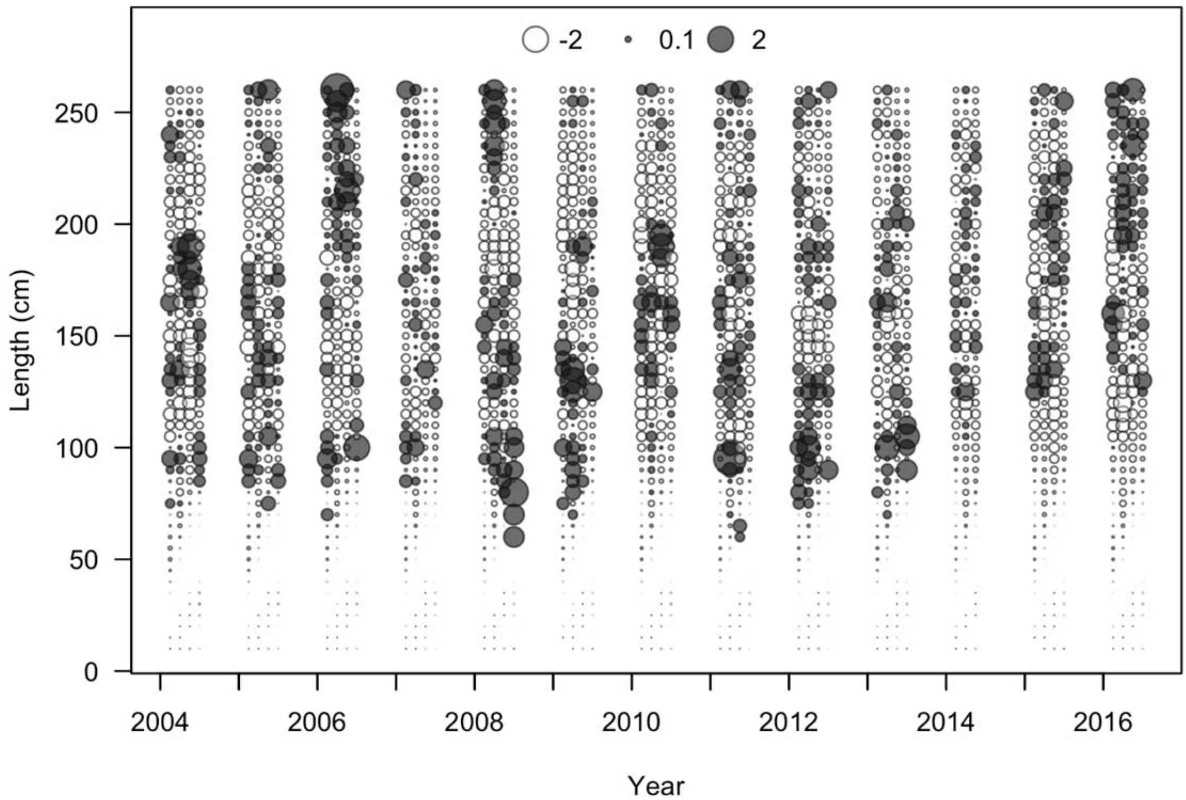

Figure 13. Pearson residuals for length composition fits for each year and quarter for F10 Taiwan LL late in the final base case model. Closed bubbles are positive residuals (observed > model-estimated) and open bubbles are negative residuals (observed < model-estimated).

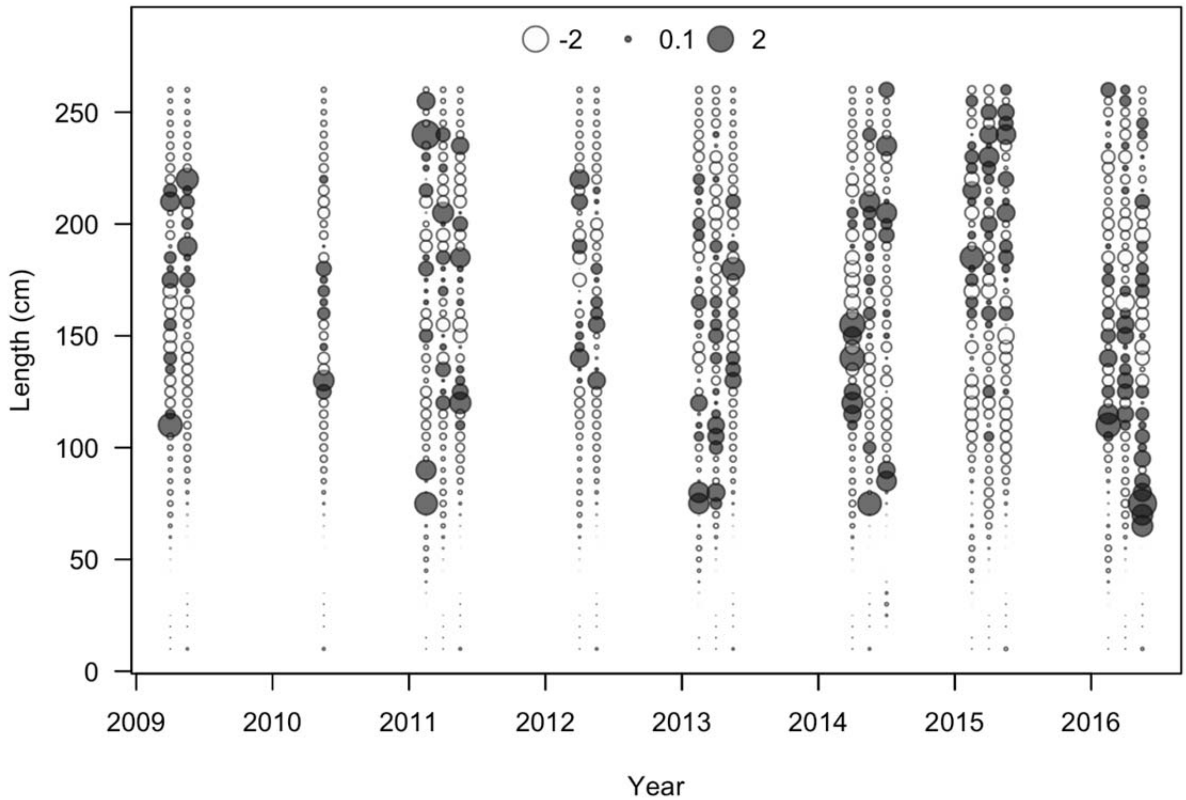

Figure 14. Pearson residuals for length composition fits for each year and quarter for F18 IATTC LL WCNPO overlap in the final base case model. Closed bubbles are positive residuals (observed > model-estimated) and open bubbles are negative residuals (observed < model-estimated).

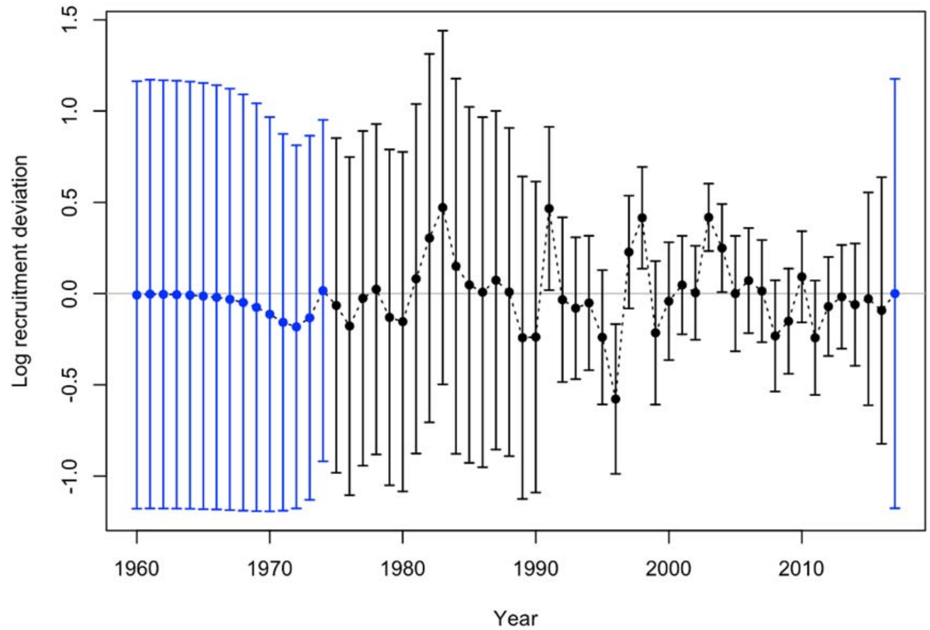

Figure 15. Residual recruitment deviations estimated from the final base case model. Blue circles indicate early recruitment deviations prior to 1975.

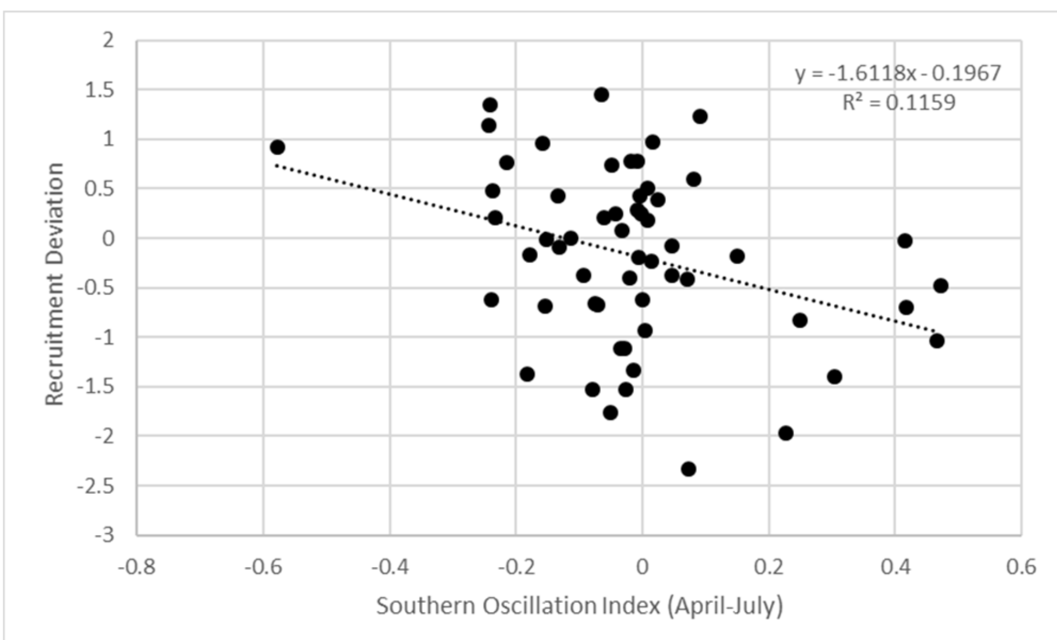

Figure 16. Recruitment deviations vs Southern Oscillation Index (average from April - July). Dotted line indicates linear regression best fit line, R^2 value and equation of the line are in the upper right hand corner.

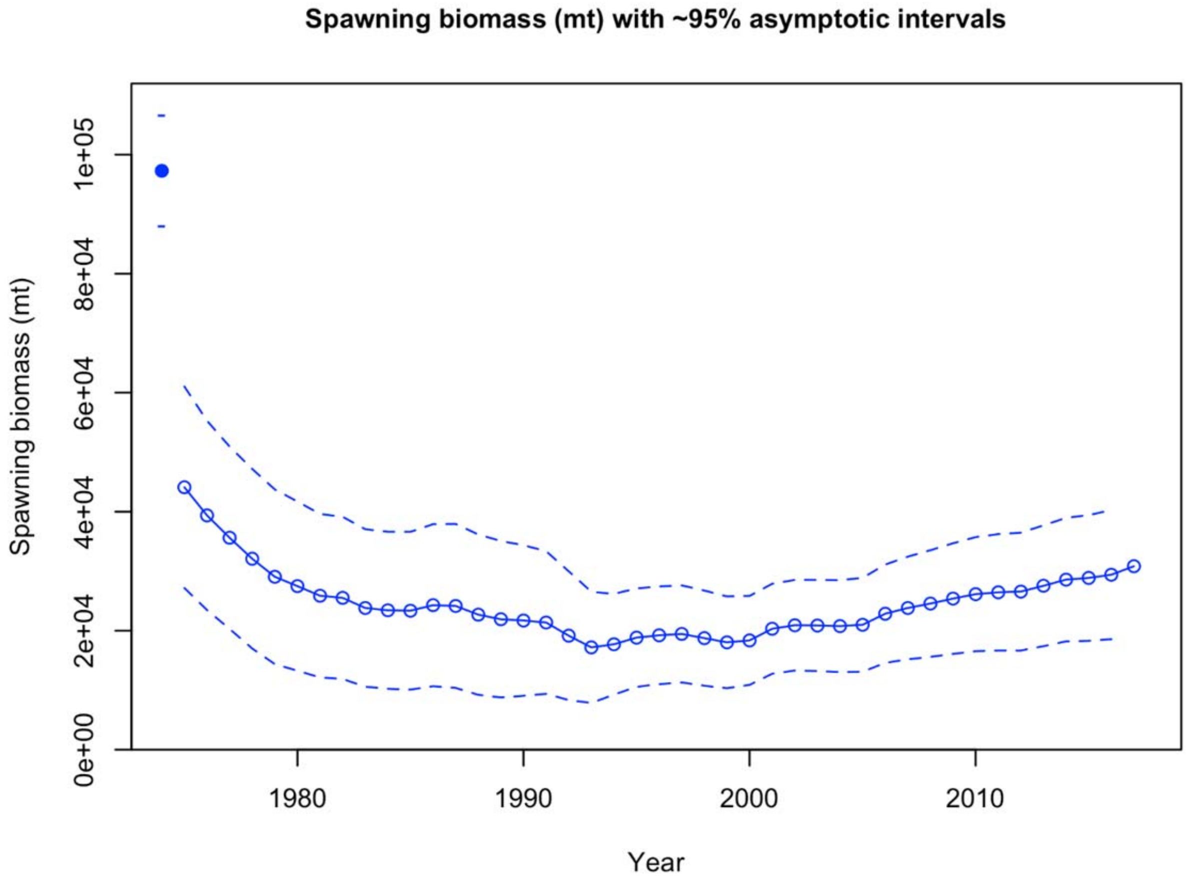

Figure 17. Annual estimates of female spawning stock biomass (open circles) with 95% confidence intervals (dashed lines). Intial female spawning stock biomass is indicated in the upper left corner (closed circle) with 95% confidence intervals (dashes).

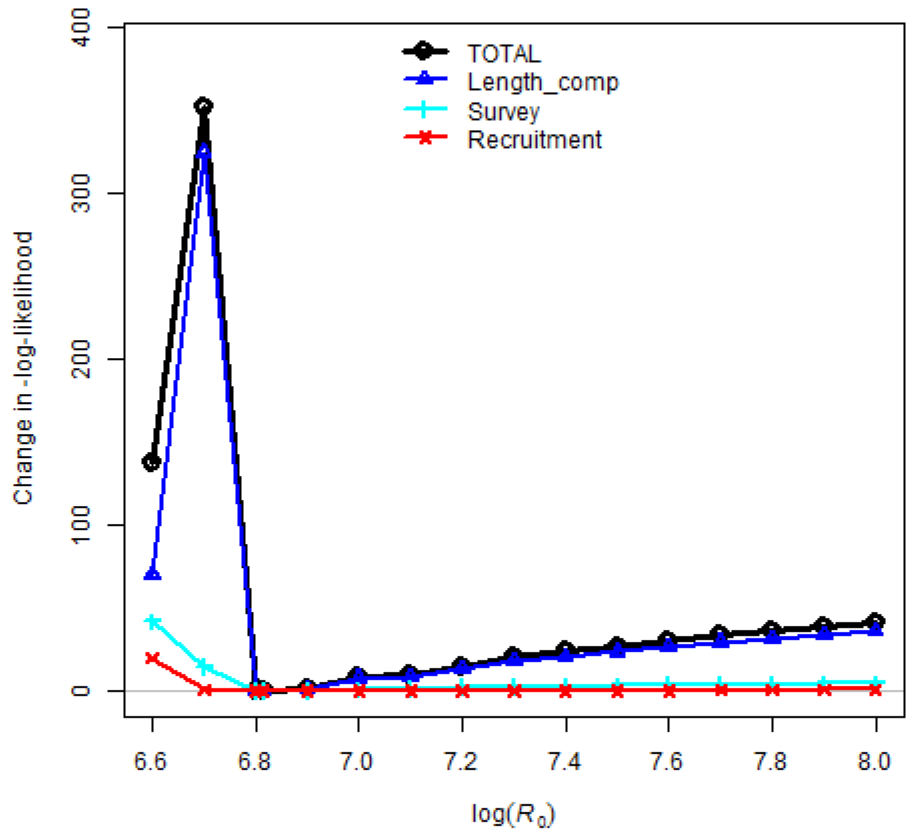

Figure 18. Likelihood profile over log(R_0) for the final base case model: total likelihood (black circles), length composition (blue triangles), survey/CPUE indices (light blue vertical bars), and recruitment index (red x's).

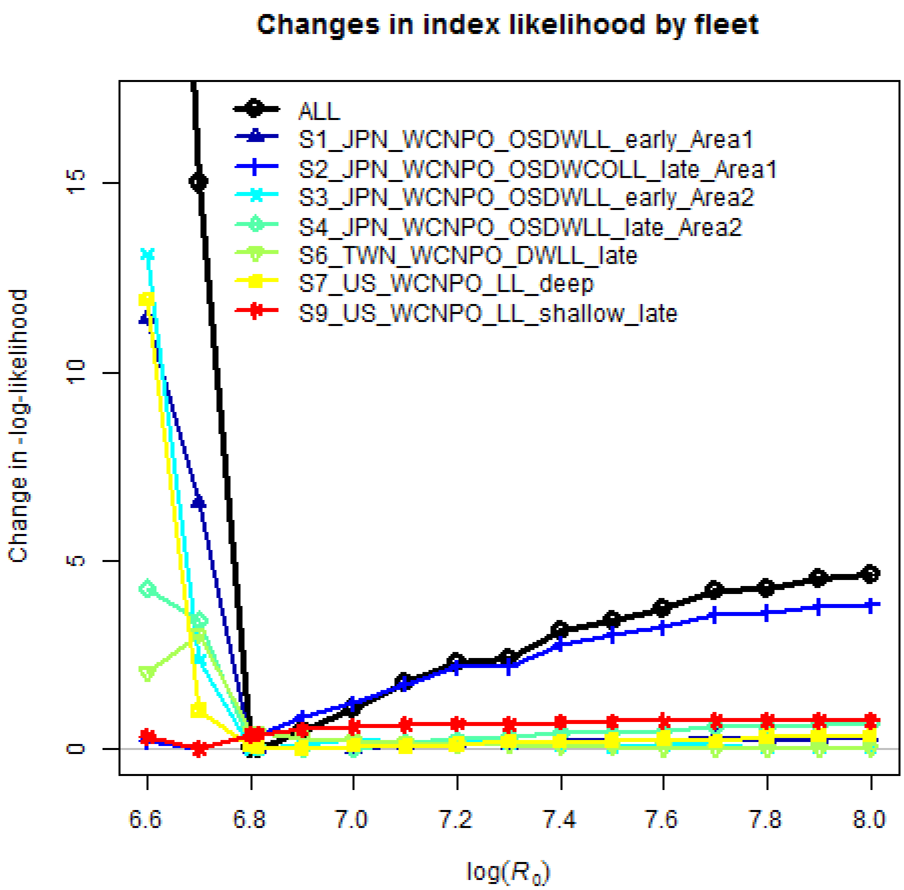

Figure 19. Likelihood profile over log(R_0) by each CPUE likelihood component: All fleets (black open cirlces); S1 (dark blue open triangles); S2 (medium blue crosses); S3 (light blue x); S4 (teal diamonds); S6 (green inverted triangles); S7 (yellow squares); S9 (red astericks). Any CPUE index which contributes to less than 0.01% of the total likelihood component was excluded.

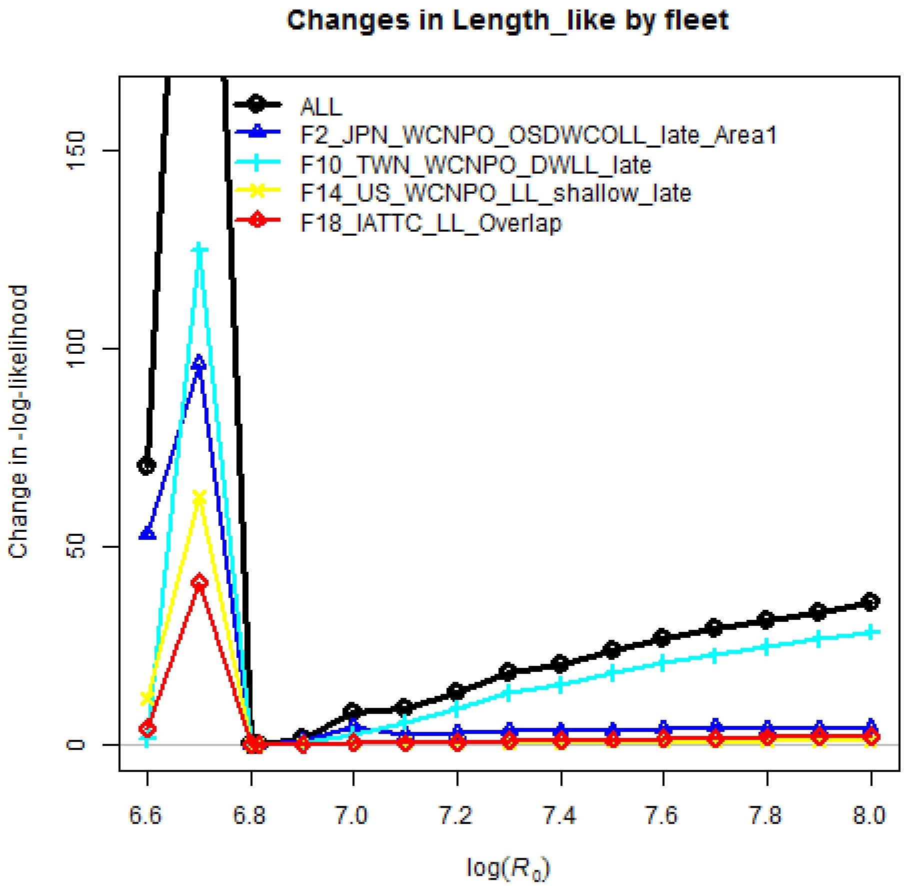

Figure 20. Likelihood profile over log(R_0) by each length composition likelihood component: All fleets (black open cirlces); F2 (dark blue open triangles); F10 (light blue crosses); F14 (yellow squares); F18 (red diamonds).

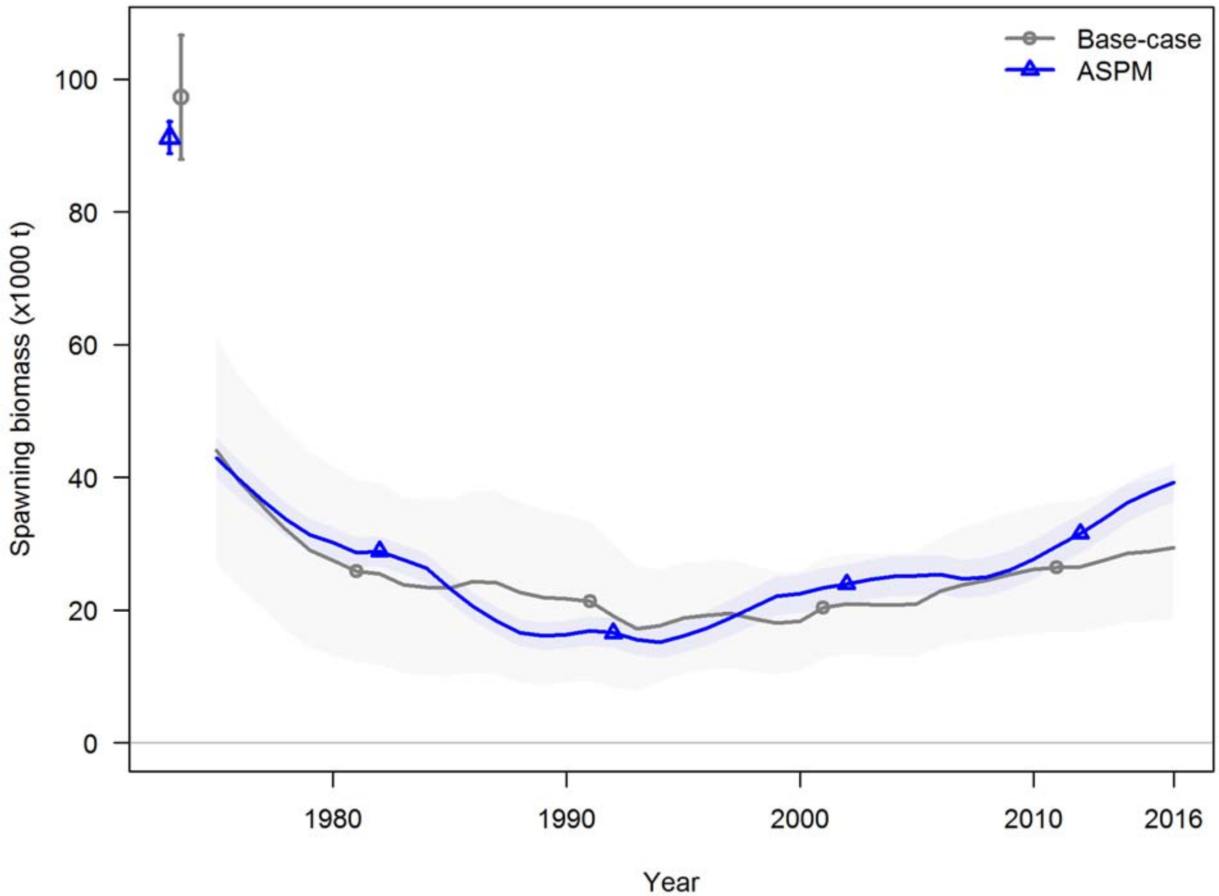

Figure 21. Plot of spawning stock biomass for the base case model (grey open circles) and the ASPM (blue triangles). Shading indicates 95% confidence interval.

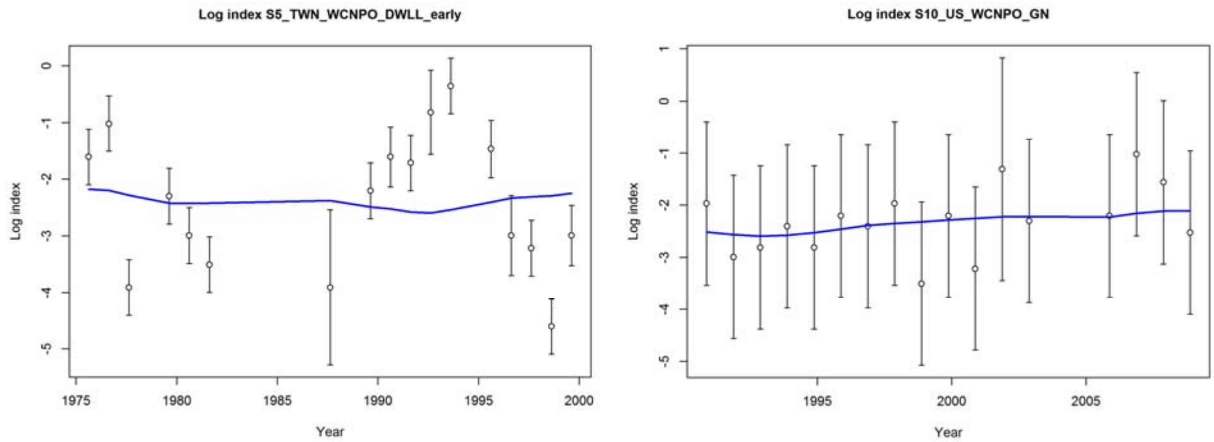

Figure 22. Alternative model 1 fits to the Taiwan LL early (left) and US gillnet (right) indices. Blue line indicates model-estimated log(CPUE), open circles indicate observed log(CPUE) with 95% confidence intervals as black vertical bars.

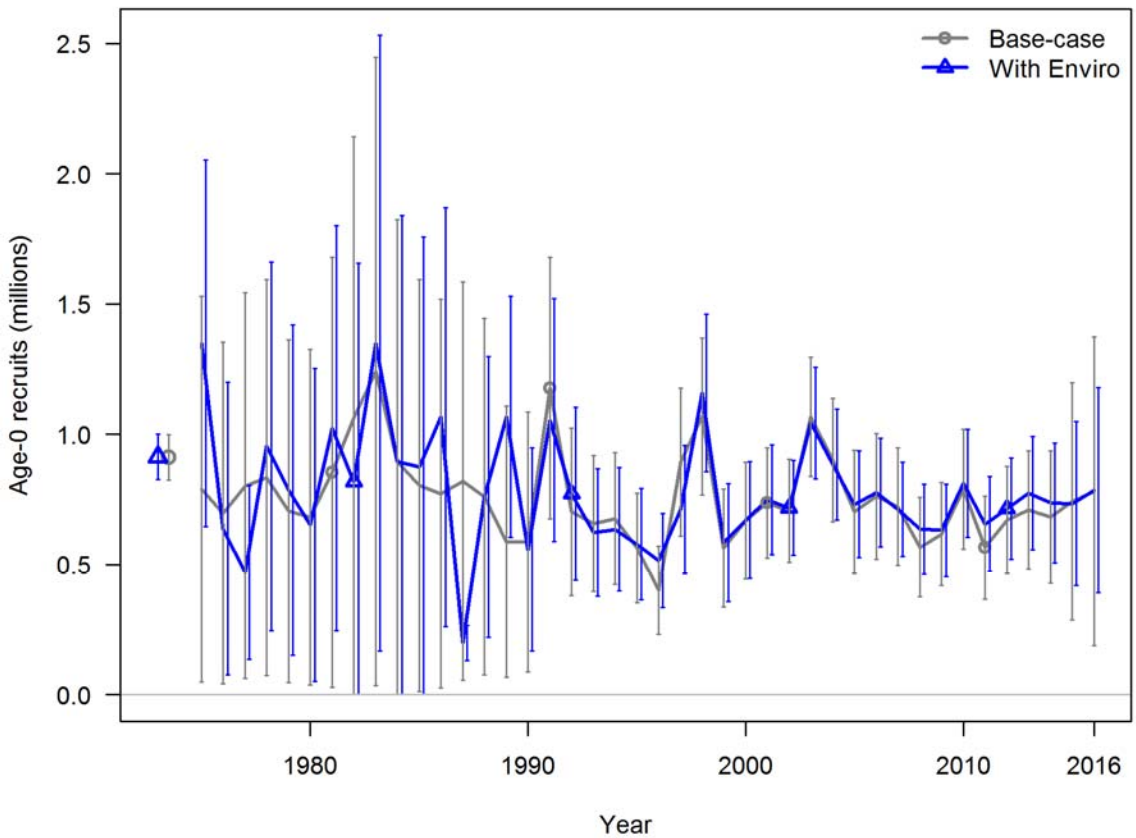

Figure 23. Numbers of age-0 recruits and 95% confidence intervals (bars) estimated from Alternative model 2 with two environmental indices (blue triangles) compared to the final base case model (grey circles).

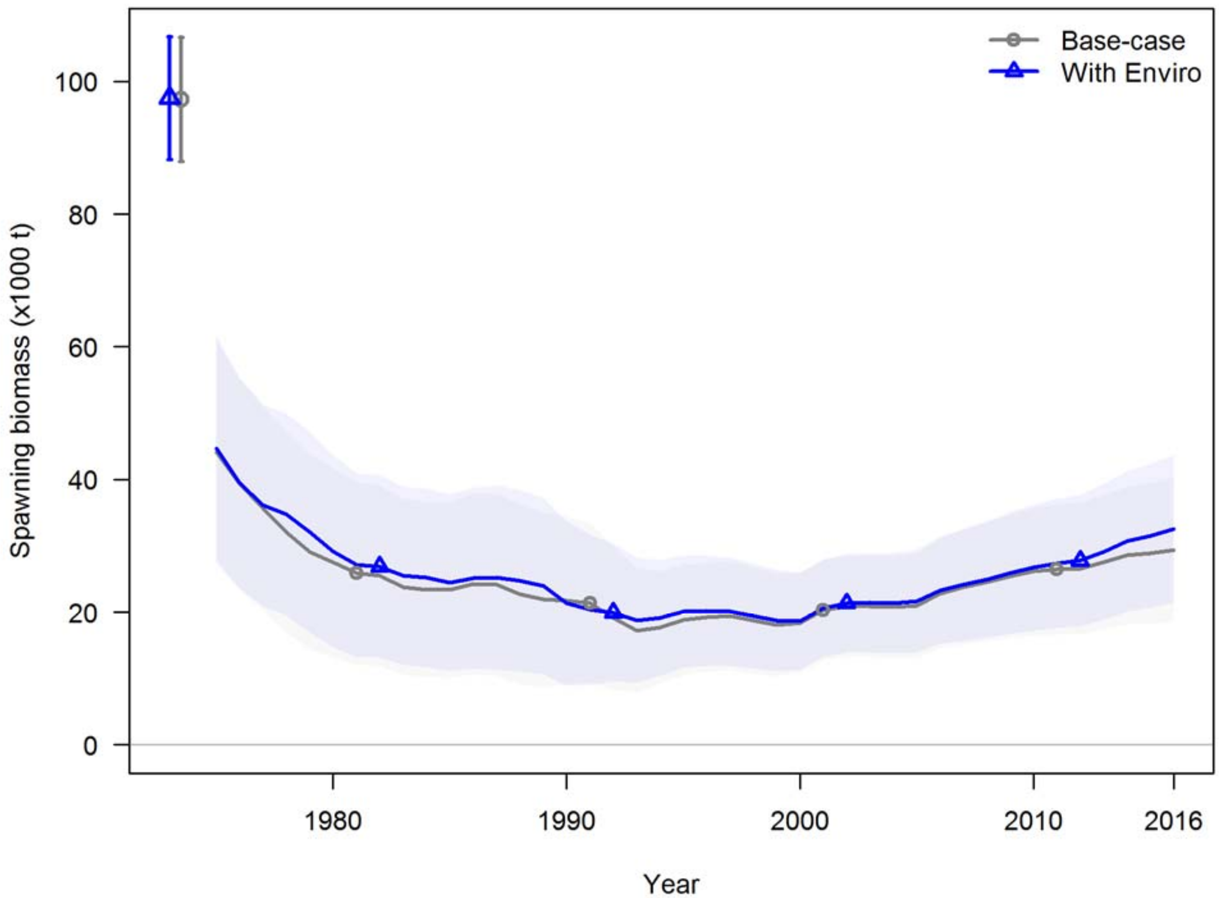

Figure 24. Female Spawning Stock Biomass (1000 mt) for the base case model (blue triangles) and the base case model with environmental covariates (grey open circles). Initial female spawning stock biomass are the first points of each series.

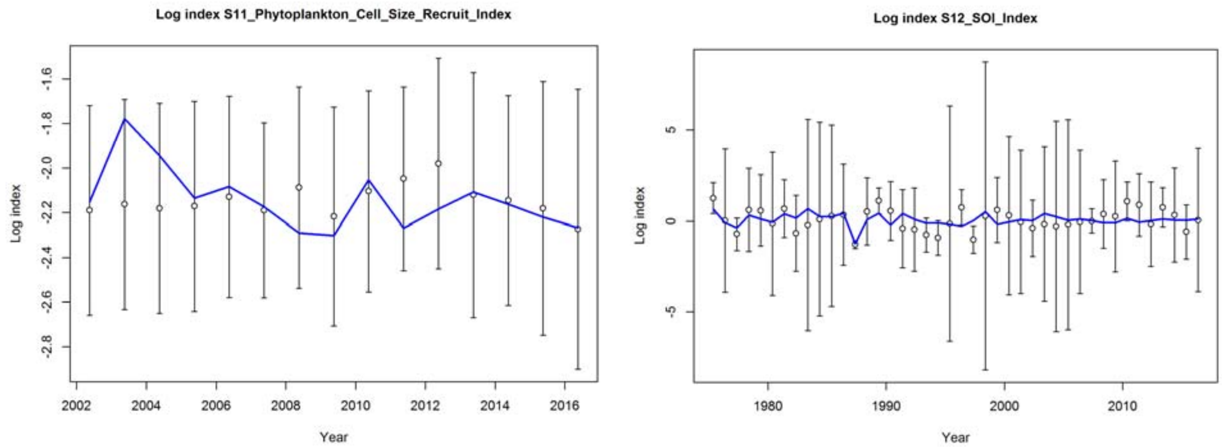

Figure 25. Alternative model 2 fits to the environmental indices phytoplankton cell size (left) and the SOI index (right). Blue indicates model-estimated log recruitment deviations and black open circles represent input log index values with 95% confidence intervals.

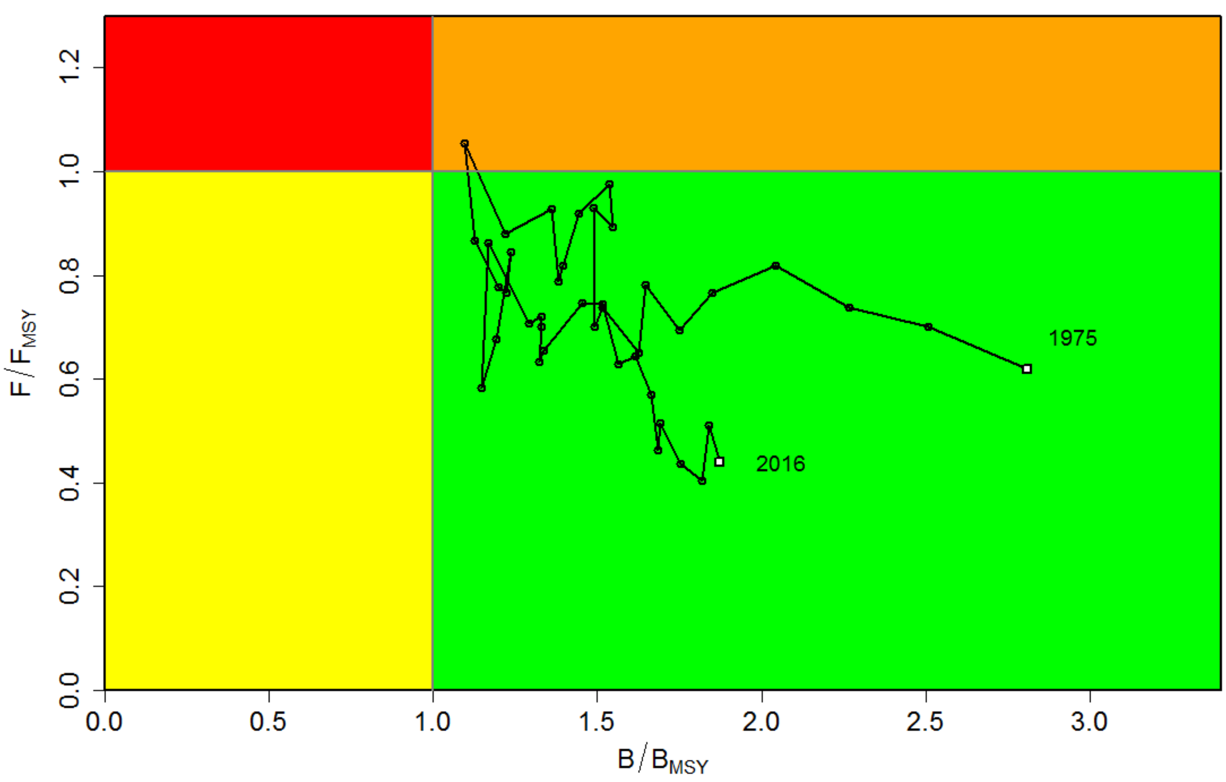

Figure 26. Kobe plot of the trends in estimates of relative fishing mortality (average of age 1-10) and spawning stock biomass for the preliminary 2018 base case Stock Synthesis swordfish assessment model from 1975-2016. White squares indicate beginning (1975) and end (2016) years of the time series