SCIENTIFIC COMMITTEE
SEVENTEENTH REGULAR SESSION

Electronic Meeting
11-19 August 2021

Stock assessment report for Pacific blue marlin (*Makaira Nigricans*) through 2019

WCPFC-SC17-2021/SA-WP-08

ISC[1]

[1] International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean

FINAL

ISC/21/ANNEX/10

ANNEX 10

21st Meeting of the
International Scientific Committee for Tuna
and Tuna-Like Species in the North Pacific Ocean
Held Virtually
July 12-20, 2021

STOCK ASSESSMENT REPORT FOR PACIFIC BLUE MARLIN (*MAKAIRA NIGRICANS*) THROUGH 2019

July 2021

Left Blank for Printing

ABSTRACT

We present the benchmark stock assessment for the Pacific blue marlin (*Makaira nigricans*) stock conducted in 2020 by the ISC Billfish Working Group (BILLWG). The 2021 assessment was based on an ensemble model composed of two plausible Stock Synthesis models with equal model weights. Both models used the best scientific information available on blues marlin catch, abundance index, and length composition data during 1971-2019. Each model used one of two proposed growth curves estimated for the Pacific blue marlin. The model-averaged results indicated that female spawning stock biomass for Pacific blue marlin decreased from 69,000 metric tons in 1971 to 24,000 metric tons in 2019. Estimated fishing mortality gradually increased from 0.08 in 1971 to a high of 0.24 $year^{-1}$ from 2003-2006, and declined to 0.13 $year^{-1}$ (2017-2019). Fishing mortality has only been above F_{MSY} from 2003 to 2006 and has been well below F_{MSY} since 2017. Compared to MSY-based reference points, the current spawning biomass (average for 2017-2019) was 13% above SSB_{MSY} and the current fishing mortality (average for ages 1 – 10 in 2017-2019) was 40% below F_{MSY}. The ensemble model indicated that under current conditions Pacific blue marlin was likely not overfished (81% probability) and was very likely not subject to overfishing (>90% probability) relative to MSY-based reference points.

EXECUTIVE SUMMARY: PACIFIC BLUE MARLIN STOCK ASSESSMENT

Stock Identification and Distribution: The Pacific blue marlin (*Makaira nigricans*) is considered a pan-Pacific stock caught primarily in tropical and sub-tropical waters. All available fishery data from the Inter-American Tropical Tuna Commission (IATTC) and the Western and Central Pacific Fisheries Commission (WCPFC) were used for this benchmark stock assessment. For modeling observations of CPUE and size composition data, it was assumed that there was instantaneous mixing of fish throughout the stock area on a quarterly basis.

Catches: Pacific blue marlin catches increased from the 1970s to the 1990s, and remained high until the 2000s when they started to decline. The relative catch by Japanese fleets has decreased and the relative catch from the Chinese Taipei and other longline fleets have increased since 2000. The catch by other fleets is small by comparison and varies (Figure S1). Overall, longline gear has accounted for the vast majority of Pacific blue marlin catches (67%), Japanese fleets dominating the catch before 2000, and Chinese Taipei and other longline fleets dominating thereafter.

Data and Assessment: Catch and size composition data were collected from three ISC countries (Japan, Chinese Taipei, and the USA), the IATTC, and the WCPFC. Standardized catch-per-unit effort data used to measure trends in relative abundance were provided by Japan, the USA, and Chinese Taipei. Pacific blue marlin was assessed using two-model ensemble of age- and length-structured Stock Synthesis models fit to time series of standardized CPUE and size composition data. The two models in the ensemble differed only in the assumption of the growth curve used. One model used the growth curve from the 2016 Pacific blue marlin assessment (hereafter referred to as the "old growth" model). The other model used a growth curve presented to the working group that was a collaboration between ISC members (hereafter the "new growth" model). The BILLWG noted some substantial differences between the two growth models, including the parameterization (von Bertalanffy vs. two-stanza growth) and the asymptotic length (L_{inf}) for old fish, which was about 50 cm larger for the old growth model. Previous work has demonstrated that stock assessment models can be highly sensitive to the L_{inf} parameter; therefore, the WG explored both models for their ability to describe the input data. Neither model could be discarded based upon model fit and diagnostics; therefore biological reference points, spawning stock biomass, and fishing mortality were averaged between the two models using the multivariate lognormal approximation method assuming equal weights. The value for stock-recruitment steepness used for the base case model was $h = 0.87$. The assessment model was fit to relative abundance indices and size composition data in a likelihood-based statistical framework with smoothing penalties for fishery selectivity. Maximum likelihood estimates of model parameters, derived outputs, and their covariances were used as inputs to the model averaging using the multivariate lognormal approach to characterize stock status and to develop stock projections. Several sensitivity analyses were conducted to evaluate the effects of changes in model parameters, including the natural mortality rate, the stock-recruitment steepness, the growth curve parameters, and the female age at 50% maturity, as well as uncertainty in the input data (i.e. CPUE indices used and the weighting of the size composition data) and model structure (i.e., initial fishing mortality).

Status of Stock: Stock status, biomass trends, and recruitment of Pacific blue marlin (*Makaira nigricans*) for both models in the ensemble hadsimilar trends, although the estimates of initial conditions are different. All reported results are the model-averaged estimates from the ensemble model unless otherwise noted. Estimates of population biomass declined until the mid-2000s, increased again until 2021, and was been relatively flat until the present. The minimum spawning stock biomass is estimated to be 17,592 mt in 2006 (5% above SSB_{MSY}, the spawning stock biomass to produce MSY, 95% C.I. 14,512-20,703 mt, SSB/SS_{MSY} 95% C.I. 0.70-1.01, Figure S2a). In 2019, SSB = 24,272 mt and the relative SSB/SSB_{MSY} = 1.17 (95% C.I. 0.87-1.51). Combined median fishing mortality on the stock (average F on ages 1-10) is currently below F_{MSY} (Figure S2b). It averaged roughly F = 0.13 during 2017-2019, or 40% below F_{MSY}, and in 2019, F=0.11 with a relative fishing mortality of F/F_{MSY} = 0.50 (95% C.I. 0.37-0.69). Median fishing mortality has been below F_{MSY} every year except 2003 to 2006. The predicted value of the spawning potential ratio (SPR, the predicted spawning output at current F as a fraction of unfished spawning output) is currently $SPR_{2017\text{-}2019}$ = 31% for the combined model, which is above the SPR required to produce MSY (17%). Recruitment was relatively consistent throughout the assessment time horizon, with occasional pulses in recruitment, but no notable periods of below-average recruitment. No target or limit reference points have been established for Pacific blue marlin under the auspices of the WCPFC. Blue marlin is expected to be highly productive due to its rapid growth and high resilience to reductions in spawning potential. Although fishing mortality has approached MSY and exceeded MSY from 2003 to 2006, the biomass of the stock has remained above MSY. With continued decreases in Pacific blue marlin catch and fishing effort, the stock is expected to remain within MSY limits. When the status of blue marlin is evaluated relative to MSY-based reference points, the 2019 spawning stock biomass of 24,272 mt is 17% above SSB_{MSY} (20,677 mt, 95% C.I. -13% to +50%) and the 2017-2019 fishing mortality is 50% of F_{MSY} (95% C.I. 37% to 69%). Therefore, relative to MSY-based reference points, overfishing was very likely not occurring (>90% probability) and Pacific blue marlin is likely not overfished (81% probability, Figure S3).

Table S1. Reported catch (mt) used in the stock assessment along with annual model-averaged estimates of female spawning biomass (mt), relative female spawning biomass (SSB/SSB_{MSY}), recruitment (thousands of age-0 fish), fishing mortality (average F, ages 1 – 10), relative fishing mortality (F/F_{MSY}), and spawning potential ratio (SPR) of Pacific blue marlin.

Year	2013	2014	2015	2016	2017	2018	2019	Mean[1]	Min[1]	Max[1]
Reported Catch	22,166	23,741	21,861	22,644	14,443	18,589	16,503	18,873	10,882	26,138
Spawning Biomass	27,707	26,321	25,476	23,693	22,942	23,222	24,279	35,007	17,601	69,331
Relative Spawning Biomass	1.33	1.26	1.22	1.15	1.11	1.12	1.18	1.70	0.84	3.51
Recruitment (thousands of age 0 fish)	960	785	608	862	870	1,399	876	895	502	1,399
Fishing Mortality	0.18	0.19	0.19	0.21	0.13	0.16	0.11	0.16	0.08	0.25
Relative Fishing Mortality	0.81	0.85	0.83	0.95	0.58	0.71	0.50	0.71	0.35	1.11
Spawning Potential Ratio	0.26	0.24	0.25	0.22	0.33	0.27	0.34	0.33	0.17	0.60

[1] During 1971-2019

Biological Reference Points: Biological reference points were computed for the combined ensemble model using a multivariate lognormal approximation that accounts for the inherent covariance between F/F_{MSY} and SSB/SSB_{MSY} (Table S2). The combined estimate of the spawning biomass to produce MSY (adult female biomass) was SSB_{MSY} = 20,677 mt. The point estimate of F_{MSY}, the fishing mortality rate to produce MSY (average fishing mortality on ages 1 – 10) was F_{MSY} = 0.23 and the corresponding equilibrium value of spawning potential ratio at MSY was SPR_{MSY} = 17%.

Projections: Deterministic stock projections were conducted with Stock Synthesis to evaluate the impact of alternative future levels of harvest intensity on female spawning stock biomass, fishing mortality, and yield for Pacific blue marlin. Future recruitment was predicted based on the stock-recruitment curve. These projections used all the multi-fleet, multi-season, size- and age-selectivity, and complexity in the assessment model to produce consistent results. The stock projections started in 2020 and continued through 2029 (10 years) under 4 levels of constant fishing mortality: (1) constant fishing mortality equal to the 2003-2005 average ($F2003{-}2005$); (2) constant fishing mortality equal to F_{MSY}; (3) constant fishing mortality equal to the 2016-2018 average defined as current; and (4) constant fishing mortality equal to $F30\%$ (F30% corresponds to the fishing mortality that produces 30% of the spawning potential ratio). Stock projections for each F scenario were run for both growth models in the ensemble and combined using the multivariate lognormal method. Using the deterministic projection result, the multivariate lognormal approximation was applied to generate 10,000 trajectories of SSB and F to calculate the model-averaged results of the new and old growth models. Results showed the projected female spawning stock biomasses, fishing mortality, and the catch biomasses under each of the combined scenarios (Table S3 and Figure S4).

Conservation information: The Pacific blue marlin stock has produced annual yields of around 18,800 mt per year since 2015, or about 90% of the MSY catch. Blue marlin stock status from the ensemble model indicates that the current median spawning biomass is above SSB_{MSY} and that the current median fishing mortality is below F_{MSY}. However, uncertainty in the stock status indicates a 19% chance of Pacific blue marlin being overfished. Both the old and new growth models show evidence of spawning biomass being above SSB_{MSY} and fishing mortality being below F_{MSY} during the last 5 years. Catch biomass has been declining for the last 5 years, and therefore the stock has a low risk of experiencing overfishing or being overfished unless fishing mortality increases to above F_{MSY} based upon stock projections. However, it is also important to note that retrospective analyses show that the assessment model tends to overestimate biomass and underestimate fishing mortality in recent years, in part due to rapid changes in longline CPUE in recent years.

Special Comments: The BILLWG achieved an ensemble model using the best available data and biological information. However, the BILLWG recognized that there is considerable uncertainty in input CPUE data in the recent years and life history parameters, especially growth. The BILLWG considered an extensive suite of model formulations and associated diagnostics for developing the assessment models. Overall, the BILLWG found issues with the new growth and old growth model diagnostics and sensitivity runs that indicated some data conflicts exist, but none of the model diagnostics suggested that the model results were invalid. To improve the stock assessment, the BILLWG also recommends continuing model development work, to reduce data conflicts and modeling uncertainties, reevaluating and improving input assessment data, and for ISC countries to participate in the International Billfish Biological Sampling program to improve estimates of life-history parameters.

Table S2. Estimates of biological reference points along with estimates of fishing mortality (F), spawning stock biomass (SSB), recent average yield (C), and spawning potential ratio (SPR) of Pacific blue marlin, derived from the assessment ensemble model, where "MSY" indicates reference points based on maximum sustainable yield.

Reference Point	Estimate
F_{MSY} (age 1-10)	0.23
F_{2019} (age 1-10)	0.11
$F_{20\%SSB0}$	0.18
SSB_{MSY}	20,677 mt
SSB_{2019}	24,241 mt
$SSB_{20\%SSB0}$	20,729 mt
MSY	24,600 mt
$C_{2017-2019}$	16,512 mt
SPR_{MSY}	17%
SPR_{2019}	34%
$SPR_{20\%SSB0}$	23%

Table S3. Projected median values of Pacific blue marlin spawning stock biomass (SSB, mt) and catch (mt) under four constant fishing mortality rate (F) scenarios during 2020-2029.

Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Scenario 1: F = $F_{2003-2005}$										
SSB	25,459	23,462	21,752	20,498	19,262	18,689	18,252	17,835	17,583	17,475
Catch	33,111	30,527	28,638	27,331	26,431	25,806	25,363	25,044	24,811	24,641
Scenario 2: F = F_{MSY}										
SSB	25,318	23,351	21,583	20,255	19,216	18,405	18,186	17,809	17,513	17,466
Catch	32,875	30,436	28,662	27,439	26,606	26,037	25,645	25,370	25,177	25,039
Scenario 3: F = $F_{2016-2018}$										
SSB	26,930	28,182	28,764	28,675	28,428	28,731	28,052	28,142	27,861	28,081
Catch	23,321	23,546	23,591	23,561	23,513	23,472	23,443	23,422	23,407	23,397
Scenario 4: F = $F_{30\%}$										
SSB	27,757	30,064	30,624	30,976	31,072	31,624	31,415	31,800	31,753	32,132
Catch	20,828	21,404	21,764	22,001	22,167	22,294	22,393	22,471	22,532	22,580

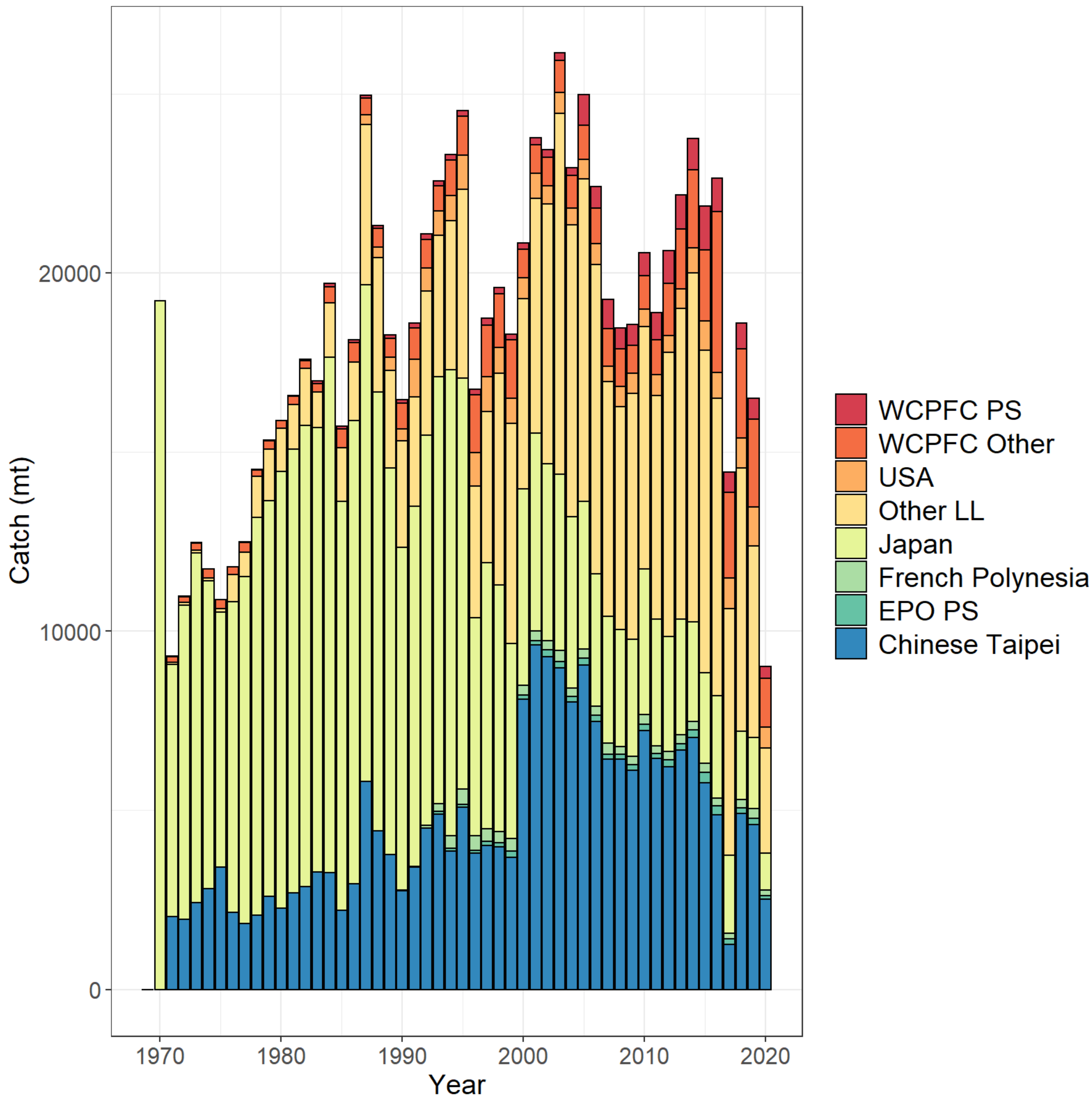


Figure S1. Annual catch biomass (mt) of Pacific blue marlin (*Makaira nigricans*) by country for Japan, Chinese Taipei, the U.S.A., and all other countries during 1975-2019.

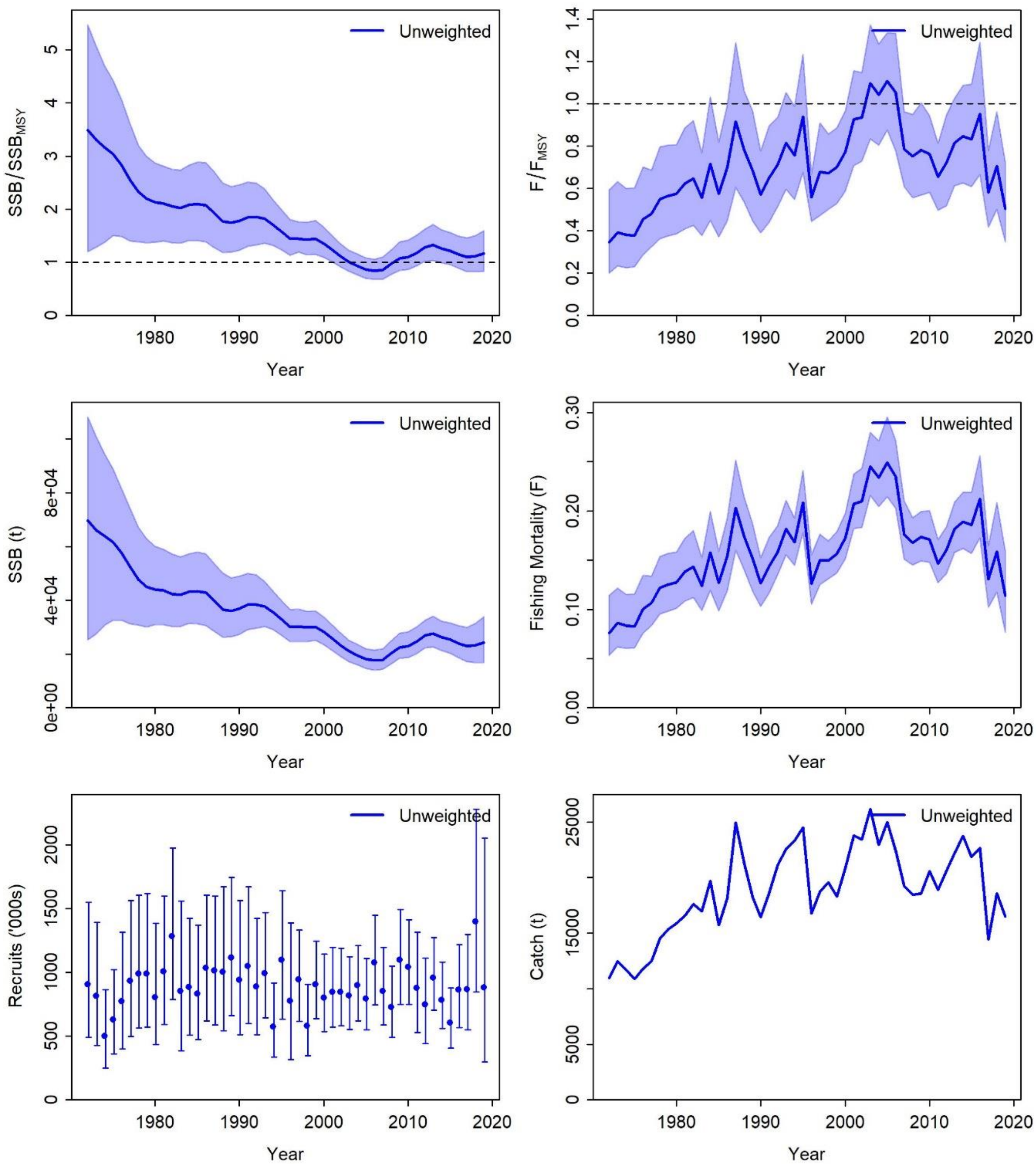


Figure S2. Time series of estimates of female spawning stock biomass over female spawning stock biomass at MSY (top left), fishing mortality over fishing mortality at MSY (top right), spawning stock biomass (center left), instantaneous fishing mortality (ages 1-10 year^{-1}, center right), recruitment (age-0 fish, bottom left), and catch (bottom right) for Pacific blue marlin (*Makaira nigricans*) derived from the 2021 stock assessment model ensemble. Lines (or points for recruitment) indicate the median value estimated from the joint multivariate delta-lognormal estimation, shaded areas (or error bars for recruitment) indicate the 95% confidence intervals. Unweighted indicates that both models have equal weights in the ensemble.

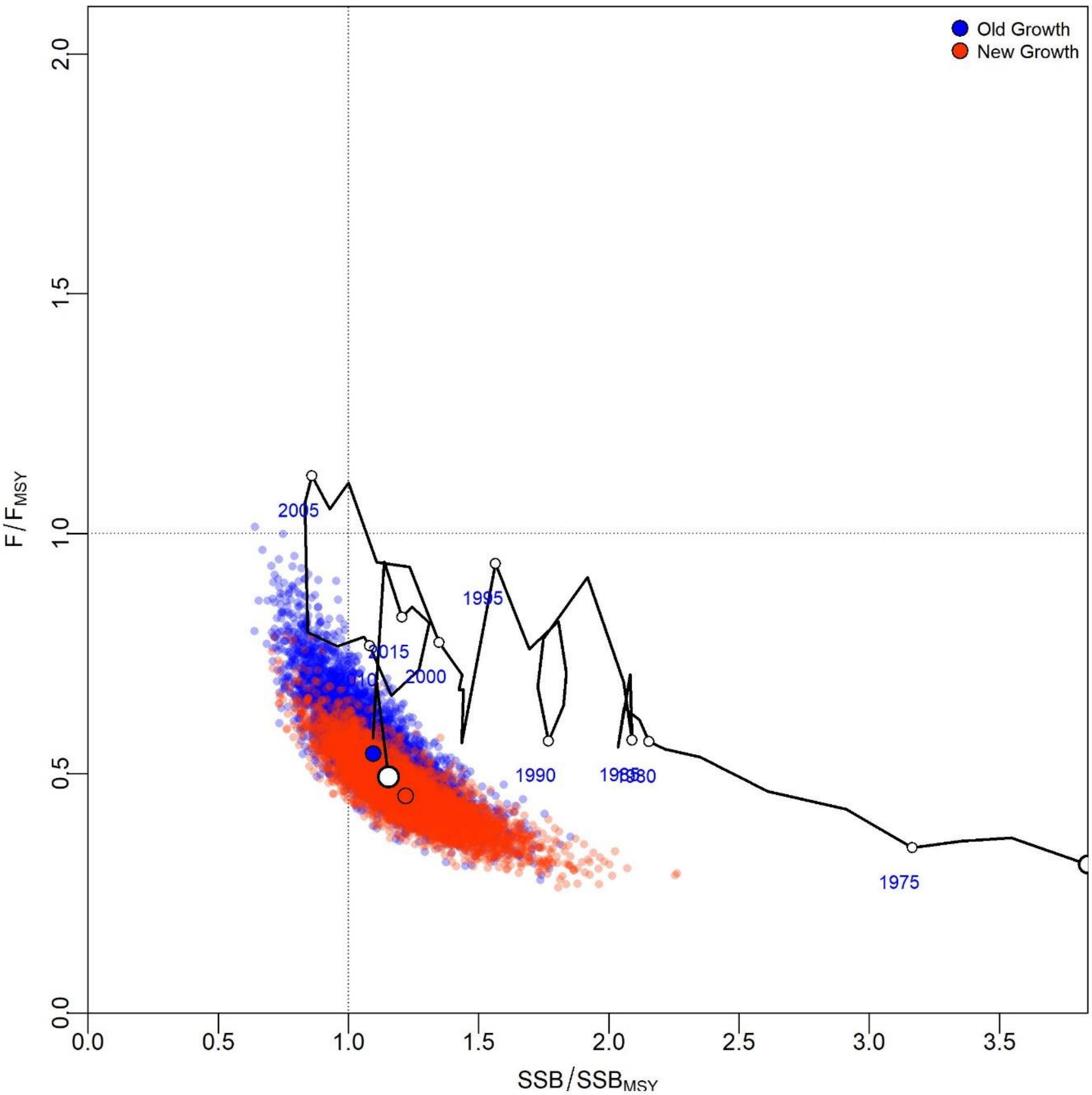


Figure S3. Kobe plot of the time series of estimates of relative fishing mortality (average of age 1-10) and relative spawning stock biomass of Pacific blue marlin (*Makaira nigricans*) during 1971-2019. The white circle denotes the delta-lognormal multivariate estimate of the ensemble model in 2019, blue dots indicate the final year stock status of the old growth model with the 10,000 multivariate draws, and red dots indicate the final year stock status of the new growth model with the 10,000 multivariate draws.

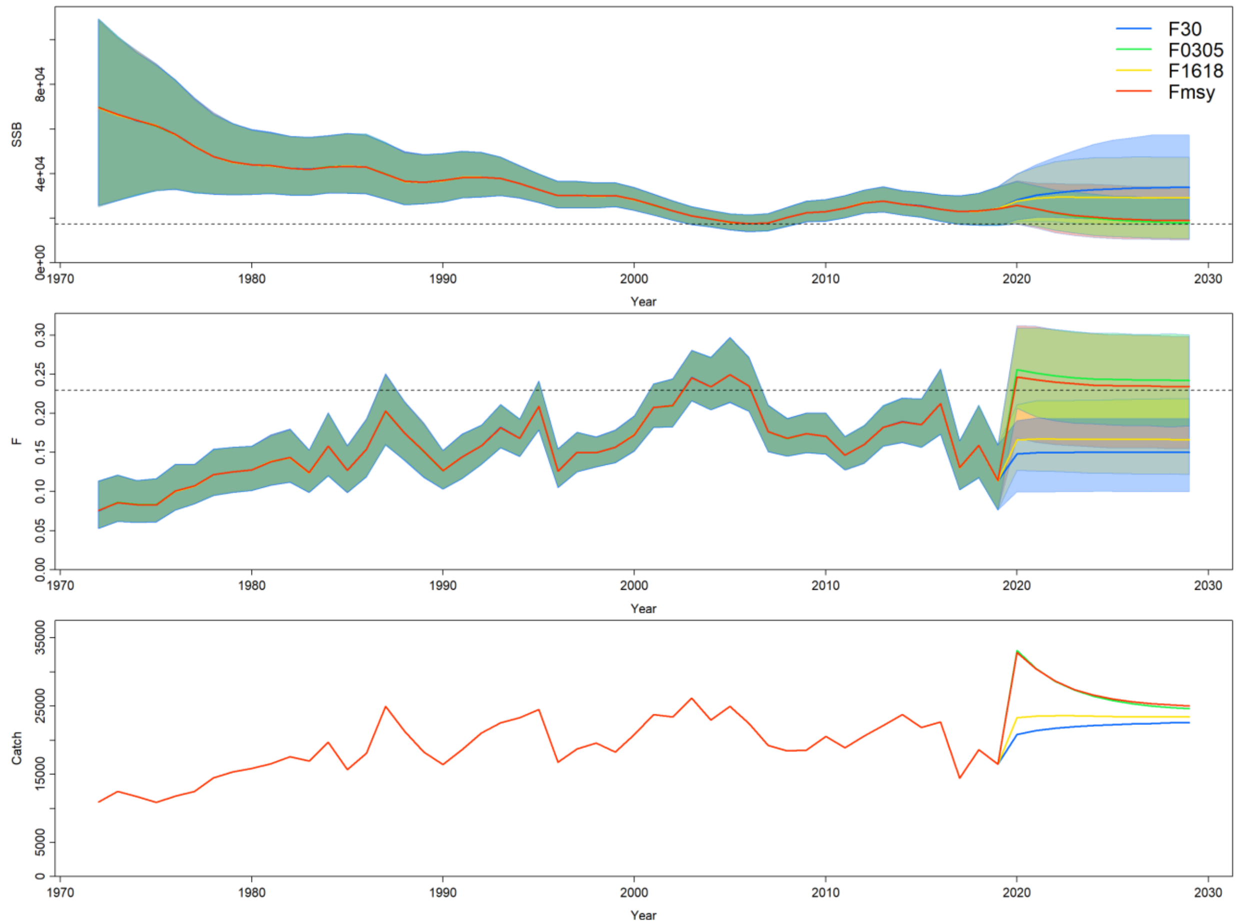


Figure S4. Historical and projected trajectories of spawning biomass and total catch from the Pacific blue marlin ensemble models based upon the four F scenarios: projected spawning biomass, dotted line indicates SSB_{MSY}, shading indicates 95% confidence intervals (top); projected instantaneous fishing mortality (ages 1-10 year^{-1}), dotted line indicates F_{MSY}, shading indicates 95% confidence intervals (center); and projected catch (mt. bottom). Green indicates scenario 1, $F_{2003\text{-}2005}$; red indicates scenario 2, F_{MSY}; yellow indicates scenario 3, $F_{2016\text{-}2018}$; and blue indicates scenario 4, $F_{30\%}$. The list of projection scenarios can be found in Table S3.

1. INTRODUCTION

The International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) Billfish Working Group (BILLWG) completed a benchmark assessment on Pacific blue marlin (*Makaira nigricans*, BUM) in 2013 and updated the assessment in 2016. The status of the 2016 BUM stock was that overfishing was likely not occurring and the stock was likely not overfished relative to MSY-based reference points but cautioned that should catch increase about recent (2012-2014) levels, the stock would be at risk of overfishing (ISC, 2016).

This report describes the 2021 stock assessment for Pacific blue marlin (*Makaira nigricans*), which is considered a pan-Pacific stock caught primarily in tropical and sub-tropical waters. The best available scientific information including the up-to-date catch, catch-per-unit-effort (CPUE), and composition data from 1971-2019 were provided by individual ISC countries, the Western and Central Pacific Fisheries Commission (WCPFC), and the Inter-American Tropical Tuna Commission (IATTC). It was decided to run the assessment using a two-sex, single-stock model in Stock Synthesis version 3.30.16 with a quarterly time step (Methot and Wetzel, 2013). Biological parameters were discussed by the BILLWG at the data preparatory meeting in November 2020, where a new growth curve was presented (Chang et al., 2020). The BILLWG could not achieve agreement on which growth curve was appropriate for the Pacific BUM because the two growth curves were significantly different in terms of function (von Bertalanffy vs two-stanza growth) and L_{inf}, a parameter that has been shown to be very sensitive in assessment models. Due to this, the BILLWG agreed to explore two possible base-case models for BUM, one using the life history parameters from the 2016 assessment (hereafter, the old growth model) and one using the new growth curve and updated natural mortality based upon the new growth curve (the new growth model). After extensive analysis and diagnostic evaluation, the BILLWG agreed that both models fit the available data sufficiently, and neither model could be discarded based upon their diagnostics. Therefore, the BILLWG has decided to put forth a model ensemble of both growth curves, weighting each model equally using a multivariate delta-lognormal approach (Carvalho *et al.* 2021, Walter *et al.* 2019, Walter and Henning 2020, Winker *et al.* 2019).

2. MATERIALS AND METHODS

2.1. Spatial and Temporal Stratification

The Pacific blue marlin (*Makaira nigricans*) is assessed as a Pacific-wide stock (i.e., Williams *et al.* 2020). Blue marlin (BUM) are found in tropical and sub-tropical waters throughout the entire region. The working group agreed to run the model from 1971 to 2019 when catch, CPUE, and size-frequency data are all available, although there has been fishing on the stock historically, with industrial catch records as early as the 1950s. Three types of data were used: fishery-specific catches, relative abundance indices, and length or weight measurements. The fishery data were compiled for 1971-2019, noting that the catch data and length composition data were compiled and modeled quarterly. Available data, sources of data, and temporal coverage of the datasets used in the updated stock assessment are summarized in Figure 1. Further details are presented below.

2.2. Definition of Fisheries

Twenty different fleets are available for inclusion in the assessment model, 16 catch time series, 4 CPUE indices, 7 fleets with length composition data, and one fleet with weight composition data. The fleet names and numbers are detailed in Table 1. The acronyms in the fleet names are defined as follows: WCPFC is Western and Central Pacific Fisheries Commission; EPO is Eastern Pacific Ocean; LL is longline; CLL is coastal longline; early is the early time period; late is the late time period, DRIFT is high-seas and coastal driftnet gear; Oth is other fishing gear (e.g. troll, handline, net, harpoon, and others); PS is purse seine gear; Bait is bait fishing.

2.3. Catch

Catch was input into the model quarterly (i.e., by calendar year and quarter) from 1971 to 2019 for the sixteen individual catch fisheries. Catch was reported in terms of catch biomass (mt) for all fisheries.

Three countries (i.e., Japan, Chinese Taipei, and the USA) provided national catch data (Hirotaka Ijima, NRIFSF, personal communication; Yi-Jay Chang, NTU, personal communication; Russell Ito, NOAA PIFSC, personal communication). Blue marlin catches for all other fishing countries were collected from WCPFC category I and II data (Peter Williams, SPC, personal communication) and IATTC category I and II data (Shane Griffiths, IATTC, personal communication).

The resulting best available data on blue marlin catches by fishery from 1971-2019 were tabulated and are shown in Figure 2 and Table 2. The historical maximum and minimum annual blue marlin catches were 26,138 metric tons in 2003 and 10,882 metric tons in 1975, respectively. Catches increased from the 1970s to the 1990s and remained high until the 2000s when they started to decline. The catch by Japanese fleets has decreased since 1975 while catch from the Chinese Taipei and other longline fleets has increased. The catch by other fleets are small by comparison and varies (Figure 2). Overall, longline gear has accounted for the vast majority of Pacific blue marlin catches, Japanese fleets dominating the catch before 2000 and Chinese Taipei and other longline fleets dominating thereafter. The annual catch of blue marlin in the Pacific averaged 18,808 metric tons in the period since the terminal year of the last assessment (2015-2019).

2.4. Abundance Indices

Relative abundance indices for Pacific blue marlin based on standardized CPUE were prepared for this assessment and are shown in Figure 3 and Tables 3 and 4. Japanese CPUE indices were updated using the 2016 habitat model with updated environmental data and removing coastal longline data. Japanese CPUE data were split into two indices (1971-1993, 1994-2019) corresponding to a change in the logbook reporting requirements (Ijima, 2020).

Operational fishing data collected in the Hawaii-based longline fishery by fishery observers in 1995-2019 were used for CPUE standardization of US longline fleets (Sculley and Brodziak 2020). The fishery operates in two sectors; a shallow-set sector targeting swordfish and a deep-set sector targeting tunas. Blue marlin is caught as bycatch in both sectors. Only the index based upon the data from the deep-set sector were included in the assessment due to poor diagnostics and high variability in the shallow-set CPUE standardization.

CPUE were standardized from the Chinese Taipei distant water tuna longline fishery using a spatio-temporal model. The distant-water longline fleet was standardized from 171-2019 (Hsu and

Chang, 2020). This index was split into three time periods (1971-1978, 1979-1999, 2000-2019) to account for changes in the fishery operations and logbook data quality.

Correlations among CPUE indices were analyzed in the 2021 assessment using the diags component of the FLCore package (Version 2.6.6, Kell *et al.* 2007) in R (version 3.4.0, R Core Team, 2018). These packages provide a standardized method to plot and summarize CPUE data so that modelers can better evaluate their input data into assessment models. Each CPUE index was fit using a Loess smoother with only year as an explanatory variable using the default phase and number of nodes in the R package gam (Hastie, 2018), and the residuals from that smoother were examined graphically. Patterns in correlations among CPUE indices for the assessment were generally positive, except between Hawaii longline fleet and the Taiwanese longline fleet, which had a strong negative correlation. Upon further inspection, it was noted that the Hawaii longline index showed a strong decline over time with a slight flattening in the last few years. Based on the graphical inspection of relative CPUEs and the correlation analysis, the data supported the exclusion of the Hawaii Longline CPUE index in the models. In addition, the R_0 likelihood profile indicated conflict between the Japanese and Chinese Taipei late indices, and sensitivity runs excluding each of these indices were conducted to evaluate their influence on the model results.

2.5. Size Composition Data

Quarterly fish length composition data from 1971–2019 for eight fisheries were available for the assessment; seven were ultimately used, and are summarized in Table 3. Length composition data for the French Polynesia longline fleet were not included because it accounted for <8% of the total catch in the fishery, and required an additional 18 parameters to fit. Additionally, due to difficulties in estimating the selectivity of the early length composition data and the poor data quality of samples measured in this period, Chinese Taipei length composition data before 2005 were excluded from the models, which is consistent with the 2016 BUM assessment.

Length frequency data were compiled using 5-cm length bins from 50 to 320 cm for all fleets except French Polynesia and Other longline, which provided data in 10-cm bins, and the Japanese driftnet fishery, which provided data in weight (kg). The lower boundary of each bin was used to define each bin for all composition data, and each observation consisted of the actual number of blue marlin measured. The composition data were agreed upon at the BILLWG data workshop as the best available scientific information for the 2021 stock assessment.

Figures 4 and 5 show the quarterly length and size compositions. Most of the fisheries' mean size caught approximately 150cm EFL individuals. The purse seine fleets caught larger fish with a mean of around 200 cm EFL, while the driftnet caught fish with a mean weight of around 125 kg.

The aggregate length composition distributions were relatively consistent between fleets, except the Chinese Taipei longline, Japanese driftnet, and EPO purse seine, which all caught more large fish than the other fleets (Figures 6 and 7).

2.6. Model Description

The assessment was conducted with Stock Synthesis (SS) version 3.30.16.00-SAFE released 09/30/2020 using Otter Research ADMB 12.2 (Methot and Wetzel 2013). The 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.

2.7. Biological Parameters

Biological parameters such as the length-weight relationship, maturity, fecundity, and stock-recruitment were the same for both the old growth and new growth models. The old growth model parameters were based upon those used in the 2016 BUM assessment. The maximum age was set to 26 years, the age at size L_1 (the age in which growth is no longer linear but follows the growth curve) is 1, and a von Bertalanffy growth curve is used (Figure 8). In addition, age-specific natural mortality is based upon the growth curve and is the same as the 2016 assessment, with natural mortality at 0.42 for age 0 individuals, 0.37 for age 1+ males and 1-3 females, and 0.22 for 4+ females (Figure 9).

Parameters for growth in the new growth model are based upon work by CCMs presented at the November BILLWG data preparatory meeting (Chang et al. 2020, ISC 2021). In this working paper, growth was modeled as a two-stanza growth curve; to simulate a similar pattern, the BILLWG agreed to model linear growth until age 0.5 and used a Richards parameterization after age 0.5 (Figure 8). Furthermore, the maximum age was set to 20, and natural mortality was age-specific based upon the new growth curve parameters with natural mortality at 0.44 for age 0 individuals, 0.38 for age 1+ males, and stepping down from 0.44 to 0.26 for age 4+ females (Figure 9). In the new growth curve, K, or Brody growth rate coefficient was smaller than the old growth curve and the length at A_{max} was also smaller (Table 5). The CV of both growth curves was set to be equal to the CV used in the 2016 assessment.

For both models, the sex ratio at birth was assumed to be 1:1, and a Beverton-Holt spawner-recruit relationship with steepness (h) fixed at 0.87 was used. SigmaR (σ_r) was initially fixed at 0.6, but in the new growth model, it was rescaled to 0.4 based upon modeled results. The maturity ogive fixed the length at 50% maturity at 179 cm EFL with a slope of -0.20 (Figure 10). The weight-length curve was also sex-specific with females reaching larger weights earlier than males (Figure 11).

Other than the biological parameters, the modeling approach and input data for both the old growth and new growth models were identical. Therefore, the following description of the assessment model is for both models.

2.8. Data Observation Models

The assessment model fit three data components: 1) total catch; 2) relative abundance indices; and 3) composition data. The observed total catches were assumed to be unbiased and relatively precise and were fitted assuming a lognormal error distribution with a standard error (SE) of 0.05. The relative abundance indices were assumed to have log-normally distributed errors with SE in log-space (log(SE)) which was $\log(SE)=\mathrm{sqrt}(\log(1+CV^2))$, where CV is the standard error of the observation divided by the mean value of the observation and sqrt is the square root function.

CPUEs were assigned to quarters with the highest average catch. Japanese longline fleets (S1, and S2) and Chinese Taipei longline fleets (S4 – S6) were assigned to quarter one. The US Hawaii longline fleet was assigned to quarter four, but this index was ultimately excluded from the models. This CPUE index was excluded from the models because it was shown to conflict with the other input data based upon *a priori* CPUE comparison analysis, and the general declining trend from the index. This decision is consistent with the 2016 BUM assessment. 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 (Table 6). The minimum CV was scaled to a minimum of 0.2 and then reweighted based upon the Francis method using the root-mean-square error (RMSE, i.e., square root of the residual variance, Francis 2011). Ultimately, no additional variance based upon the RSME was added to the assessment models.

The composition data were assumed to have multinomial error distributions with the error variances determined by the effective sample sizes. Measurements of fish are usually not random samples from the entire population. Rather, they tend to be highly correlated within a set or trip (Pennington *et al*., 2002). The effective sample size is usually substantially lower than the actual number of fish measured because the variance within each set or trip is substantially lower than the variance within a population. All fleets had effective sample sizes equal to 1/10 of the total number of samples in each quarter, in alignment with previous assessments (ISC 2016). In addition, quarters with fewer than 25 total samples were removed from the time series due to limited sample size, and the maximum number of samples was set to 50, as agreed upon by the modeling sub-group. Size data from fleet F13 (French Polynesia) was ultimately excluded from the model due to its small contribution to the fishery and the number of parameters required to fit the data.

Length composition data were weighted using the 2-stage process based upon the Francis (2011) method. 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). Size composition data were only re-weighted if the Francis method suggested down weighting the fleet, to ensure that the CPUE indices contributed substantially to the likelihood. Based upon the R_0 likelihood profiles, it was ultimately determined to down weight the Japanese early LL size data (F1) and the Hawaii longline size data (F7). This resulted in better fits to the CPUE indices and reduced the conflict in the likelihood profiles.

2.9. Estimation of Fishery Selectivity

Selectivity was estimated as a double-normal curve for Japanese driftnet (F4), Chinese Taipei longline (F10), and other longline fleets (F12). Three fleets were estimated using a cubic spline function, the Japanese early longline fleet (F1) was estimated with four parameters, and the Japanese late longline fleet (F2) and US Hawaii longline fleet (F7) were estimated with three parameters. The EPO purse seine fishery was estimated as asymptotic lognormal (Figure 12). In addition, the Japanese longline late fleet (F2) and the US Hawaii longline fleet (F7) included time-varying selectivity (Figures 13 and 14). All other fleets were mirrored to the fleet that was believed to have the most similar selectivity pattern (Table 7).

2.10. Data Weighting

Index data were prioritized in this assessment based on the principles that relative abundance indices should be fitted well because abundance indices are a direct measure of population trends and scale, and that other data components such as composition data should not induce poor fits to the abundance indices (Francis, 2011).

It is common practice to re-weight some or all data sets in two stages (Francis, 2011). Stage 2 reweighting was only applied to the size composition data from F1 and F7 to improve the fits to the CPUE data. Both were down-weighted by 0.5.

2.11. Model Diagnostics

Several diagnostics have been evaluated for their utility to identify data conflicts and model misspecification within integrated stock assessment models (Carvalho *et al*. 2017, Carvalho *et al*. 2021). However, Carvalho *et al*. (2017) determined that there was no single diagnostic that worked well in all of the cases they evaluated. Instead, the BILLWG recommend the use of a carefully selected range of diagnostics that proved to increase the ability to detect model misspecification.

Key stock assessment diagnostics identified by Carvalho *et al*. (2021) were implemented to evaluate the base case model using the R package ss3diags.

2.11.1. Retrospective analysis

Retrospective analysis is a way to detect bias and model misspecification (Hurtado-Ferro *et al*. 2014). A retrospective analysis was applied to the model results. The diagnostic was implemented here by sequentially eliminating the five most recent years of data from the full stock assessment model (a 5-year "peel") and then re-estimating all stock assessment model parameters from each peel and the full model. Then Mohn's rho was calculated for the biomass and fishing mortality peels, which measures the severity of the retrospective pattern (Hurtado-Ferro *et al*. 2014). Values higher than 0.20 and lower than -0.15 can indicate problematic retrospective patterns and may point to model misspecification, data conflicts, or poor fits to the data.

2.11.2. Prediction skill

In addition to evaluating the retrospective patterns of the model, understanding how well a model predicts future years is key to evaluating projections. To do so, hindcasting cross-validation was used to predict the next years' observed data from the retrospective peel (Carvalho *et al*. 2021). Then the forecast bias is estimated by comparing the forecasted values from the retrospective peel to the full model. To evaluate the predictive skill, the mean absolute scaled error (MASE) is used to determine if the predicted value improves the model forecast compared to the baseline (Carvalho *et al*. 2021). A MASE score of >1 indicates that the average model forecasts are worse than a random walk model, and a value of 0.5 indicates the model has prediction skill. The hindcasting cross-validation and MASE scores were calculated for the two CPUE indices in the last five years of the assessment, Japanese longline late (S2) and Chinese Taipei longline late (S6).

2.11.3. R0 likelihood profile

An R0 likelihood component profile (Lee *et al*. 2014) was applied to both models' results.

The diagnostic was implemented here by sequentially fixing the equilibrium recruitment parameter, R_0, on the natural log scale, $\log(R_0)$, to a range of values. The relative change in negative log-likelihood units over the range of fixed values for $\log(R_0)$ (the R_0 profile) was compared among the Stock Synthesis model likelihood components for CPUE, length-composition, size-composition, and recruitment deviations using two diagnostic tests. First, a relatively large change in negative log-likelihood units along the R_0 profile was diagnostic of a relatively informative data source for that particular model. Second, a difference in the location of the minimum negative log-likelihood along the R_0 profile among data sources was diagnostic of either conflict in the data or model misspecification (or both).

2.11.4. Age-structured production model

An age-structured production model (ASPM; Maunder and Piner 2015; Carvalho *et al*. 2017) was applied to the model results.

The diagnostic was implemented here by fixing selectivity to its estimated values in the fully integrated stock assessment model, fixing recruitment equal to the stock-recruitment curve obtained from the fully integrated stock assessment model, and then estimating the remaining parameters of the stock assessment model. Trends in relative spawning stock size were compared from the fully integrated stock assessment model and the ASPM.

Carvalho *et al*. (2017) suggest that if the ASPM can fit well to the indices of abundance that have good contrast (i.e. those that have declining and/or increasing trends), then this is evidence of the existence of a production function, and the indices will likely provide information about absolute abundance. On the other hand, Carvalho *et al*. (2017) suggest that if there is not a good fit to the indices, then the catch data alone cannot explain the trajectories depicted in the indices of relative abundance. This can have several causes: (i) the stock is recruitment-driven; (ii) the stock has not yet declined to the point at which catch is a major factor influencing abundance; (iii) the model is incorrect; or (iv) the indices of relative abundance are not proportional to abundance.

2.11.5. Goodness-of-Fit Indices of Abundance

Residuals are examined for patterns to evaluate whether the model assumptions have been met. Many statistics exist to evaluate the residuals for desirable properties. One way is to calculate, for each abundance index, the root-mean-square-error (RSME) was used as a goodness-of-fit diagnostic, with relatively low RMSE values (i.e., RMSE < 0.3) being indicative of a good fit.

2.11.6. Goodness-of-Fit Size Composition Data

Comparisons between the observed and expected mean values of composition data from Francis (2011) were used for model diagnostics. Pearson residuals for size composition data fits were also used as a model diagnostic.

2.11.7. Runs Test

The runs test evaluates the residuals of the CPUE indices and size composition mean length (or weight) trends. This is a nonparametric test for randomness in the sequence of residuals (Carvalho *et al*. 2021, Wald and Wolfowitz 1940). In other words, this test uses a 2-sided p-value to estimate the number of positive or negative residuals in a row (a "run"). CPUE or size composition data that fail the runs test indicate that there may be a pattern in the residuals and the model is unable to fit the data well or is misspecified.

2.12. Model ensemble

Historically, the BILLWG has produced stock status for an assessment based upon a single base-case or reference model with a suite of sensitivity runs to evaluate how alternative parameters, data, or model structure may change the status of the stock. For the 2021 Pacific blue marlin assessment, two valid but notably different growth curves were reviewed at the November 2020 data preparatory workshop. The BILLWG agreed to produce two BUM assessment models using each growth curve and use diagnostics to determine a base case model and also agreed that both models use for the base case model if there is no apparent diagnostics difference between the old and new growth curve models. The BILLWG constructed a base case model by repeating model diagnosis and model modification for the two candidate models. However, the model diagnostics and model fit did not provide a clear choice between the two models and the WG agreed instead to take a model ensemble approach. Without clear evidence to determine the weighting of these models, the WG decided to assign the models equal weights and use a multivariate log-normal approximation (MVLN). MVLN accounts for the inherent covariance between F/F_{MSY} and

SSB/SSB_{MSY} to estimate the stock status, spawning stock biomass, recruitment, and fishing mortality, and the associated uncertainty from the two models (Walter *et al.* 2019, Winker *et al.* 2019).

To average the two models, 10,000 draws from a multivariate log-normal distribution were pulled for each model and exponentiated to obtain the probability distributions around SSB/SSB_{MSY} and F/F_{MSY} using the R package ss3diags (Walter et al. 2019, Carvalho et al. 2021). Then these two distributions are combined equally to calculate the mean and 95% confidence intervals of the stock status in each year. This fully incorporates the uncertainty from both models into the ensemble model result.

2.13. Stock Projections

Consistent with the 2016 assessment, deterministic stock projections were conducted in SS to evaluate the impact of various levels of fishing mortality on future SSB and yield. No recruitment deviations and log-bias adjustment were applied to the future projection in this study. Instead, the absolute future recruitments were based on the expected stock-recruitment relationship. The future projection routine calculated the future SSB and yield that would occur while the specific fishing mortality, selectivity patterns, and relative fishing mortality proportions depended on the specific harvest scenarios. In this study, the last three model years' (2016-2018) selectivity patterns and relative fishing mortality rates were used in population future projection. The projection started in 2020 and continued through 2029 under four different harvest scenarios:

1. **High F Scenario (F_{03-05})**: Select the 3 years with the highest average F (age 1-10) and apply this fishing mortality rate to the stock estimates beginning in 2020; this corresponds to 2003-2005;
2. **F_{MSY} Scenario (F_{MSY})**: Apply the estimate of the FMSY fishing mortality rate to the stock estimates beginning in 2020;
3. **Status Quo F Scenario (F_{16-18})**: This will be the average F (age 1-10) during 2016-2018 (F2016-2018);
4. **Low F Scenario (F_{30})**: Apply an $F_{30\%}$ fishing mortality rate to the stock estimates beginning in 2020.

Each constant fishing mortality scenario was run for both the old growth and new growth models for eight projection runs. Then each scenario was combined using the MVLN method used to combine the growth models to produce four ensemble projections in total (Walter *et al.* 2019, Walter and Winker 2020).

3. RESULTS

Diagnostics and results for the old growth model will be presented first, followed by the new growth model. Then the model ensemble results, projections, and sensitivity runs will be described.

3.1. Old Growth Model

3.1.1. Model Convergence

All estimated parameters in the base case model were within the set bounds, and the final gradient of the model was <0.0001 and the hessian matrix for the parameter estimates was positive definite, which indicated that the model converged to a local or global minimum. Results from 30 model runs with different random initial starting values for estimated parameters using the internal "jitter" routine in SS supported the result that a global minimum was obtained (i.e., there was no evidence of a lack of convergence to a global minimum, Figure 15).

3.1.2. Model Diagnostics

Figure 16 and Table 8 presents the results of the likelihood profiling on the logarithm of the unfished recruitment parameter R_0, i.e. log(R_0), for each data component. Detailed information on changes in negative log-likelihoods among the various fishery data sources are shown in Tables 9 and 10 and Figures 17 and 18.

Changes in the likelihood of each data component indicated how informative that data component was to the overall estimated model fit. Ideally, relative abundance indices should be the primary sources of information on the population scale in a model (Francis, 2011).

There was a relatively large change in the R_0 profile for estimated recruitment deviations (Recruitment) for R_0 values below the maximum likelihood estimate (MLE) relative to the data likelihood components for survey (CPUE) and length composition data (Figure 16). At R_0 values above the MLE, the length composition data contributed the most to the change in log-likelihood. This result indicated that the estimation of the recruitment deviations and length composition data were relatively informative within the likelihood. Overall, none of the model components appeared to conflict, with MLEs around 7.1.

The Japanese early and late CPUE index and Chinese Taipei late CPUE index contributed the most to the survey component of the likelihood. Some conflict was apparent between the Japanese late CPUE index, which had a minimum of around 6.9, and the Chinese Taipei longline index, which had a minimum of around 7.9 (Table 9). The other Chinese Taipei indices were relatively flat and did not contribute substantially to the likelihood (Figure 17). Sensitivity runs were conducted to evaluate the effect of removing each of the late longline indices from the model.

Similar to the abundance indices data, some conflict between the length composition data was apparent in the likelihood profile (Figure 18). The two Japanese longline fleets drove the likelihood at low R_0 values, although the early data suggested a minimum around 7.0 and the late data suggested a minimum around 7.4. The EPO purse seine size data was the largest contributor to the change in likelihood above the MLE and had a minimum of around 6.3 (Table 10, Figure 18). The other length data had relatively flat likelihood profiles and did not contribute substantially to the likelihood. Down-weighting the US HI longline data reduced its influence on the likelihood. Down weighting the Japanese early longline length data also reduced its influence on the likelihood, although it was still an important component. The generalized size composition data were not a significant component of the overall likelihood profile. Only the other longline data (F12) contributed to the likelihood profile, with an R0 minimum of 7.0 (Figure 16, Table 10).

3.1.3. Goodness-of-Fit Indices of Abundance

Goodness-of-fit diagnostics (RMSE) were presented in Table 6, and plots of predicted and observed CPUE by fishery for the base case model were shown in Figure 19. Overall, fits to the CPUE indices were reasonable. Two indices have RMSE<0.3, Japanese LL early (S1) and Chinese Taipei longline early (S4). Chinese Taipei longline late (S6) had an RMSE of ~0.3, and Japanese longline late and Chinese Taipei longline mid (S2 and S4) both had poor fits to the data, with RMSE much greater than 0.3. Some of the poor fit of S2 and S6 could be contributed to the conflict between the two indices identified in the R0 profiles. In the last few years of the assessment, index S2 declines while index S6 increases.

3.1.4. Residuals Analysis of Size Composition Data

Comparisons between the observed and expected mean values of composition data from Francis (2011) were used for model diagnostics. Figure 20 shows the 95% credible intervals for the mean value for the length and generalized size composition data sets. The model fit passed through almost all of the credible intervals.

Fits to the annual length compositions by fleet could be improved (Figure 21), with a few obvious systematic patterns observed in the residuals (e.g., patterns of positive or negative residuals) making it difficult to objectively determine how to improve the fits. This is an important area for future model development.

Assuming standardized residuals were normally distributed, 95% of the measurements would fall within 2 standard deviations of the mean. The majority of Pearson residuals did not meet this criterion (Figure 21); only the Japanese driftnet weight data fit this criterion. In addition, pulses of larger than expected numbers of small fish are observed in several different fleets, which suggest periodic strong recruitments. Overall, the model fit the length modes in composition data aggregated by fishery fairly well using the input effective sample sizes (Figures 22 and 23).

3.1.5. Runs test

Four of the five CPUE indices included in the model passed the runs test (Figure 24). Japanese longline early CPUE was the only index with did not. This indicates that in general, the model can fit the CPUE indices well. Of the five length composition data time series available, three passed the runs test (Figure 25). Japanese longline late, US Hawaii longline, and Chinese Taipei longline length composition data all passed, but the Japanese longline early, and EPO purse seine data did not. It should be noted that the Japanese longline early length composition data were down-weighted in the model, which would cause a degradation in fit. This suggests that additional work will be necessary to improve the fit to the length composition data. This pattern is partly driven by the conflict between the Japanese late longline CPUE index and the Chinese Taipei late longline CPUE index, as removing one of these indices reduces the retrospective pattern (i.e. Mohn's rho is closer to zero), although the pattern would still be considered significant.

3.1.6. Retrospective Analysis

A retrospective analysis of the Pacific blue marlin stock old growth assessment model was conducted for the last 5 years of the assessment time horizon to evaluate whether there were any strong changes in parameter estimates through time. The results of the retrospective analysis are shown in Figure 26. The trajectories of estimated spawning stock biomass and fishing mortality showed there was a tendency for the base case model to overestimate spawning biomass in recent years and underestimate fishing intensity. In addition, the Mohn's rho for biomass (0.39) and fishing mortality (-0.28) fall outside of the range of acceptable values, suggesting that the retrospective pattern is substantial.

3.1.7. Predictive Skill

Two CPUE indices and four length composition time series had at least one data point within the last five years of the assessment, the hindcasting evaluation period. Both of the CPUE indices had MASE scores greater than 1, indicating a poor predictive ability of the model (Figure 27). The

Japanese longline late CPUE index had a worse predictive ability, as three of the five predicted years were outside of the 95% confidence interval of the CPUE index and had a MASE score of 2.03. The last three years of the assessment when the index declines were the worst years of the hindcast. All but one of the predicted CPUE values from the Chinese Taipei late longline index was inside of the 95% confidence interval of the observed CPUE, however, the MASE score was still 1.18, indicating poor predictive skill.

Predictive skill for the length composition data was slightly better than the CPUE data (Figure 28). Two length composition time series had MASE scores below one, Japanese longline late (MASE=0.79) and Chinese Taipei longline data (MASE = 0.72). Both the US Hawaii longline data and the EPO purse seine data had MASE values over one (MASE = 2.16 and 2.61, respectively). Unlike the CPUE data hindcast, all of the predicted length composition data points were within the 95% confidence intervals of the original input data.

3.1.8. Age-structured production model

ASPM results are provided in Figure 29. The models relatively consistent SSB trends during the modeled timeframe, with a slight exception in the first 10 years of the assessment. The asymptotic 95% confidence interval from the fully integrated stock assessment overlapped with the SSB trend from the ASPM for most of the modeled years.

3.1.9. Recruitment deviations

Recruitment appeared to vary between 0.6 and -0.6 without long periods of high or low recruitment throughout the time-period modeled (Figure 30). Variability of recruitment was higher in the early part of the model, from 1970 to around 1995, and was much smaller after 1995 due to the increase in data available to inform recruitment. Bias adjustment was applied to the recruitment deviations, with no bias adjustment prior to 1970, a ramp up in adjustment from 1970-1971, and full bias adjustment applied 1971-2018 (Figure 31).

3.2. New Growth Model

3.2.1. Model Convergence

All estimated parameters in the base case model were within the set bounds, and the final gradient of the model was <0.0001 and the hessian matrix for the parameter estimates was positive definite, which indicated that the model converged to a local or global minimum. Results from 30 model runs with different random initial starting values for estimated parameters using the internal “jitter” routine in SS supported the result that a global minimum was obtained (i.e., there was no evidence of a lack of convergence to a global minimum, Figure 32).

3.2.2. Model Diagnostics

Figure 33 and Table 11 presents the results of the likelihood profiling on the logarithm of the unfished recruitment parameter R_0, i.e. $\log(R_0)$, for each data component. Detailed information on changes in negative log-likelihoods among the various fishery data sources are shown in Tables 12 and 13 and Figures 34 and 35.

There was a relatively large change in the R_0 profile for estimated recruitment deviations (Recruitment) for R_0 values below the maximum likelihood estimate (MLE) relative to the data likelihood components for survey (CPUE), generalized size composition data, and length composition data (Figure 34, Table 11). At R_0 values above the MLE, the length composition and generalized size composition data contributed the most to the change in log-likelihood. This result indicated that the estimation of the recruitment deviations and length composition data were relatively informative within the likelihood. Overall, the generalized size composition data suggested an R_0 value lower than the MLE at 6.92.

The Japanese early and late CPUE index and Chinese Taipei late CPUE index contributed the most to the survey component of the likelihood. Some conflict was apparent between the Japanese late CPUE index, which had a minimum of around 6.7, and the Chinese Taipei late longline index, which had a minimum of around 8.0 (Table 12). The other indices all had minimum R0 values at or above 7.0, while the MLE was around 6.9 (Figure 34). Sensitivity runs were conducted to evaluate the effect of removing each of the late longline indices from the model.

Similar to the abundance indices data, some conflict between the length composition data was apparent in the likelihood profile (Figure 35). The two Japanese longline fleets drove the likelihood at low R_0 values, although the early longline data suggested a minimum around 6.7 and the late data suggested a minimum of around 7.2. The EPO purse seine size data was the largest contributor to the change in likelihood above the MLE and had a minimum of around 6.8 (Table 13, Figure 35). The other length data had relatively flat likelihood profiles and did not contribute substantially to the likelihood. Down-weighting the US HI longline data reduced its influence on the likelihood. Down weighting the Japanese early longline length data also reduced its influence on the likelihood, although it was still an important component. The generalized size composition data were not a significant component of the overall likelihood profile. Only the other longline data (F12) contributed to the likelihood profile, with an R_0 minimum of 6.0 (Figure 33).

3.2.3. Goodness-of-Fit Indices of Abundance

Goodness-of-fit diagnostics were presented in Table 6, and plots of predicted and observed CPUE by fishery for the base case model were shown in Figure 36. Overall, fits to the CPUE indices were reasonable. Two indices have RMSE<0.3, Japanese LL early (S1) and Chinese Taipei longline early (S4). Chinese Taipei longline late (S6) had an RMSE of ~0.3, and Japanese longline late and Chinese Taipei longline mid (S2 and S4) both had poor fits to the data, with RMSE much greater than 0.3. Some of the poor fit of S2 and S6 could be contributed to the conflict between the two indices identified in the R0 profiles. In the last few years of the assessment, index S2 declines while index S6 increases.

3.2.4. Residuals Analysis of Size Composition Data

Comparisons between the observed and expected mean values of composition data from Francis (2011) were used for model diagnostics. Figure 37 shows the 95% credible intervals for the mean value for the length and generalized size composition data sets. The model fit passed through almost all of the credible intervals.

Fits to the annual length compositions by fleet could be improved (Figure 38), with a few obvious systematic patterns observed in the residuals (e.g., patterns of positive or negative residuals) making it difficult to objectively determine how to improve the fits. This is an important area for future model development.

Assuming standardized residuals were normally distributed, 95% of the measurements would fall within 2 standard deviations of the mean. The majority of Pearson residuals did not meet this criterion (Figure 38), only the Japanese driftnet weight data and EPO purse seine length composition data fit this criterion. In addition, pulses of larger than expected numbers of small fish are observed in several different fleets, which suggest periodic strong recruitments. Overall, the model fit the length modes in composition data aggregated by fishery fairly well using the input effective sample sizes (Figures 39 and 40).

3.2.5. Runs test

Four of the five CPUE indices included in the model passed the runs test (Figure 41). Japanese longline early CPUE was the only index with did not. This indicates that in general, the model can fit the CPUE indices well. Of the five length composition data time series available, only two passed the runs test (Figure 42). Japanese longline early and late length composition data passed, but the US Hawaii longline, Chinese Taipei longline, and EPO purse seine data did not. This suggests that additional work will be necessary to improve the fit to the length composition data. It is important to note, that some of the misfit to the US Hawaii longline data may be due to this fleet being down-weighted in the likelihood, which would degrade the fit to the data in favor of fitting other data components better. While the Japanese longline early data were also down-weighted, it did not affect the fit in the model.

3.2.6. Retrospective Analysis

A retrospective analysis of the Pacific blue marlin stock old growth assessment model was conducted for the last 5 years of the assessment time horizon to evaluate whether there were any strong changes in parameter estimates through time. The results of the retrospective analysis are shown in Figure 43. The trajectories of estimated spawning stock biomass and fishing mortality showed there was a tendency for the base case model to overestimate spawning biomass in recent years and underestimate fishing intensity. In addition, the Mohn's rhos for biomass (0.31) and fishing mortality (-0.24) fall outside of the range of acceptable values, suggesting that the retrospective patterns were substantial.

3.2.7. Predictive Skill

Two CPUE indices and four length composition time series had at least one data point within the last five years of the assessment, the hindcasting evaluation period. Both of the CPUE indices had MASE scores greater than or equal to one, indicating a poor predictive ability of the model (Figure 44). The Japanese longline late CPUE index had a worse predictive ability, as three of the five predicted years were outside of the 95% confidence interval of the CPUE index and had a MASE score of 1.88. The last three years of the assessment when the index declines were the worst years of the hindcast. All of the predicted CPUE values from the Chinese Taipei late longline index were inside of the 95% confidence interval of the observed CPUE, however, the MASE score was still 0.99, indicating poor predictive skill.

Predictive skill for the length composition data was slightly better than the CPUE data (Figure 45). Two length composition time series had MASE scores below one, Japanese longline late (MASE=0.56) and Chinese Taipei longline data (MASE = 0.78). Both the US Hawaii longline data and the EPO purse seine data had MASE values over one (MASE = 1.73 and 2.48, respectively). Unlike the CPUE data hindcast, all but one of the predicted length composition data points were within the 95% confidence intervals of the original input data.

3.2.8. Age-structured production model

ASPM results are provided in Figure 46. The models relatively consistent SSB trends during the modeled timeframe, with a slight exception in the first 10 years of the assessment. The asymptotic 95% confidence interval from the fully integrated stock assessment overlapped with the SSB trend from the ASPM for all of the modeled years.

3.2.9. Recruitment deviations

Recruitment appeared to vary between 0.5 and -0.5 without long periods of high or low recruitment throughout the time-period modeled (Figure 47). Variability of recruitment was fairly consistent throughout the time series, but may be slightly lower after 2000 due to the increase in data available to inform recruitment. Bias adjustment was applied to the recruitment deviations, with no bias adjustment prior to 1965, a ramp up in adjustment from 1965-1971, and full bias adjustment applied 1971-2018 (Figure 48).

3.3. Stock Assessment Results

3.3.1. Comparison between the old growth and new growth models

While the trends in spawning stock biomass and fishing mortality are similar between the old growth and new growth models, several important differences can be observed (Figure 49). The new growth model had a larger virgin SSB than the old growth model. The first 10 years of the assessment has a sharp decline in SSB for the new growth model, until the 1980s when the scale of the SSB is the same as the old growth model. After 1980, the trend in SSB is almost identical. Fishing mortality is very similar between the two models as well. The biggest difference between the two models is the estimates of F_{MSY} and SSB_{MSY}. SSB_{MSY} in the new growth model is lower than the old growth model; therefore, there are fewer years in which the stock is overfished in the new growth model. Similarly, F_{MSY} is smaller in the old growth model, and therefore there are more years in which overfishing is occurring in the old growth model and none in the new growth model. However, the current stock statuses for both models are the same.

3.3.2. Ensemble model results

Estimates of spawning stock population biomass declined until the mid-2000s, increased again until 2021, and was been relatively flat until the present with a minimum of 17,582 metric tons in 2006, and the current SSB of 24,241 metric tons in 2019 (Table 14 and Figure 50, top and center left panels). The time-series of SSB at the beginning of the spawning cycle (quarter 2) averaged 57,883 metric tons during 1971-1979; 41,345 metric tons during 1980−1989; 34,032 metric tons during 1990−1999; 21,437 metric tons during 2000−2009, and 24,822 metric tons in 2010−2019. The precision of SSB estimates gradually improved over time. Overall, SSB exhibited a long-term decline from the early 1970s to the 2000s and has since exhibited a moderate increase.

Recruitment (age-0 fish) estimates indicated a long-term fluctuation around a mean of approximately 896,000 (Table 14 and Figure 50, bottom left panel). Recruitment was low in the early part of the time series (1971-1976) with an average of 725,000 recruits. The model estimated that several strong year classes (> 1000 thousand recruits) recruited to the fisheries in 1981-1982, 1986-1989, 1991, 1995, 2006, 2009-2010, and 2018 followed by several weak year classes. While the overall pattern of recruitment from 1971-2014 was variable, there was no apparent long-term trend in recruitment strength (Table 14 and Figure 50, bottom left panel).

Throughout the assessment time horizon, estimated fishing mortality (arithmetic average of F for ages 1 - 10) gradually increased from the early 1970s to the 1990s, peaked at 0.25 $year^{-1}$ in 2005 in response to higher catches, and afterward declined to 0.13 $year^{-1}$ in the most recent years (2017-2019) (Table 14 and Figure 50, top and center right panels).

3.4. Biological Reference Points

Biological reference points were computed from the two alternative growth Stock Synthesis models using the MVLN method. Since most life-history parameters for Pacific blue marlin, including steepness, were considered reasonably well defined and no target or limit reference points have been established for Pacific blue marlin under the auspices of the WCPFC, MSY-based biological reference points were used to assess relative stock status (Table 15). The point estimate of the maximum sustainable yield was MSY = 24,600 metric tons. The point estimate of the spawning stock biomass to produce MSY was SSB_{MSY} = 20,677 metric tons. The point estimate of F_{MSY}, the fishing mortality rate to produce MSY on ages 1-10 fish was F_{MSY} = 0.23 and the corresponding equilibrium value of spawning potential ratio at MSY was SPR_{MSY} = 17%.

3.5. Stock Status

Compared to MSY-based reference points, the current spawning biomass (average of 2017-2019) was 13% above SSB_{MSY} and the current fishing mortality (average for ages 1-10 in 2017-2019) was 40% below F_{MSY}. The Kobe plot indicates that the Pacific blue marlin stock is likely not overfished and is likely not subject to overfishing relative to MSY-based reference points (Figure 51). Using the 95% confidence intervals for the ensemble model, there is a 19% probability that the stock is overfished and overfishing is not occurring (SSB/SSB_{MSY} <1, F/F_{MSY}<1), and an 81% probability that the stock is not overfished and overfishing is not occurring (SSB/SSB_{MSY}>1, F/F_{MSY}<1).

3.6. Sensitivity Analyses

In the April 2021 BILLWG workshop, it was agreed that life-history parameters, alternative model structure, and alternative data weighting would be evaluated in sensitivity analyses in the 2021 assessment (Table 16) to examine the effects of plausible alternative model assumptions and data input. These analyses were:

(1) **Sensitivity analysis on growth**: The BILLWG agreed to conduct four sensitivity analyses for growth. These were an alternative growth curve with a 10% larger maximum size and 10% smaller maximum size, as well as a 10% larger Brody growth rate coefficient (K) and a 10% smaller K for each growth curve.

(2) **Sensitivity analysis on natural mortality**: The BILLWG agreed to conduct two sensitivity analyses for natural mortality at age for each growth curve. These were a low natural mortality scenario where M at age was 10% lower than the base case for each age group and a high natural mortality scenario where M at age was 10% higher than the base case for each age.

(3) **Sensitivity analysis on steepness**: The BILLWG agreed to run three additional sensitivity runs on steepness for each growth curve. Steepness was fixed at h=0.95, h=0.8, and h=0.65.

(4) **Sensitivity analysis on maturity**: The group agreed to run two sensitivity analyses for the maturity ogive. These were alternative maturity ogives with L_{50} = 161.8 cm (10% lower) and alternative maturity ogives with L_{50}=197.7 cm (10% higher).

(5) **Sensitivity analysis on initial equilibrium catch**: The group agreed to run one sensitivity analysis for the initial equilibrium catch for each growth model. This fixed the equilibrium catch to the value used in the 2016 assessment instead of estimating it, which was down in this assessment.

(6) **Sensitivity analysis on the inclusion of CPUE indices**: The group agreed to run two sensitivity analyses on the inclusion of CPUE indices in each growth model. Either the Japanese longline late index (S2) or the Chinese Taipei longline late index (S6) was excluded for the run, which resulted in a single CPUE index for the last 20 years of the assessment.

(7) **Sensitivity analysis on size composition weighting**: The group agreed to run a sensitivity analysis on the weighting of the size composition data for each growth model. In this run, all size composition data (length and weight) were down-weighted by 0.5 to reduce their influence on the model results.

For each sensitivity run, comparisons of spawning stock biomass and fishing intensity (1-SPR) trajectories were completed (Figure 52). Additionally, the BILLWG produced a Kobe plot, that showed the patterns of the base case and terminal year estimates for the key sensitivity runs (Figure 53).

For one of the 30 sensitivity runs, the stock status indicated that the stock was overfished and experiencing overfishing, this is the old growth model with steepness = 0.65 (Figure 53). Five runs indicated that the stock was overfished, but overfishing was not occurring. These are the new growth with low steepness (Run 23), new growth with 10% higher L_{inf} (Run 11), old growth with 10% higher L_{inf} (Run 12), old growth with 10% higher length at 50% maturity (Run 20) and old growth with steepness = 0.8 (Run 26). All other runs estimated that overfishing was not occurring and the stock was not overfished. However, runs 9, 16, 25, and 30 all were close to $SSB/SSB_{MSY} = 1$. Overall, the old growth model and its sensitivity runs were more pessimistic on stock status than the new growth models. Finally, the assessment models appear to be most sensitive to changes in steepness, changes in L_{inf}, and changes to the Brody growth rate coefficient K.

Overall, the results of the sensitivity analyses confirmed the robustness of the base case model, and it was concluded that other sensitivity runs were not necessary for this stock assessment.

3.7. Stock Projections

Projection results showed the average spawning stock biomass the average catch, and the probability of achieving the spawning stock biomass target in 2018-2022, 2027, and 2037 for each of the four constant fishing mortality model ensemble scenarios (Table 17 and Figures 54 and 55). Results show that for fishing mortalities at or above F_{MSY} (scenarios 1 and 2), SSB biomass decreases towards SSB_{MSY}, but remains above it in 2029. For scenarios 3 and 4, F is below F_{MSY} therefore SSB remains above SSB_{MSY} and increases from 2019 levels to 2029. Catch at the beginning of the projection period is much higher for scenarios 1 and 2; however, by 2029 the difference in catch between the highest F scenario (F_{0305}, scenario 1) and the lowest F scenario (F_{30}, scenario 4) is only about 2500 metric tons. All of the constant F projections have at least a 50% probability of being above SSB_{MSY} and below F_{MSY} in 2029.

3.8. Assessment Challenges

The BILLWG identified several challenges in developing the Pacific blue marlin stock assessment model that contributed to uncertainty in the assessment results. The four major sources of uncertainty were detailed by the BILLWG and should be carefully evaluated in the future.

3.8.1. Fleets S2 and S4 conflict – concerns about the CPUE standardization method

The BILLWG expressed concerns over the use of the Japanese longline late CPUE index in the assessment models. Both the new and old growth models had problems fitting the last few years of the Japanese longline late CPUE index. The BILLWG agreed that it would be important to understand the effects of removing the Japanese longline late CPUE index on the strong retrospective patterns. The BILLWG noted that the Japanese longline fleet exhibited a different spatial distribution of fishing effort in the past decade compared to the rest of the time series. In particular, graphs of fishing efforts showed that the Japanese longline fishing grounds in the EPO had a decreasing pattern starting around 2010 (Ijima 2020). The BILLWG discussed how to treat the Japanese longline late CPUE index from 2010-2019 and whether it would be appropriate to split the CPUE index before and after the year 2010. Some BILLWG members noted that if the index was not a consistent abundance index, then it should be excluded, while others pointed out that there was probably useful information on abundance trends in the index at least before 2010. The BILLWG agreed that additional work should be undertaken to understand fully the trends from the Japanese longline catch and effort distribution, and noted that a new CPUE standardization is being undertaken on the Japanese CPUE data, and the preliminary results were presented to the BILLWG at the November 2020 data preparatory workshop (Ijima and Koike 2021).

3.8.2. Life History Parameters

The BILLWG noted that two valid growth curves were available for Pacific blue marlin for this assessment. Neither growth curve could be excluded based upon the assessment model diagnostics, and trends in SSB and fishing mortality were very similar between the two models. Also, the BILLWG agreed that there was no clear basis to choose between the growth models based on biological realism. It was suggested that the old growth model might be somewhat biased at younger ages while the new growth model may be somewhat biased at the older ages. These apparent biases were due to differences in the growth parameter estimates under both models. The BILLWG noted that the old growth model had an estimated Brody growth coefficient (k) that was too low, while the new growth model had an estimated asymptotic length (Linf) that was too low. The biggest difference between the two growth curves is that the new growth curves have much lower L_{inf} values than the old growth curve. The result of using this curve is that large females are no longer part of the SSB in the new growth curve models, which results in larger virgin population sizes and lower female relative SSB. The BILLWG agreed that additional work to determine the appropriate growth curve, or if time-varying growth should be considered, is important to this stock. Some of this work is already in progress as part of the ISC International Billfish Biological Sampling program, which hopes to answer many questions about the biology of swordfish and marlins.

3.8.3. Retrospective analysis

The BILLWG noted the strong pattern in the retrospective analysis for both the old growth and new growth models. There were no substantial retrospective patterns in the 2016 BUM stock assessment model. Underreporting of BUM catch could lead to an overestimation of biomass and different CPUE standardizations and data weighting in the current assessment compared to the 2016 assessment could change the retrospective pattern. The BILLWG also suggested that misspecification of life-history parameters could be another cause of retrospective patterns. Overestimation of biomass could be driven by slower growth or higher natural mortality, like those in the new growth model. However, the life history parameters under the old growth model were identical to those used in the 2016 benchmark assessment. Since the 2016 benchmark assessment model did not exhibit a retrospective pattern, it seemed unlikely that life-history parameter misspecification was a substantial factor in the strong retrospective patterns in the current old growth model. The conflict between the Japanese longline late CPUE index and the Chinese Taipei longline CPUE index also contributed to the retrospective pattern. Removing the Japanese longline late index decreased the retrospective pattern, although it did not remove it completely, which suggests that other sources are contributing to the pattern observed. Retrospective patterns that show a positive Mohn's ρ for biomass (negative Mohn's ρ for F) like this assessment imply consistent overestimation of biomass and the highest risk for overfishing (Hurtado-Ferro *et al.* 2015).

3.8.4. Projection skill

Similar to the concerns with the retrospective pattern, it was noted that the prediction skill of the model over the last 5 years is relatively poor, which causes concern over the validity of the future projections. The CPUE indices had the worst MASE scores, which again points back to the conflict between the Japanese longline late CPUE index and the Chinese Taipei longline CPUE index. Work to resolve this conflict, and fully validate the CPUE standardizations for each fleet will help to understand better the patterns observed in this assessment.

3.9. Special Comments

BILLWG achieved an ensemble model using the best available data and biological information. However, the BILLWG recognized that there is considerable uncertainty in input CPUE data in the recent years and life history parameters, especially growth. The BILLWG considered an extensive suite of model formulations and associated diagnostics for developing the assessment models. Overall, the BILLWG found issues with the new growth and old growth model diagnostics and sensitivity runs that indicated some data conflicts exist (see sections Assessment Challenges and Sensitivity Analyses), but none of the model diagnostics suggested that the model results were invalid. To improve the stock assessment, the BILLWG also recommends continuing model development work, to reduce data conflicts and modeling uncertainties, reevaluating and improving input assessment data, and for ISC countries to participate in the International Billfish Biological Sampling program to improve estimates of life-history parameters.

3.10. Conservation information

The Pacific blue marlin stock has produced annual yields of around 18,800 mt per year since 2015, or about 90% of the MSY catch amount. Blue marlin stock status from the ensemble model indicates that the current median spawning biomass is above SSB_{MSY} and that the current median fishing mortality is below F_{MSY}. However, uncertainty in the stock status indicates a 19% chance of Pacific blue marlin being overfished. Both the old and new growth models show evidence of spawning biomass being above SSB_{MSY} and fishing mortality being below F_{MSY} during the last 5 years. Catch biomass has been declining for the last 5 years, and therefore the stock has a low risk of experiencing overfishing or being overfished unless fishing mortality increases to above F_{MSY} based upon stock projections. However, it is also important to note that retrospective analyses show that the assessment model tends to overestimate biomass in recent years, in part due to rapid changes in longline CPUE in recent years.

4. ACKNOWLEDGMENTS

We thank the fishery stakeholders, data providers, and participants in the ISC Billfish Working Group meetings for their help in preparing and providing information for this assessment of Pacific blue marlin.

REFERENCES

Andrews, A.H., Humphreys Jr, R.L. and Sampaga, J.D., (2018). Blue marlin (*Makaira nigricans*) longevity estimates confirmed with bomb radiocarbon dating. Canadian Journal of Fisheries and Aquatic Sciences, 75(1), 17-25.

Brodziak, J., (2013). Combining Information on Length-Weight Relationships for Pacific Blue Marlin. ISC/13/BILLWG-1/01

Brodziak, J. (2021). Natural mortality rates of Pacific blue marlin. ISC/21/BILLWG-01/03.

Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., and Piner, K. R. (2017). Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research, 192: 28-40.

Carvalho, F.,Winker, H.,Courtney, D.,Kapur, M.,Kell, L.,Cardinale, M.,Schirripa, M.,Kitakado, T.,Yemane, D.,Piner, K.R.,Maunder, M.N.,Taylor, I.,Wetzel, C.R.,Doering, K.,Johnson, K.F., and Methot, R.D. (2021). A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research 240, 105959. https://doi.org/10.1016/j.fishres.2021.105959.

Chang, Y.J., Lee, H.H., Brodziak, J., DiNardo, G., and Sun, C.L., (2013). A Bayesian hierarchical meta-analysis of blue marlin (*Makaira nigricans*) growth in the Pacific Ocean. ISC/13/BILLWG-1/02

Chang, Y.J., Shimose, T., Kanaiwa, Miyuki, Chang, X.B., Masubuchi, T. Yamamoto, K, and Kanaiwa Minoru. (2020). Estimation of the two-stanza growth curves with ageing uncertainty for the Pacific blue marlin (*Makaira nigricans*). ISC/20/BILLWG-02/03.

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

Hastie, T. (2018) gam: Generalized Additive Models, Version 1.15. Available at: https://CRAN.R-project.org/package=gam

Hsu, J. and Chang, Y.J. (2021). CPUE standardization of blue marlin (*Makaira nigricans*) for the Taiwanese distant-water tuna longline fishery in the Pacific Ocean during 1971-2019. ISC/21/BILLWG-03/03.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A. Whitten, A.R., and Punt, A.E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72: 99-110.

Ijima, H. (2021). Update Japanese longline abundance index of Pacific blue marlin (*Makaira nigricans*) estimate by the habitat model. ISC/21/BILLWG-03/01.

Ijima, H. and Koike, H. (2021). Preliminary analysis for the CPUE standardization of the Pacific blue marlin using Japanese longline logbook and R software package R-INLA. ISC/21/BILLWG-03/02.

International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific [ISC]. (2016). Annex 10: Stock assessment update for blue marlin (*Makaira nigricans*) in the Pacific Ocean through 2014, Report of the Billfish Working Group Stock Assessment Workshop. July, Sapporo, Japan. ISC/SAR/BUM/2016. Available at: http://isc.fra.go.jp/pdf/ISC16/ISC16_Annex_10_Stock_Assessment_Update_for_Blue_Marlin_in_the_Pacific_Ocean_through_2014(ISC2016).pdf

International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific [ISC Billfish WG]. (2021). Report of the Billfish Working Group Workshop, 6-7th, 10, 13th November 2020. Honolulu, HI, USA. ISC/21/BILLWG-01/REPORT.

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J-M., Garcia, D., Hillary, R., Jardim, E., Mardle, S., Pastoors, M.A., Poos, J.J., Scott, F., and Scott, R.D. (2007). FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science, 64 (4): 640-646. Doi: 10.1093/icesjms/fsm012

Lee, H.H and Chang, Y.J., (2013). Age-structured Natural Mortality for Pacific Blue Marlin Based on Meta- analysis and an Ad Hoc Model. ISC/13/BILLWG-1/07

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.

Maunder, M. N., and Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science, 72: 7-18.

Methot, R.D. and Wetzel, C.R. (2013). Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142, 86–99.

Pennington, M., Burmeister, L. M., and Hjellvik, V. (2002). Assessing the precision of frequency distributions estimated from trawl-survey samples. Fishery Bulletin, US, 100: 74–81.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sculley, M. and Brodziak, J. (2021). Standardization of Pacific blue marlin catch per unit effort in the Hawaii longline fishery from 1995-2019. ISC/21/BILL-03/04.

Sun, C.L., Chang, Y.J., Tszeng, C.C., Yeh, S.Z., and Su, N.J., (2009). Reproductive biology of blue marlin (*Makaira nigricans*) in the western Pacific Ocean. Fishery Bulletin. 107:420-432.

Wald, A., and Wolfowitz, J. (1940). On a test whether two samples are from the same population. Annals of Mathematical Statistics, Institute of Mathematical Statistics11: 147-162.

Walter, J., Hiroki, Y., Satoh, K., Matsumoto, T., Winker, H., Urtizberea Ijurco, A., and Schirripa., M. (2019). Atlantic bigeye tuna Stock Synthesis projections and Kobe 2 matrices. Collected Volumes of Scientific Papers, ICCAT, 75(7): 2283-2300.

Walter J. and Winker H. (2020). Projections to create Kobe 2 strategy matrix using the multivariate log-normal approximation for Atlantic yellowfin tuna. Collected Volumes of Scientific Papers, ICCAT, 76(8): 725-739.

Williams, S.M., Wyatt, J., and Ovenden, J.R., (2020). Investigating the genetic stock structure of blue marlin (*Makaira nigricans*) in the Pacific Ocean. Fisheries Research 228, 105565. https://doi.org/10.1016/j.fishres.2020.105565.

Winker H., Kell L., Fu D., Sharma R., Courtney D., Carvalho F., Schirripa M., and Walter J. (2019). A rapid approach to approximate Kobe posteriors from Stock Synthesis assessment models with applications to north Atlantic shortfin mako. ICCAT SCRS/2019/093.

Table 1. Descriptions of fisheries catch and abundance indices included in the base case model for the stock assessment including fishing countries, time period, and reference sources for CPUE standardizations.

Catch Index	Abundance Index	Fleet Name	Time Period	Source
F1	S1	JPNEarlyLL	1971-1993	Ijima (2020)
F2	S2	JPNLateLL	1994-2019	Ijima (2020)
F3	-	JPNCLL	1971-2019	
F4	-	JPNDRIFT	1971-2019	
F5	-	JPNBait	1971-2019	
F6	-	JPNOth	1971-2019	
F7	S3	HWLL	1991-2019	Sculley and Brodziak (2020)
F8	-	ASLL	1996-2019	
F9	-	HWOth	1975-2017	
F10	S4, S5, S6	TWNLL	1987-2019	Hsu and Chang (2020)
F11	-	TWNOth	1971-2019	
F12	-	OthLL	1971-2019	
F13	-	PYFLL	1971-2019	
F14	-	EPOPS	1990-2019	
F15	-	WCPFCPS	1993-2019	
F16	-	WCPFCOth	1971-2019	

Table 2. Time series of catch by fleet submitted for the 2021 North Pacific blue marlin stock assessment in metric tons. See Table 1 for an explanation of fleet numbers.

		Fleet							
Year	**Quarter**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1971	1	1897.7	-	28.2	0.0	1.5	12.3	-	-
1971	2	1667.4	-	28.2	0.0	1.5	12.3	-	-
1971	3	1894.5	-	28.2	0.0	1.5	12.3	-	-
1971	4	1404.1	-	28.2	0.0	1.5	12.3	-	-
1972	1	2546.8	-	52.8	2.0	1.7	13.0	-	-
1972	2	2241.3	-	52.8	2.0	1.7	13.0	-	-
1972	3	2123.2	-	52.8	2.0	1.7	13.0	-	-
1972	4	1581.5	-	52.8	2.0	1.7	13.0	-	-
1973	1	2855.0	-	52.8	65.9	5.7	33.5	-	-
1973	2	2606.6	-	52.8	65.9	5.7	33.5	-	-
1973	3	1661.1	-	52.8	65.9	5.7	33.5	-	-
1973	4	2001.9	-	52.8	65.9	5.7	33.5	-	-
1974	1	2493.9	-	45.3	56.6	15.2	13.0	-	-
1974	2	2081.2	-	45.3	56.6	15.2	13.0	-	-
1974	3	1740.5	-	45.3	56.6	15.2	13.0	-	-
1974	4	1757.2	-	45.3	56.6	15.2	13.0	-	-
1975	1	1585.3	-	116.1	195.5	36.4	20.4	-	-
1975	2	1269.2	-	116.1	195.5	36.4	20.4	-	-

		Fleet							
Year	**Quarter**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1975	3	1614.8	-	116.1	195.5	36.4	20.4	-	-
1975	4	1187.9	-	116.1	195.5	36.4	20.4	-	-
1976	1	1469.6	-	106.1	142.9	49.9	80.7	-	-
1976	2	1817.9	-	106.1	142.9	49.9	80.7	-	-
1976	3	2050.3	-	106.1	142.9	49.9	80.7	-	-
1976	4	1807.7	-	106.1	142.9	49.9	80.7	-	-
1977	1	2100.5	-	129.2	245.5	47.7	38.6	-	-
1977	2	2033.5	-	129.2	245.5	47.7	38.6	-	-
1977	3	1838.0	-	129.2	245.5	47.7	38.6	-	-
1977	4	1877.5	-	129.2	245.5	47.7	38.6	-	-
1978	1	2329.9	-	206.6	217.4	49.2	98.4	-	-
1978	2	2629.8	-	206.6	217.4	49.2	98.4	-	-
1978	3	2129.0	-	206.6	217.4	49.2	98.4	-	-
1978	4	1705.3	-	206.6	217.4	49.2	98.4	-	-
1979	1	2269.3	-	186.9	126.2	41.3	66.4	-	-
1979	2	2761.7	-	186.9	126.2	41.3	66.4	-	-
1979	3	2148.6	-	186.9	126.2	41.3	66.4	-	-
1979	4	2184.3	-	186.9	126.2	41.3	66.4	-	-
1980	1	3410.3	-	170.9	213.5	34.4	29.5	-	-
1980	2	2755.6	-	170.9	213.5	34.4	29.5	-	-
1980	3	2145.3	-	170.9	213.5	34.4	29.5	-	-
1980	4	2075.4	-	170.9	213.5	34.4	29.5	-	-
1981	1	2785.4	-	199.6	286.5	46.2	36.2	-	-
1981	2	3085.0	-	199.6	286.5	46.2	36.2	-	-
1981	3	2281.9	-	199.6	286.5	46.2	36.2	-	-
1981	4	1951.2	-	199.6	286.5	46.2	36.2	-	-
1982	1	3073.8	-	175.7	234.9	42.3	61.7	-	-
1982	2	3152.1	-	175.7	234.9	42.3	61.7	-	-
1982	3	2542.3	-	175.7	234.9	42.3	61.7	-	-
1982	4	2049.5	-	175.7	234.9	42.3	61.7	-	-
1983	1	2997.2	-	257.5	229.0	56.8	109.9	-	-
1983	2	2753.7	-	257.5	229.0	56.8	109.9	-	-
1983	3	1918.1	-	257.5	229.0	56.8	109.9	-	-
1983	4	2116.5	-	257.5	229.0	56.8	109.9	-	-
1984	1	3968.5	-	317.9	59.8	45.7	107.0	-	-
1984	2	3272.0	-	317.9	59.8	45.7	107.0	-	-
1984	3	2547.4	-	317.9	59.8	45.7	107.0	-	-
1984	4	2465.4	-	317.9	59.8	45.7	107.0	-	-

		Fleet							
Year	Quarter	1	2	3	4	5	6	7	8
1985	1	3206.3	-	252.4	98.6	74.5	90.8	-	-
1985	2	2718.3	-	252.4	98.6	74.5	90.8	-	-
1985	3	1665.4	-	252.4	98.6	74.5	90.8	-	-
1985	4	1761.9	-	252.4	98.6	74.5	90.8	-	-
1986	1	3360.8	-	218.6	43.3	91.5	38.4	-	-
1986	2	3616.6	-	218.6	43.3	91.5	38.4	-	-
1986	3	2301.7	-	218.6	43.3	91.5	38.4	-	-
1986	4	2075.9	-	218.6	43.3	91.5	38.4	-	-
1987	1	2743.7	-	371.4	63.0	70.3	32.0	-	-
1987	2	3506.6	-	371.4	63.0	70.3	32.0	-	-
1987	3	3153.7	-	371.4	63.0	70.3	32.0	-	-
1987	4	2296.0	-	371.4	63.0	70.3	32.0	-	-
1988	1	3796.3	-	353.9	89.3	57.3	37.6	-	-
1988	2	2883.8	-	353.9	89.3	57.3	37.6	-	-
1988	3	1952.4	-	353.9	89.3	57.3	37.6	-	-
1988	4	1475.8	-	353.9	89.3	57.3	37.6	-	-
1989	1	2269.0	-	306.9	72.1	97.2	33.2	-	-
1989	2	2446.9	-	306.9	72.1	97.2	33.2	-	-
1989	3	2100.2	-	306.9	72.1	97.2	33.2	-	-
1989	4	1931.5	-	306.9	72.1	97.2	33.2	-	-
1990	1	2357.7	-	293.0	62.0	62.5	44.3	-	-
1990	2	2171.8	-	293.0	62.0	62.5	44.3	-	-
1990	3	1316.5	-	293.0	62.0	62.5	44.3	-	-
1990	4	1868.0	-	293.0	62.0	62.5	44.3	-	-
1991	1	2417.1	-	326.6	43.8	42.3	16.7	256.2	0.0
1991	2	2675.6	-	326.6	43.8	42.3	16.7	195.6	0.0
1991	3	1468.9	-	326.6	43.8	42.3	16.7	69.8	0.0
1991	4	1774.1	-	326.6	43.8	42.3	16.7	127.4	0.0
1992	1	2769.6	-	403.3	39.6	37.6	14.3	111.0	0.0
1992	2	2748.5	-	403.3	39.6	37.6	14.3	114.2	0.0
1992	3	1790.6	-	403.3	39.6	37.6	14.3	58.2	0.0
1992	4	1599.0	-	403.3	39.6	37.6	14.3	68.7	0.0
1993	1	2621.9	-	509.2	35.9	46.7	21.9	56.0	0.0
1993	2	2704.8	-	509.2	35.9	46.7	21.9	116.6	0.0
1993	3	2026.3	-	509.2	35.9	46.7	21.9	84.6	0.0
1993	4	2111.9	-	509.2	35.9	46.7	21.9	84.5	0.0
1994	1	-	3036.5	207.6	38.6	34.9	17.5	59.2	0.0
1994	2	-	3004.1	350.6	38.6	34.9	17.5	100.6	0.0

Year	Quarter	Fleet 1	2	3	4	5	6	7	8
1994	3	-	2433.1	590.6	38.6	34.9	17.5	151.0	0.0
1994	4	-	2660.1	362.1	38.6	34.9	17.5	68.7	0.0
1995	1	-	2743.9	177.2	34.9	42.8	16.7	49.5	0.0
1995	2	-	2659.9	510.6	34.9	42.8	16.7	172.2	0.0
1995	3	-	2175.6	603.9	34.9	42.8	16.7	182.0	0.0
1995	4	-	1737.2	493.9	34.9	42.8	16.7	195.3	0.0
1996	1	-	1342.0	233.6	26.3	44.3	10.6	102.5	0.1
1996	2	-	1308.9	325.6	26.3	44.3	10.6	144.8	0.3
1996	3	-	1056.1	282.7	26.3	44.3	10.6	131.5	2.2
1996	4	-	951.5	254.8	26.3	44.3	10.6	112.0	2.3
1997	1	-	1207.9	174.1	18.7	58.3	8.6	52.2	3.6
1997	2	-	1615.1	249.5	18.7	58.3	8.6	181.8	3.2
1997	3	-	1679.5	278.4	18.7	58.3	8.6	203.4	3.3
1997	4	-	1642.9	249.0	18.7	58.3	8.6	99.1	1.2
1998	1	-	1609.2	138.0	13.5	70.6	7.1	68.0	3.8
1998	2	-	1487.6	372.9	13.5	70.6	7.1	108.4	2.5
1998	3	-	1257.3	302.7	13.5	70.6	7.1	123.7	4.7
1998	4	-	1067.8	275.0	13.5	70.6	7.1	128.6	2.6
1999	1	-	1167.4	223.7	18.9	42.6	3.0	72.0	1.6
1999	2	-	989.2	318.6	18.9	42.6	3.0	126.2	2.6
1999	3	-	997.0	264.8	18.9	42.6	3.0	110.8	4.6
1999	4	-	934.6	283.3	18.9	42.6	3.0	55.6	3.4
2000	1	-	1003.6	183.0	5.2	48.5	8.1	28.5	4.4
2000	2	-	797.1	414.1	5.2	48.5	8.1	76.2	6.2
2000	3	-	1198.4	336.2	5.2	48.5	8.1	137.5	4.9
2000	4	-	1025.0	275.0	5.2	48.5	8.1	82.7	2.5
2001	1	-	924.6	150.8	39.8	33.9	5.4	26.7	1.7
2001	2	-	991.1	455.4	39.8	33.9	5.4	93.4	4.2
2001	3	-	1091.7	282.3	39.8	33.9	5.4	152.4	2.4
2001	4	-	1054.1	266.8	39.8	33.9	5.4	130.6	5.3
2002	1	-	1098.6	151.6	26.1	37.1	6.9	60.4	19.7
2002	2	-	1036.7	347.3	26.1	37.1	6.9	81.7	5.8
2002	3	-	842.4	205.9	26.1	37.1	6.9	69.5	3.7
2002	4	-	811.7	148.3	26.1	37.1	6.9	43.7	2.9
2003	1	-	1235.8	158.7	9.1	43.8	5.2	45.9	2.0
2003	2	-	947.8	288.9	9.1	43.8	5.2	141.0	2.9
2003	3	-	712.4	276.9	9.1	43.8	5.2	60.2	2.3
2003	4	-	811.8	252.8	9.1	43.8	5.2	114.1	2.9

						Fleet			
Year	**Quarter**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2004	1	-	1043.6	229.5	4.9	48.0	8.4	72.3	1.7
2004	2	-	747.1	394.2	4.9	48.0	8.4	84.3	0.5
2004	3	-	693.0	352.8	4.9	48.0	8.4	67.4	0.8
2004	4	-	911.5	162.8	4.9	48.0	8.4	58.0	1.6
2005	1	-	1111.7	108.9	9.1	48.0	5.9	64.0	5.5
2005	2	-	697.3	305.1	9.1	48.0	5.9	134.2	2.3
2005	3	-	639.7	301.3	9.1	48.0	5.9	67.9	1.5
2005	4	-	437.7	264.4	9.1	48.0	5.9	69.8	12.2
2006	1	-	589.7	133.4	7.9	34.7	5.4	46.3	5.1
2006	2	-	719.1	292.7	7.9	34.7	5.4	121.5	4.9
2006	3	-	600.2	306.0	7.9	34.7	5.4	131.7	4.4
2006	4	-	597.1	256.4	7.9	34.7	5.4	114.3	6.0
2007	1	-	786.9	156.4	18.7	39.8	8.1	41.4	13.3
2007	2	-	537.5	356.9	18.7	39.8	8.1	103.4	4.4
2007	3	-	452.4	283.1	18.7	39.8	8.1	61.3	1.8
2007	4	-	388.4	307.2	18.7	39.8	8.1	67.0	12.2
2008	1	-	510.5	175.7	7.9	49.9	11.8	53.8	13.0
2008	2	-	525.5	359.3	7.9	49.9	11.8	138.7	6.8
2008	3	-	429.6	361.5	7.9	49.9	11.8	70.4	2.5
2008	4	-	377.3	250.3	7.9	49.9	11.8	95.1	6.2
2009	1	-	550.1	198.5	14.3	39.4	8.6	52.9	8.9
2009	2	-	396.8	402.5	14.3	39.4	8.6	139.0	8.2
2009	3	-	398.2	294.1	14.3	39.4	8.6	112.1	3.2
2009	4	-	582.0	199.3	14.3	39.4	8.6	48.7	6.0
2010	1	-	704.7	142.9	23.4	55.6	8.4	32.9	6.4
2010	2	-	658.3	558.9	23.4	55.6	8.4	117.3	8.4
2010	3	-	452.9	480.6	23.4	55.6	8.4	97.2	3.6
2010	4	-	420.7	299.8	23.4	55.6	8.4	47.5	9.1
2011	1	-	584.1	178.5	25.1	58.5	10.8	80.8	5.2
2011	2	-	564.5	479.1	25.1	58.5	10.8	130.1	7.4
2011	3	-	444.3	309.4	25.1	58.5	10.8	75.2	4.0
2011	4	-	370.2	225.0	25.1	58.5	10.8	83.7	5.4
2012	1	-	445.6	158.0	11.8	60.5	19.7	41.8	8.2
2012	2	-	458.4	359.0	11.8	60.5	19.7	120.9	5.6
2012	3	-	462.1	329.6	11.8	60.5	19.7	61.4	4.1
2012	4	-	471.5	155.1	11.8	60.5	19.7	68.8	9.4
2013	1	-	541.3	173.1	3.4	43.3	19.9	51.8	1.9
2013	2	-	532.1	445.4	3.4	43.3	19.9	124.8	4.7
2013	3	-	427.5	347.0	3.4	43.3	19.9	102.6	3.2

Year	Quarter	Fleet 1	2	3	4	5	6	7	8
2013	4	-	288.0	208.7	3.4	43.3	19.9	109.2	11.0
2014	1	-	453.7	163.3	2.5	43.8	9.1	82.7	4.5
2014	2	-	492.9	322.7	2.5	43.8	9.1	185.5	6.0
2014	3	-	398.8	256.2	2.5	43.8	9.1	112.7	6.6
2014	4	-	350.7	116.6	2.5	43.8	9.1	118.8	7.3
2015	1	-	431.6	102.6	6.6	43.0	14.5	79.4	5.1
2015	2	-	403.7	247.6	6.6	43.0	14.5	240.2	7.8
2015	3	-	351.3	236.0	6.6	43.0	14.5	78.6	4.4
2015	4	-	334.2	166.3	6.6	43.0	14.5	198.4	6.8
2016	1	-	663.3	107.1	4.2	39.8	12.5	126.9	8.3
2016	2	-	476.9	355.3	4.2	39.8	12.5	162.7	7.3
2016	3	-	275.6	237.9	4.2	39.8	12.5	99.8	3.5
2016	4	-	303.9	198.6	4.2	39.8	12.5	146.9	10.0
2017	1	-	384.0	101.2	3.7	30.3	12.5	134.1	9.6
2017	2	-	366.2	164.7	3.7	30.3	12.5	208.6	9.7
2017	3	-	309.7	248.0	3.7	30.3	12.5	169.4	4.9
2017	4	-	237.2	178.8	3.7	30.3	12.5	145.3	13.1
2018	1	-	292.9	77.2	1.2	43.0	9.3	141.3	10.1
2018	2	-	309.9	224.9	1.2	43.0	9.3	210.7	2.9
2018	3	-	245.1	190.6	1.2	43.0	9.3	128.7	5.1
2018	4	-	179.1	166.0	1.2	43.0	9.3	160.0	9.1
2019	1	-	246.7	82.2	1.2	43.0	9.3	155.7	8.8
2019	2	-	318.2	302.3	1.2	43.0	9.3	299.7	9.4
2019	3	-	221.1	266.1	1.2	43.0	9.3	196.2	4.9
2019	4	-	139.0	184.3	1.2	43.0	9.3	231.3	4.0

		Fleet							
Year	**Quarter**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1971	1	-	26	483.75	15	-	-	2	40.75
1971	2	-	26	483.75	15	-	-	2	40.75
1971	3	-	26	483.75	15	-	-	2	40.75
1971	4	-	26	483.75	15	-	-	2	40.75
1972	1	-	50.75	439.75	15.75	-	-	2.25	42.5
1972	2	-	50.75	439.75	15.75	-	-	2.25	42.5
1972	3	-	50.75	439.75	15.75	-	-	2.25	42.5
1972	4	-	50.75	439.75	15.75	-	-	2.25	42.5
1973	1	-	56.25	550.5	18.75	-	-	3.5	50.75
1973	2	-	56.25	550.5	18.75	-	-	3.5	50.75
1973	3	-	56.25	550.5	18.75	-	-	3.5	50.75
1973	4	-	56.25	550.5	18.75	-	-	3.5	50.75
1974	1	-	40.25	662.5	21.75	-	-	1.75	58.75
1974	2	-	40.25	662.5	21.75	-	-	1.75	58.75
1974	3	-	40.25	662.5	21.75	-	-	1.75	58.75
1974	4	-	40.25	662.5	21.75	-	-	1.75	58.75
1975	1	-	37	814.75	24.28	-	-	1.75	60
1975	2	-	37	814.75	24.28	-	-	1.75	60
1975	3	-	37	814.75	24.28	-	-	1.75	60
1975	4	-	37	814.75	24.28	-	-	1.75	60
1976	1	-	44	493.25	193.7	-	-	1.5	51.25
1976	2	-	44	493.25	193.7	-	-	1.5	51.25
1976	3	-	44	493.25	193.7	-	-	1.5	51.25
1976	4	-	44	493.25	193.7	-	-	1.5	51.25
1977	1	-	36.25	421.75	167.6	-	-	2.25	71
1977	2	-	36.25	421.75	167.6	-	-	2.25	71
1977	3	-	36.25	421.75	167.6	-	-	2.25	71
1977	4	-	36.25	421.75	167.6	-	-	2.25	71
1978	1	-	15.75	505	288.7	-	-	2	45.25
1978	2	-	15.75	505	288.7	-	-	2	45.25
1978	3	-	15.75	505	288.7	-	-	2	45.25
1978	4	-	15.75	505	288.7	-	-	2	45.25
1979	1	-	105.5	543.5	360.1	-	-	3.25	57
1979	2	-	105.5	543.5	360.1	-	-	3.25	57
1979	3	-	105.5	543.5	360.1	-	-	3.25	57
1979	4	-	105.5	543.5	360.1	-	-	3.25	57
1980	1	-	122.5	445.75	303	-	-	3.25	51.5
1980	2	-	122.5	445.75	303	-	-	3.25	51.5
1980	3	-	122.5	445.75	303	-	-	3.25	51.5
1980	4	-	122.5	445.75	303	-	-	3.25	51.5

Year	Quarter	Fleet 9	10	11	12	13	14	15	16
1981	1	-	115.75	557.75	311.9	-	-	7.5	56.75
1981	2	-	115.75	557.75	311.9	-	-	7.5	56.75
1981	3	-	115.75	557.75	311.9	-	-	7.5	56.75
1981	4	-	115.75	557.75	311.9	-	-	7.5	56.75
1982	1	-	76	640.5	398.3	-	-	10.5	53.5
1982	2	-	76	640.5	398.3	-	-	10.5	53.5
1982	3	-	76	640.5	398.3	-	-	10.5	53.5
1982	4	-	76	640.5	398.3	-	-	10.5	53.5
1983	1	-	68	753.75	245.2	-	-	16.75	60.25
1983	2	-	68	753.75	245.2	-	-	16.75	60.25
1983	3	-	68	753.75	245.2	-	-	16.75	60.25
1983	4	-	68	753.75	245.2	-	-	16.75	60.25
1984	1	-	95.5	720.5	381.3	-	-	21.5	111.75
1984	2	-	95.5	720.5	381.3	-	-	21.5	111.75
1984	3	-	95.5	720.5	381.3	-	-	21.5	111.75
1984	4	-	95.5	720.5	381.3	-	-	21.5	111.75
1985	1	-	53	499.25	374.9	-	-	17.25	130
1985	2	-	53	499.25	374.9	-	-	17.25	130
1985	3	-	53	499.25	374.9	-	-	17.25	130
1985	4	-	53	499.25	374.9	-	-	17.25	130
1986	1	-	46	690.75	409	-	-	16.5	137.75
1986	2	-	46	690.75	409	-	-	16.5	137.75
1986	3	-	46	690.75	409	-	-	16.5	137.75
1986	4	-	46	690.75	409	-	-	16.5	137.75
1987	1	70.75	49.5	1403.25	1119	-	-	18.25	116.25
1987	2	70.75	49.5	1403.25	1119	-	-	18.25	116.25
1987	3	70.75	49.5	1403.25	1119	-	-	18.25	116.25
1987	4	70.75	49.5	1403.25	1119	-	-	18.25	116.25
1988	1	74	80	1024.25	936.9	-	-	17.75	129.25
1988	2	74	80	1024.25	936.9	-	-	17.75	129.25
1988	3	74	80	1024.25	936.9	-	-	17.75	129.25
1988	4	74	80	1024.25	936.9	-	-	17.75	129.25
1989	1	91.25	111.25	829.25	682.1	-	-	21.5	134.5
1989	2	91.25	111.25	829.25	682.1	-	-	21.5	134.5
1989	3	91.25	111.25	829.25	682.1	-	-	21.5	134.5
1989	4	91.25	111.25	829.25	682.1	-	-	21.5	134.5
1990	1	84.25	109.25	581.75	745.1	0.75	-	23.75	180.5
1990	2	84.25	109.25	581.75	745.1	0.75	-	23.75	180.5
1990	3	84.25	109.25	581.75	745.1	0.75	-	23.75	180.5

Year	Quarter	Fleet 9	10	11	12	13	14	15	16
1990	4	84.25	109.25	581.75	745.1	0.75	-	23.75	180.5
1991	1	96.75	180	674	761.9	5.75	-	33.75	222
1991	2	96.75	180	674	761.9	5.75	-	33.75	222
1991	3	96.75	180	674	761.9	5.75	-	33.75	222
1991	4	96.75	180	674	761.9	5.75	-	33.75	222
1992	1	75.25	30.5	1095	1004	19.25	-	35.25	200
1992	2	75.25	30.5	1095	1004	19.25	-	35.25	200
1992	3	75.25	30.5	1095	1004	19.25	-	35.25	200
1992	4	75.25	30.5	1095	1004	19.25	-	35.25	200
1993	1	84.75	112.25	1110.75	985.4	51.25	21.25	35.5	174.25
1993	2	84.75	112.25	1110.75	985.4	51.25	21.25	35.5	174.25
1993	3	84.75	112.25	1110.75	985.4	51.25	21.25	35.5	174.25
1993	4	84.75	112.25	1110.75	985.4	51.25	21.25	35.5	174.25
1994	1	83.5	150.75	815.5	1038	87.25	17.5	35.25	248.75
1994	2	83.5	150.75	815.5	1038	87.25	17.5	35.25	248.75
1994	3	83.5	150.75	815.5	1038	87.25	17.5	35.25	248.75
1994	4	83.5	150.75	815.5	1038	87.25	17.5	35.25	248.75
1995	1	87.75	81.5	1192.75	1318	104	17.5	36	272.5
1995	2	87.75	81.5	1192.75	1318	104	17.5	36	272.5
1995	3	87.75	81.5	1192.75	1318	104	17.5	36	272.5
1995	4	87.75	81.5	1192.75	1318	104	17.5	36	272.5
1996	1	110.25	46.75	906.5	918.5	105.5	15	40	402.75
1996	2	110.25	46.75	906.5	918.5	105.5	15	40	402.75
1996	3	110.25	46.75	906.5	918.5	105.5	15	40	402.75
1996	4	110.25	46.75	906.5	918.5	105.5	15	40	402.75
1997	1	105.5	26	977.5	1054	84.25	31.5	44.75	361.5
1997	2	105.5	26	977.5	1054	84.25	31.5	44.75	361.5
1997	3	105.5	26	977.5	1054	84.25	31.5	44.75	361.5
1997	4	105.5	26	977.5	1054	84.25	31.5	44.75	361.5
1998	1	66	52.25	940.5	1479	76.75	32.25	45.5	375.75
1998	2	66	52.25	940.5	1479	76.75	32.25	45.5	375.75
1998	3	66	52.25	940.5	1479	76.75	32.25	45.5	375.75
1998	4	66	52.25	940.5	1479	76.75	32.25	45.5	375.75
1999	1	83	32.75	888	1535	88.75	45.5	38.25	408.25
1999	2	83	32.75	888	1535	88.75	45.5	38.25	408.25
1999	3	83	32.75	888	1535	88.75	45.5	38.25	408.25
1999	4	83	32.75	888	1535	88.75	45.5	38.25	408.25
2000	1	58.75	28.5	1997.25	1330	65.25	29.75	46	198
2000	2	58.75	28.5	1997.25	1330	65.25	29.75	46	198

		Fleet							
Year	**Quarter**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
2000	3	58.75	28.5	1997.25	1330	65.25	29.75	46	198
2000	4	58.75	28.5	1997.25	1330	65.25	29.75	46	198
2001	1	73.5	146.25	2257.5	1634	66.25	29.75	47.25	199
2001	2	73.5	146.25	2257.5	1634	66.25	29.75	47.25	199
2001	3	73.5	146.25	2257.5	1634	66.25	29.75	47.25	199
2001	4	73.5	146.25	2257.5	1634	66.25	29.75	47.25	199
2002	1	57.75	123.75	2199.75	1812	63.75	47.5	51.25	200.25
2002	2	57.75	123.75	2199.75	1812	63.75	47.5	51.25	200.25
2002	3	57.75	123.75	2199.75	1812	63.75	47.5	51.25	200.25
2002	4	57.75	123.75	2199.75	1812	63.75	47.5	51.25	200.25
2003	1	52.75	301.75	1940	2517	75.75	46	53.5	225.25
2003	2	52.75	301.75	1940	2517	75.75	46	53.5	225.25
2003	3	52.75	301.75	1940	2517	75.75	46	53.5	225.25
2003	4	52.75	301.75	1940	2517	75.75	46	53.5	225.25
2004	1	47	364	1643	2037	60.75	35.25	56.5	224.75
2004	2	47	364	1643	2037	60.75	35.25	56.5	224.75
2004	3	47	364	1643	2037	60.75	35.25	56.5	224.75
2004	4	47	364	1643	2037	60.75	35.25	56.5	224.75
2005	1	47	376.5	1885	2250	62.75	52.25	212	237
2005	2	47	376.5	1885	2250	62.75	52.25	212	237
2005	3	47	376.5	1885	2250	62.75	52.25	212	237
2005	4	47	376.5	1885	2250	62.75	52.25	212	237
2006	1	40.25	419.5	1452	2155	66.5	40.75	152.75	246.75
2006	2	40.25	419.5	1452	2155	66.5	40.75	152.75	246.75
2006	3	40.25	419.5	1452	2155	66.5	40.75	152.75	246.75
2006	4	40.25	419.5	1452	2155	66.5	40.75	152.75	246.75
2007	1	32.25	317.75	1290.25	1635	81.75	30.75	206	261.75
2007	2	32.25	317.75	1290.25	1635	81.75	30.75	206	261.75
2007	3	32.25	317.75	1290.25	1635	81.75	30.75	206	261.75
2007	4	32.25	317.75	1290.25	1635	81.75	30.75	206	261.75
2008	1	45.25	227.5	1380.75	1552	56	31.25	148	261.75
2008	2	45.25	227.5	1380.75	1552	56	31.25	148	261.75
2008	3	45.25	227.5	1380.75	1552	56	31.25	148	261.75
2008	4	45.25	227.5	1380.75	1552	56	31.25	148	261.75
2009	1	45	334.5	1196.75	1714	55.75	39.25	144.75	197
2009	2	45	334.5	1196.75	1714	55.75	39.25	144.75	197
2009	3	45	334.5	1196.75	1714	55.75	39.25	144.75	197
2009	4	45	334.5	1196.75	1714	55.75	39.25	144.75	197
2010	1	37.5	372.5	1435.5	1692	65	44	161	235.5

		Fleet							
Year	**Quarter**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
2010	2	37.5	372.5	1435.5	1692	65	44	161	235.5
2010	3	37.5	372.5	1435.5	1692	65	44	161	235.5
2010	4	37.5	372.5	1435.5	1692	65	44	161	235.5
2011	1	50.25	332.75	1278	1560	50.25	37.5	188	242.25
2011	2	50.25	332.75	1278	1560	50.25	37.5	188	242.25
2011	3	50.25	332.75	1278	1560	50.25	37.5	188	242.25
2011	4	50.25	332.75	1278	1560	50.25	37.5	188	242.25
2012	1	35.75	321	1235	1984	60.25	44.25	229.5	362.25
2012	2	35.75	321	1235	1984	60.25	44.25	229.5	362.25
2012	3	35.75	321	1235	1984	60.25	44.25	229.5	362.25
2012	4	35.75	321	1235	1984	60.25	44.25	229.5	362.25
2013	1	34.75	263.75	1407.75	2169	60.75	43	236.5	416
2013	2	34.75	263.75	1407.75	2169	60.75	43	236.5	416
2013	3	34.75	263.75	1407.75	2169	60.75	43	236.5	416
2013	4	34.75	263.75	1407.75	2169	60.75	43	236.5	416
2014	1	41	306.25	1451.5	2436	60	52	217.25	546.5
2014	2	41	306.25	1451.5	2436	60	52	217.25	546.5
2014	3	41	306.25	1451.5	2436	60	52	217.25	546.5
2014	4	41	306.25	1451.5	2436	60	52	217.25	546.5
2015	1	49.75	507	1243.25	2251	60.25	76.25	304.5	496.25
2015	2	49.75	507	1243.25	2251	60.25	76.25	304.5	496.25
2015	3	49.75	507	1243.25	2251	60.25	76.25	304.5	496.25
2015	4	49.75	507	1243.25	2251	60.25	76.25	304.5	496.25
2016	1	40.75	442	1036.25	2076	54.75	62	233.75	1122.3
2016	2	40.75	442	1036.25	2076	54.75	62	233.75	1122.3
2016	3	40.75	442	1036.25	2076	54.75	62	233.75	1122.3
2016	4	40.75	442	1036.25	2076	54.75	62	233.75	1122.3
2017	1	39.5	314.75	1192	1721	40.75	37.75	140.25	598.5
2017	2	39.5	314.75	1192	1721	40.75	37.75	140.25	598.5
2017	3	39.5	314.75	1192	1721	40.75	37.75	140.25	598.5
2017	4	39.5	314.75	1192	1721	40.75	37.75	140.25	598.5
2018	1	41.75	227.5	999.25	1840	56	41.5	180.25	618.75
2018	2	41.75	227.5	999.25	1840	56	41.5	180.25	618.75
2018	3	41.75	227.5	999.25	1840	56	41.5	180.25	618.75
2018	4	41.75	227.5	999.25	1840	56	41.5	180.25	618.75
2019	1	45.75	171.25	980	1339	70	41.5	148	609.5
2019	2	45.75	171.25	980	1339	70	41.5	148	609.5
2019	3	45.75	171.25	980	1339	70	41.5	148	609.5
2019	4	45.75	171.25	980	1339	70	41.5	148	609.5

Table 3. List of fleets with catch used in the assessment model along with CPUE indices provided for the 2021 Pacific blue marlin stock assessment, their source, and whether the indices were used in the assessment model.

Length Comp – Used?	**Relative Abundance Index – Used?**	**Fleet Name**	**Time Series**	**Source**
F1 – Y	S1 – Y	JPNEarlyLL	1971-1993	Ijima (2020)
F2 - Y	S2 - Y	JPNLateLL	1994-2019	Ijima (2020)
F3 - N	-	JPNCLL	1971-2019	Hirotaka Ijima, pers. comm.
F4 – Y	-	JPNDRIFT	1971-2019	Hirotaka Ijima, pers. comm.
F5 – N	-	JPNBait	1971-2019	Hirotaka Ijima, pers. comm.
F6 – N	-	JPNOth	1971-2019	Hirotaka Ijima, pers. comm.
F7 – Y	S3 - N	HWLL	1991-2019	Sculley and Brodziak (2020)
F8 – N	-	ASLL	1996-2019	Russell Ito, pers. comm.
F9 – N	-	HWOth	1975-2017	Russell Ito, pers. comm.
F10 – Y	S4, S5, S6 - Y	TWNLL	1987-2019	Hsu and Chang (2020)
F11 – N	-	TWNOth	1971-2019	Yi-Jay Chang, pers. comm.
F12 – Y	-	OthLL	1971-2019	Peter Williams, pers. comm.
F13 – N	-	PYFLL	1971-2019	Peter Williams, pers. comm.
F14 – Y	-	EPOPS	1990-2019	Shane Griffiths, pers. comm.
F15 – N	-	WCPFCPS	1993-2019	Peter Williams, pers. comm.
F16 – N	-	WCPFCOth	1971-2019	Peter Williams, pers. comm.

Table 4. Standardized catch-per-unit-effort (CPUE; in number per 1000 hooks) indices and input standard error (SE) in log-scale (i.e., log(SE)) of lognormal error of CPUE for Pacific blue marlin used in the stock assessment. Index descriptions can be found in Table 3.

Fleet	**S1**		**S2**		**S3**		**S4**		**S5**		**S6**	
Year	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**
1971	-	-	-	-	-	-	0.689	0.40	-	-	-	-
1972	-	-	-	-	-	-	1.145	0.38	-	-	-	-
1973	-	-	-	-	-	-	0.912	0.37	-	-	-	-
1974	-	-	-	-	-	-	0.721	0.37	-	-	-	-
1975	0.33	0.19	-	-	-	-	0.965	0.39	-	-	-	-
1976	0.33	0.19	-	-	-	-	1.377	0.42	-	-	-	-
1977	0.25	0.19	-	-	-	-	0.581	0.40	-	-	-	-
1978	0.40	0.19	-	-	-	-	0.696	0.44	-	-	-	-
1979	0.46	0.20	-	-	-	-	-	-	1.926	0.29	-	-
1980	0.47	0.20	-	-	-	-	-	-	1.6	0.28	-	-
1981	0.55	0.20	-	-	-	-	-	-	1.255	0.27	-	-
1982	0.55	0.20	-	-	-	-	-	-	1.127	0.31	-	-
1983	0.44	0.20	-	-	-	-	-	-	1.019	0.35	-	-
1984	0.70	0.21	-	-	-	-	-	-	1.287	0.34	-	-
1985	0.48	0.20	-	-	-	-	-	-	1.313	0.36	-	-
1986	0.49	0.20	-	-	-	-	-	-	1.131	0.38	-	-
1987	0.48	0.20	-	-	-	-	-	-	0.674	0.36	-	-
1988	0.46	0.20	-	-	-	-	-	-	0.857	0.39	-	-
1989	0.48	0.20	-	-	-	-	-	-	1.09	0.37	-	-
1990	0.46	0.20	-	-	-	-	-	-	0.826	0.38	-	-
1991	0.44	0.20	-	-	-	-	-	-	0.846	0.37	-	-
1992	0.45	0.20	-	-	-	-	-	-	0.432	0.52	-	-
1993	0.57	0.20	-	-	-	-	-	-	1.058	0.41	-	-
1994	-	-	1.43	0.19	-	-	-	-	1.721	0.34	-	-
1995	-	-	1.35	0.19	1.21	0.21	-	-	0.911	0.38	-	-
1996	-	-	0.69	0.19	1.06	0.22	-	-	0.768	0.38	-	-
1997	-	-	2.01	0.19	1.01	0.22	-	-	0.3	0.40	-	-
1998	-	-	1.13	0.19	0.98	0.22	-	-	0.342	0.46	-	-
1999	-	-	0.76	0.20	0.90	0.23	-	-	0.517	0.43	-	-
2000	-	-	0.84	0.19	0.82	0.23	-	-	-	-	0.382	0.33
2001	-	-	0.77	0.19	0.96	0.22	-	-	-	-	0.745	0.29
2002	-	-	0.84	0.19	0.78	0.24	-	-	-	-	0.715	0.31
2003	-	-	0.99	0.19	0.89	0.23	-	-	-	-	1.008	0.27
2004	-	-	1.45	0.19	0.76	0.24	-	-	-	-	0.827	0.28
2005	-	-	0.66	0.20	0.75	0.24	-	-	-	-	1.21	0.25
2006	-	-	1.70	0.20	0.76	0.24	-	-	-	-	1.017	0.26
2007	-	-	0.58	0.20	0.66	0.25	-	-	-	-	0.948	0.27
2008	-	-	0.63	0.20	0.67	0.25	-	-	-	-	0.94	0.28

Fleet	**S1**		**S2**		**S3**		**S4**		**S5**		**S6**	
Year	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**	**CPUE**	**CV**
2009	-	-	1.31	0.20	0.66	0.25	-	-	-	-	0.937	0.26
2010	-	-	0.62	0.20	0.62	0.25	-	-	-	-	0.952	0.26
2011	-	-	0.64	0.20	0.65	0.25	-	-	-	-	0.92	0.26
2012	-	-	1.13	0.20	0.57	0.26	-	-	-	-	0.987	0.27
2013	-	-	0.73	0.20	0.58	0.26	-	-	-	-	1.196	0.27
2014	-	-	1.38	0.21	0.61	0.25	-	-	-	-	1.434	0.27
2015	-	-	1.40	0.20	0.65	0.25	-	-	-	-	1.364	0.26
2016	-	-	1.26	0.20	0.58	0.26	-	-	-	-	1.377	0.26
2017	-	-	0.70	0.20	0.59	0.26	-	-	-	-	1.127	0.26
2018	-	-	0.62	0.20	0.59	0.26	-	-	-	-	0.798	0.28
2019	-	-	0.36	0.30	0.62	0.25	-	-	-	-	1.117	0.28

Table 5. Key life history parameters and model structures for Pacific blue marlin used in the stock assessment. 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) from Table 2 in the BILLWG Data Preparatory report (ISC 2021).

Parameter	Old Growth	New Growth	Reference
Growth_Age_for_L1	1	0.5	Chang et al. (2013), Chang et al (2020)
Growth_Age_for_L2	26	20	Chang et al. (2013), Chang et al (2020) Andrews (2018)
NatM_Fem_GP_1	M_0 =0.42, M_1 = 0.37, M_{4+} =0.22	M_0 = 0.41, M_1 = 0.35, M_2 = 0.33, M_3 = 0.32, M_{4+} = 0.3	Lee and Chang (2013), Brodziak (2021)
L_at_Amin_Fem_GP_1	144	136.13	Chang et al. (2013) , Chang et al (2020)
L_at_Amax_Fem_GP_1	304.178	249.1	Chang et al. (2013), Chang et al (2020)
VonBert_K_Fem_GP_1	0.107	0.31	Chang et al. (2013), Chang et al (2020)
Richards_Fem_GP_1	NA	0.000468	Chang et al (2020)
CV_young_Fem_GP_1	0.14	0.13	Chang et al. (2013), Chang et al (2020)
CV_old_Fem_GP_1	0.15	0.15	Chang et al. (2013), Chang et al (2020)
NatM_Mal_GP_1	M_0 = 0.42, M_{1+} = 0.37	M_0 = 0.41, M_{1+} = 0.35	Lee and Chang (2013), Brodziak (2021)
L_at_Amin_Mal_GP_1	144	136.13	Chang et al. (2013), Chang et al (2020)
L_at_Amax_Mal_GP_1	226	206.4	Chang et al. (2013), Chang et al (2020)
VonBert_K_Mal_GP_1	0.211	0.18	Chang et al. (2013), Chang et al (2020)
Richards_Mal_GP_1	NA	0.000468	Chang et al (2020)
CV_young_Mal_GP_1	0.14	0.2	Chang et al. (2013), Chang et al (2020)
CV_old_Mal_GP_1	0.1	0.1	Chang et al. (2013), Chang et al (2020)
Wtlen_1_Fem	1.84E-05	1.84E-05	Brodziak 2013
Wtlen_2_Fem	2.956	2.956	Brodziak 2013
Mat50%_Fem	179.76	179.76	Sun et al. (2009)
Mat_slope_Fem	-0.2039	-0.2039	Sun et al. (2009)
Fecundity	Proportional to spawning biomass	Proportional to spawning biomass	Sun et al. (2009)
Wtlen_1_Mal	1.37E-05	1.37E-05	Brodziak 2013
Wtlen_2_Mal	2.975	2.975	Brodziak 2013
Spawning season	2	2	Sun et al. (2009)
R0	0.6	0.4	Rescaled
Steepness	0.9	0.9	Estimated

Table 6. Mean input standard error (SE) in log-space (i.e., log(SE)) of lognormal error and root-mean-square-errors (RMSE) for the relative abundance indices for Pacific blue marlin used in the models.

Model	**Fleet**	***N***	**Input log(SE)**	**RMSE**
Old Growth	S1_JPN_LL_early	19	0.20	0.20
Old Growth	S2_JPN_LL_late	26	0.20	0.39
Old Growth	S4_TWN_DWLL_Early	8	0.40	0.27
Old Growth	S5_TWN_DWLL_Mid	20	0.37	0.49
Old Growth	S6_TWN_DWLL_Late	21	0.28	0.32
New Growth	S1_JPN_LL_early	19	0.20	0.18
New Growth	S2_JPN_LL_late	26	0.20	0.41
New Growth	S4_TWN_DWLL_Early	8	0.40	0.29
New Growth	S5_TWN_DWLL_Mid	20	0.37	0.46
New Growth	S6_TWN_DWLL_Late	21	0.28	0.33

Table 7. Fishery-specific selectivity assumptions for the Pacific blue marlin stock assessment. The selectivity curves for fisheries lacking length composition data were assumed to be the same as (i.e., mirror gear) closely related fisheries or fisheries operating in the same area.

Fleet	Selectivity Function
F1 JPN LL Early	4-parameter cubic spline
F2 JPN LL Late	3-parameter cubic spline
F3 JPN CLL	Mirror F2
F4 JPN DRIFT	Double normal
F5 JPN Bait	Mirror F2
F6 JPN Oth	Mirror F2
F7 HW LL	3-parameter cubic spline
F8 AS LL	Mirror F7
F9 HW Oth	Mirror F7
F10 TWN LL	Double normal
F11 TWN Oth	Mirror F10
F12 Oth LL	Double Normal
F13 PYF LL	Mirror F12
F14 EPO PS	Asymptotic logistic
F15 WCPFC PS	Mirror F14
F16 WCPFC Oth	Mirror F10
S1 JPN LL Early	Mirror F1
S2 JPN LL Late	Mirror F2
S4 TWN LL Early	Mirror F10
S5 TWN LL Mid	Mirror F10
S6 TWN LL Late	Mirror F10

Table 8. Relative negative log-likelihoods of each data component in the old growth 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). The maximum likelihood estimate of $\log(R_0)$ was 7.102.

log(R0)	Recruitment	Length composition	Abundance Indices	Size Frequency
6	188.1	69.2	28.5	5.8
6.1	168.0	58.0	24.9	4.5
6.2	146.5	47.6	24.5	1.2
6.3	120.2	44.0	22.2	1.2
6.4	93.3	35.8	18.4	1.9
6.5	73.1	31.4	16.2	1.2
6.6	55.0	25.8	13.8	1.0
6.7	38.3	20.3	10.9	0.6
6.8	23.9	13.9	7.7	0.4
6.9	11.6	8.0	4.2	0.1
7	3.3	2.6	1.4	0
7.1	0	0	0	1.1
7.102	0	0	0	1.2
7.2	1.9	0.3	1.6	3.0
7.3	5.2	4.8	5.0	6.1
7.4	7.2	15.5	8.4	9.5
7.5	7.4	29.8	10.5	12.9
7.6	7.2	44.2	11.9	15.6
7.7	6.4	72.5	13.4	17.9
7.8	7.2	68.8	13.6	19.8
7.9	7.4	78.8	14.3	21.3
8	8.4	100.0	15.9	22.1

Table 9. Relative negative log-likelihoods of abundance index data components in the old growth 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). The maximum likelihood estimate of log(R_0) was 7.102. See Table 3 for a description of the abundance indices. S3 was not included in the total likelihood.

log(R0)	S1	S2	S4	S5	S6
6	13.1	5.4	0	2.8	13.6
6.1	13.1	2.6	0	2.4	13.1
6.2	13.0	3.5	0.1	1.9	12.4
6.3	12.5	3.0	0.1	1.5	11.4
6.4	11.4	1.3	0.2	1.7	10.2
6.5	10.7	0.6	0.3	1.4	9.5
6.6	9.3	0.4	0.4	1.1	9.0
6.7	7.2	0.2	0.5	0.7	8.5
6.8	5.4	0	0.5	0.3	7.8
6.9	3.4	0	0.4	0	6.7
7	2.3	0.5	0.3	0.1	4.6
7.1	1.2	2.1	0.2	0.2	2.6
7.102	1.2	2.1	0.2	0.2	2.6
7.2	0.4	5.4	0.2	0.4	1.6
7.3	0	9.2	0.3	0.6	1.2
7.4	0	12.5	0.3	0.7	1.2
7.5	0.2	14.3	0.3	1.0	1.0
7.6	0.4	15.4	0.3	1.3	0.8
7.7	0.6	16.3	0.3	1.8	0.7
7.8	0.8	16.5	0.3	2.1	0.2
7.9	0.9	16.8	0.3	2.6	0
8	1.1	17.4	0.3	2.9	0.5

Table 10. Relative negative log-likelihoods of length composition data components and generalized size composition data components in the old growth 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). The maximum likelihood estimate of $\log(R_0)$ was 7.102. F4 was excluded because it did not meet the threshold of likelihood contribution to include. See Table 3 for a description of the composition data.

ln(R0)	F1	F2	F7	F10	F14	F12
6	26.5	35.3	7.2	8.2	4.2	4.9
6.1	26.0	32.7	1.5	8.0	2.1	3.6
6.2	25.3	28.9	0.5	5.0	0	0.3
6.3	23.9	27.5	0.3	4.6	0	0.4
6.4	16.8	23.4	3.0	4.2	0.7	1.2
6.5	14.5	21.2	2.9	3.9	1.1	0.6
6.6	10.6	20.0	2.6	3.6	1.2	0.4
6.7	6.9	18.8	2.4	3.4	1.1	0.2
6.8	3.6	16.9	1.8	3.4	0.5	0.1
6.9	1.0	14.9	1.2	3.0	0.3	0
7	0	11.1	0.7	2.0	1.1	0
7.1	0.6	7.9	0.3	1.3	2.2	1.2
7.102	0.7	7.8	0.3	1.3	2.2	1.2
7.2	2.0	4.5	0	1.3	4.8	2.9
7.3	3.4	0.8	0	0	12.8	6.0
7.4	4.0	0	0	0.2	23.5	9.3
7.5	4.1	1.9	0.1	0.6	35.5	12.7
7.6	4.0	5.1	0.1	1.1	46.1	15.4
7.7	3.8	24.5	0.3	1.3	54.8	17.7
7.8	3.8	12.2	0.2	2.2	62.6	19.6
7.9	3.8	15.4	0.2	2.6	69.0	21.1
8	3.6	34.1	0.5	2.1	72.0	21.9

Table 11. Relative negative log-likelihoods of each data component in the new growth 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). The maximum likelihood estimate of log(R_0) was 6.92.

log(R0)	Recruitment	Length composition	Abundance Indices	Size Frequency
6	215.2	82.5	35.0	0
6.1	181.2	73.1	31.6	1.5
6.2	153.5	58.4	26.0	1.5
6.3	122.7	46.3	22.2	1.7
6.4	89.0	40.3	21.9	0.9
6.5	62.7	29.9	14.2	2.7
6.6	36.0	20.5	12.8	2.2
6.7	14.9	13.8	8.0	2.9
6.8	3.6	7.4	3.4	4.5
6.9	0.4	2.2	1.9	7.1
6.92	0.1	1.7	1.8	7.7
7	0	0	0.9	9.8
7.1	0.2	3.6	0	12.4
7.2	0.6	6.6	0.9	15.2
7.3	0.8	12.4	0.9	17.7
7.4	1.1	18.6	0.9	19.9
7.5	1.2	24.7	0.9	21.9
7.6	1.4	30.5	0.9	23.6
7.7	1.5	35.9	0.9	25.0
7.8	1.7	40.7	0.9	26.3
7.9	1.8	45.1	1.0	27.4
8	2.0	49.1	1.0	28.3

Table 12. Relative negative log-likelihoods of abundance index data components in the new growth 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). The maximum likelihood estimate of log(R_0) was 6.9. See Table 3 for a description of the abundance indices. S3 was not included in the total likelihood.

log(R0)	**S1**	**S2**	**S4**	**S5**	**S6**
6	13.8	3.9	0	6.0	19.0
6.1	12.7	6.3	0.1	4.7	15.5
6.2	11.6	2.6	0.2	4.2	15.1
6.3	10.2	1.6	0.2	3.9	14.0
6.4	11.5	0.8	0.4	4.1	12.8
6.5	5.5	0.5	0.4	3.6	11.9
6.6	6.3	0.1	0.5	3.0	10.5
6.7	4.0	0	0.5	2.3	8.8
6.8	2.2	1.1	0.5	1.2	6.0
6.9	1.3	3.3	0.6	0.2	4.0
6.92	1.2	3.9	0.6	0.1	3.7
7	0.7	3.8	0.5	0.1	3.4
7.1	0.4	3.4	0.5	0.1	3.4
7.2	0.4	5.7	0.5	0	2.0
7.3	0.3	6.2	0.5	0.1	1.6
7.4	0.2	6.5	0.4	0.3	1.2
7.5	0.1	6.7	0.4	0.4	0.9
7.6	0.1	6.8	0.4	0.6	0.6
7.7	0.1	6.9	0.4	0.8	0.4
7.8	0	6.9	0.4	1.0	0.3
7.9	0	7.0	0.4	1.2	0.1
8	0	7.0	0.4	1.3	0

Table 13. Relative negative log-likelihoods of length composition data components and generalized size composition data components in the new growth 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). The maximum likelihood estimate of $\log(R_0)$ was 6.92. F4 was excluded because it did not meet the threshold of likelihood contribution to include. See Table 3 for a description of the composition data.

ln(R0)	F1	F2	F7	F10	F14	F12
6	29.3	40.6	4.6	11.5	8.5	0
6.1	24.9	40.5	4.1	7.8	7.8	1.7
6.2	20.7	35.6	2.5	6.0	5.6	1.9
6.3	15.3	32.0	1.9	4.7	4.5	2.3
6.4	15.7	28.9	1.5	3.7	2.5	2.6
6.5	7.6	27.2	1.1	3.2	2.8	2.6
6.6	2.2	24.9	0.9	2.6	2.1	3.0
6.7	0	22.3	0.6	2.0	1.1	3.7
6.8	1.0	17.3	0.2	0.9	0	5.2
6.9	2.4	10.7	0.1	0.3	0.7	7.7
6.92	2.8	9.5	0	0.3	1.1	8.2
7	3.8	4.5	0	0	3.8	10.1
7.1	5.1	2.0	0	0.9	7.6	12.6
7.2	5.7	0	0	0.6	12.3	15.4
7.3	6.2	0	0.1	1.1	17.0	17.8
7.4	6.7	0.8	0.1	1.7	21.4	19.9
7.5	7.0	2.0	0.2	2.1	25.4	21.8
7.6	7.2	3.4	0.3	2.6	29.1	23.5
7.7	7.5	4.8	0.3	3.0	32.3	24.9
7.8	7.7	6.2	0.4	3.3	35.1	26.1
7.9	7.9	7.6	0.5	3.7	37.6	27.2
8	8.0	8.9	0.5	3.9	39.8	28.2

Table 14. Time series of female spawning biomass (metric ton), age-0 recruitment (thousands of fish), instantaneous fishing mortality (year^{-1}), and fishing intensity (1- spawning potential ratio) for Pacific blue marlin estimated in the model ensemble.

Year	Female Spawning Stock Biomass (mt)	Recruitment (1000 age-0 fish)	Instantaneous fishing mortality	1-spawning potential ratio
1972	69,232	907	0.08	0.40
1973	66,240	820	0.09	0.44
1974	63,775	499	0.08	0.43
1975	61,397	631	0.08	0.43
1976	57,236	772	0.10	0.48
1977	52,333	937	0.11	0.52
1978	47,716	995	0.12	0.56
1979	45,138	994	0.13	0.57
1980	43,914	809	0.13	0.58
1981	43,607	1013	0.14	0.60
1982	42,400	1285	0.14	0.62
1983	41,950	857	0.12	0.58
1984	42,959	882	0.16	0.63
1985	43,380	835	0.13	0.57
1986	42,789	1034	0.15	0.63
1987	39,754	1022	0.20	0.73
1988	36,597	1004	0.17	0.70
1989	36,102	1112	0.15	0.64
1990	36,876	944	0.13	0.60
1991	38,331	1056	0.14	0.63
1992	38,473	887	0.16	0.67
1993	37,788	992	0.18	0.70
1994	35,563	573	0.17	0.71
1995	32,795	1099	0.21	0.75
1996	30,131	778	0.13	0.66
1997	30,364	944	0.15	0.68
1998	29,858	582	0.15	0.70
1999	30,143	905	0.16	0.69
2000	28,295	800	0.17	0.73
2001	25,882	851	0.21	0.78
2002	23,352	847	0.21	0.79
2003	21,049	820	0.25	0.82
2004	19,540	903	0.23	0.81
2005	18,119	796	0.25	0.83
2006	17,583	1084	0.23	0.81
2007	17,842	854	0.18	0.77
2008	20,257	730	0.17	0.74

Year	Female Spawning Stock Biomass (mt)	Recruitment (1000 age-0 fish)	Instantaneous fishing mortality	1-spawning potential ratio
2009	22,454	1099	0.17	0.73
2010	22,934	1046	0.17	0.74
2011	24,608	879	0.15	0.70
2012	26,923	751	0.16	0.71
2013	27,667	960	0.18	0.74
2014	26,274	786	0.19	0.76
2015	25,484	610	0.19	0.75
2016	23,935	865	0.21	0.78
2017	22,898	871	0.13	0.67
2018	23,261	1405	0.16	0.73
2019	24,241	880	0.11	0.66

Table 15. Estimates of biological reference points along with estimates of fishing mortality (F), spawning stock biomass (SSB), recent average yield (C), and spawning potential ratio (SPR) of Pacific blue marlin, derived from the assessment ensemble model, where "MSY" indicates reference points based on maximum sustainable yield.

Reference Point	**Estimate**
F_{MSY} (age 1-10)	0.23
F_{2019} (age 1-10)	0.11
$F_{20\%SSB0}$	0.18
SSB_{MSY}	20,677 mt
SSB_{2019}	24,241 mt
$SSB_{20\%SSB0}$	20,729 mt
MSY	24,600 mt
$C_{2017-2019}$	16,512 mt
SPR_{MSY}	17%
SPR_{2019}	34%
$SPR_{20\%SSB0}$	23%

Table 16. Complete list of sensitivity runs conducted for the 2021 stock assessment of Pacific blue marlin.

RUN	NAME	DESCRIPTION
	Alternative Data Inputs	
1	New Drop S2	Remove Japanese longline late CPUE index (S2) from the new growth model.
2	Old Drop S2	Remove Japanese longline late CPUE index (S2) from the old growth model.
3	New Drop S6	Remove Chinese Taipei longline late CPUE index (S6) from the new growth model.
4	Old Drop S6	Remove Chinese Taipei longline late CPUE index (S6) from the old growth model.
5	New DW Size	Down weight all length and size comp to 0.5 in the new growth model.
6	Old DW Size	Down weight all length and size comp to 0.5 in the old growth model
	Alternative Life History Parameters: Natural Mortality	
7	New High M	Alternative natural mortality rates are 10% higher than in the new growth model
8	Old High M	Alternative natural mortality rates are 10% higher than in the old growth model
9	New Low M	Alternative natural mortality rates are 10% lower than in the new growth model
10	Old Low M	Alternative natural mortality rates are 10% lower than in the old growth model
	Alternative Life History Parameters: Growth	
11	New High Linf	Alternative growth curve with 10% higher L_{Amax} for both males and females in the new growth model.
12	Old High Linf	Alternative growth curve with 10% higher L_{Amax} for both males and females in the old growth model.
13	New Low Linf	Alternative growth curve with 10% lower L_{Amax} for both males and females in the new growth model.
14	Old Low Linf	Alternative growth curve with 10% lower L_{Amax} for both males and females in the old growth model.
15	New High K	Alternative growth curve with 10% higher K for both males and females in the new growth model.
16	Old High K	Alternative growth curve with 10% higher K for both males and females in the old growth model.
17	New Low K	Alternative growth curve with 10% lower K for both males and females in the new growth model.
18	Old Low K	Alternative growth curve with 10% lower K for both males and females in the old growth model.
	Alternative Life History Parameters: Maturity Ogive	
19	New Low L_{50}	Alternative lower female length at 50% maturity for the new growth model.
20	Old Low L_{50}	Alternative lower female length at 50% maturity for the old growth model.
21	New High L_{50}	Alternative higher female length at 50% maturity for the new growth model.

22	Old High L_{50}	Alternative higher female length at 50% maturity for the old growth model.
Alternative Life History Parameters: Stock-Recruitment Steepness		
23	New Low Steepness	Alternative steepness set to 0.65 in the new growth model.
24	Old Low Steepness	Alternative steepness set to 0.65 in the old growth model.
25	New Mid Steepness	Alternative steepness set to 0.80 in the new growth model.
26	Old Mid Steepness	Alternative steepness set to 0.80 in the old growth model.
27	New High Steepness	Alternative steepness set to 0.95 in the new growth model.
28	Old High Steepness	Alternative steepness set to 0.95 in the old growth model.
Alternative Model Configuration: Initial Equilibrium Catch		
29	New Fixed EqCatch	Alternative model with Initial Equilibrium Catch set to 2016 levels in the new growth model.
30	Old Fixed EqCatch	Alternative model with Initial Equilibrium Catch set to 2016 levels in the old growth model.

Table 17. Projected median values of Pacific blue marlin spawning stock biomass (SSB, mt) and catch (mt) under four constant fishing mortality rate (F) ensemble model scenarios during 2020-2029.

Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Scenario 1: F = $F_{2003\text{-}2005}$										
SSB	25,459	23,462	21,752	20,498	19,262	18,689	18,252	17,835	17,583	17,475
Catch	33,111	30,527	28,638	27,331	26,431	25,806	25,363	25,044	24,811	24,641
Scenario 2: F = F_{MSY}										
SSB	25,318	23,351	21,583	20,255	19,216	18,405	18,186	17,809	17,513	17,466
Catch	32,875	30,436	28,662	27,439	26,606	26,037	25,645	25,370	25,177	25,039
Scenario 3: F = $F_{2016\text{-}2018}$										
SSB	26,930	28,182	28,764	28,675	28,428	28,731	28,052	28,142	27,861	28,081
Catch	23,321	23,546	23,591	23,561	23,513	23,472	23,443	23,422	23,407	23,397
Scenario 4: F = $F_{30\%}$										
SSB	27,757	30,064	30,624	30,976	31,072	31,624	31,415	31,800	31,753	32,132
Catch	20,828	21,404	21,764	22,001	22,167	22,294	22,393	22,471	22,532	22,580

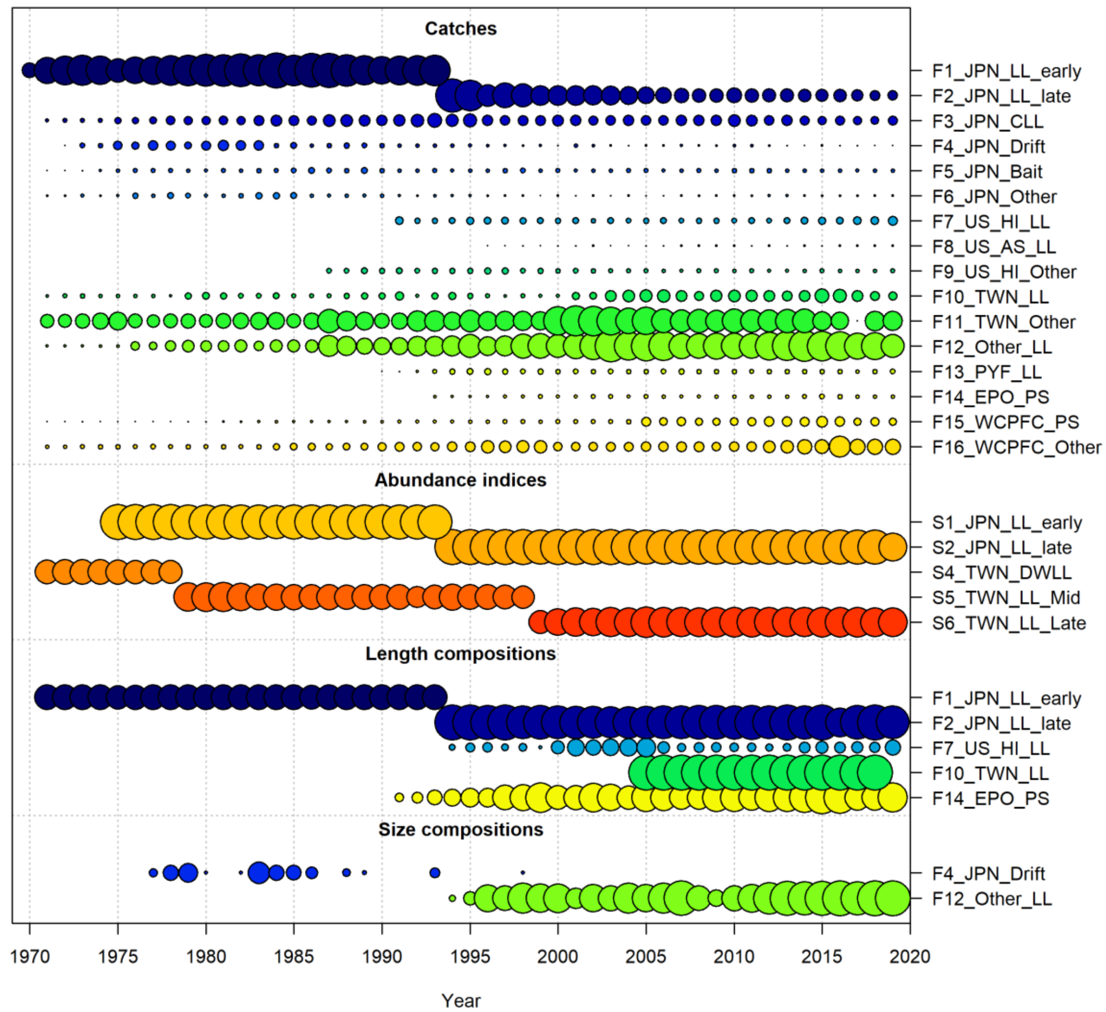


Figure 1. Available temporal coverage and sources of catch, CPUE (abundance indices), and length and size composition for the 2021 stock assessment of the Pacific blue marlin. The size of the circle for catch and length/size composition data indicate the relative amount of catch or number of fish measured, respectively.

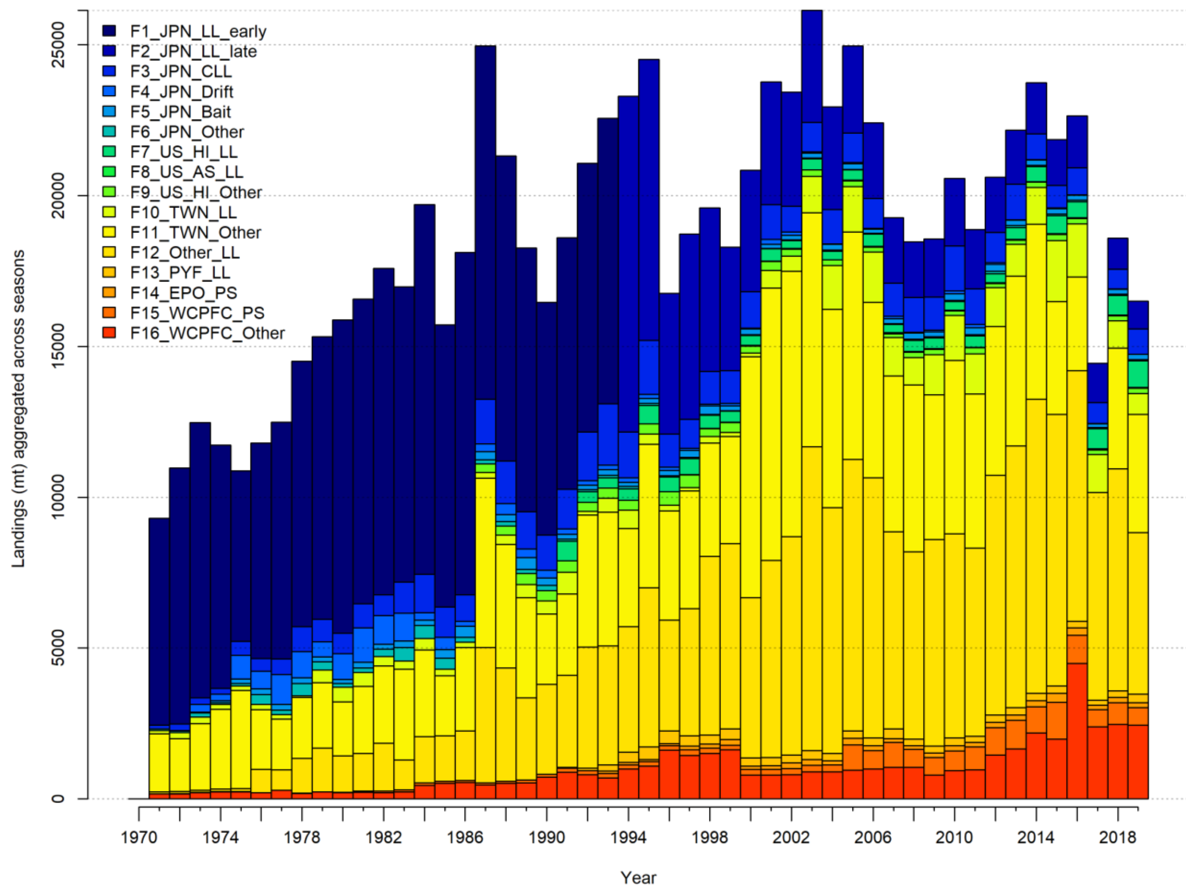


Figure 2. Total annual catch of the Pacific blue marlin by all fisheries harvesting the stock during 1971-2019. See Table 1 for the reference code for each fishery.

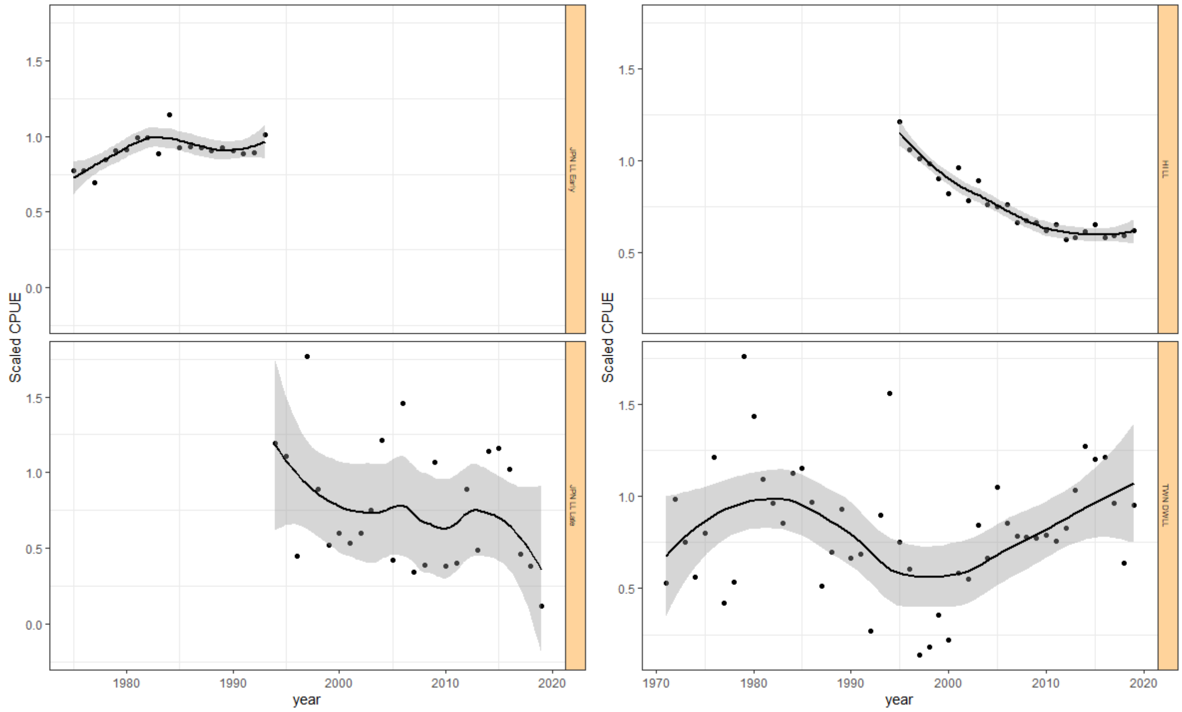
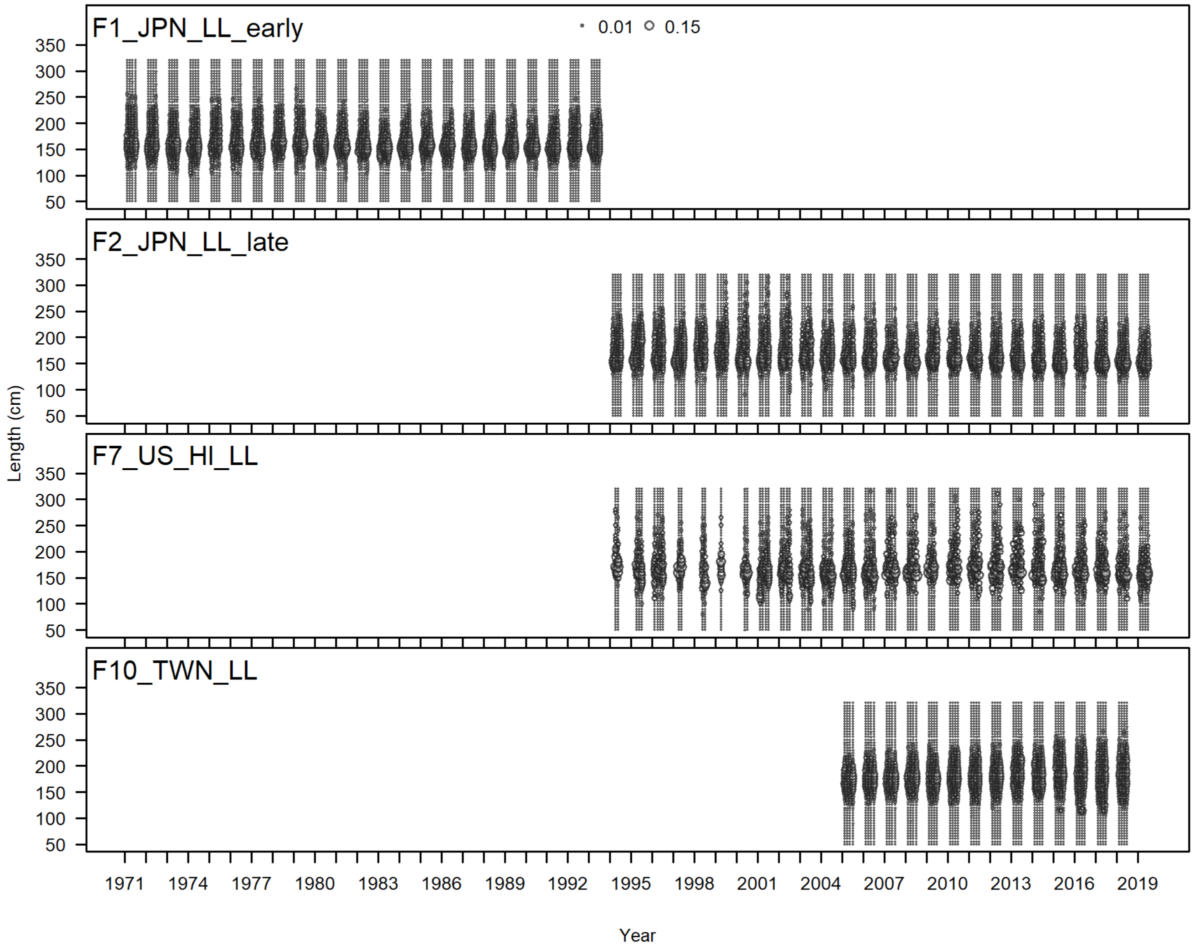


Figure 3. Time series of annual standardized indices of catch-per-unit-effort (CPUE) for each fleet in the assessment models for the Pacific blue marlin as described in Table 3. Index values were rescaled by the mean of each index for comparison purposes. A loess curve was fit to the data to show the general trend with the shaded area representing 95% confidence intervals. In the assessment model, the Chinese Taipei longline index (bottom right) was split into three indices based upon changes in the fishery operations and data reporting.



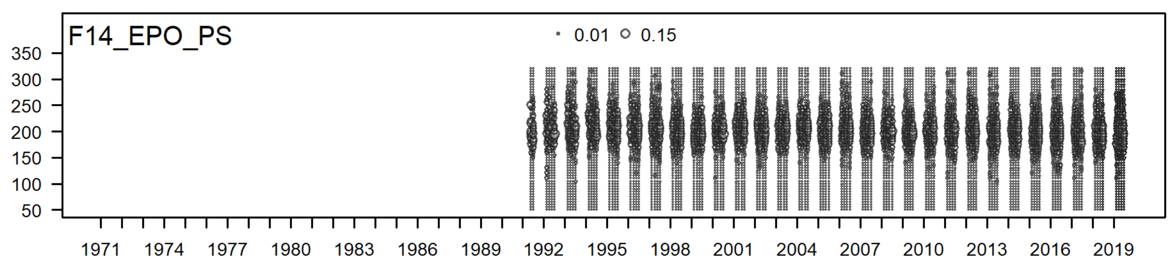
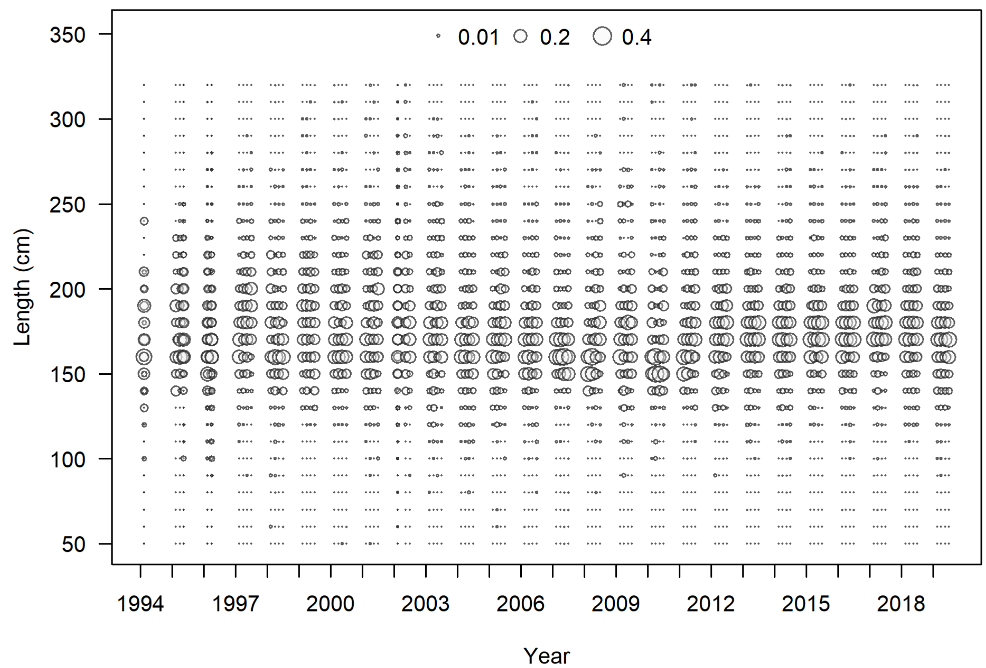


Figure 4. Quarterly length composition data by fishery used in the stock assessment (see Table 3). The sizes of the circles are proportional to the number of observations. All measurements were eye- fork lengths (EFL, cm).



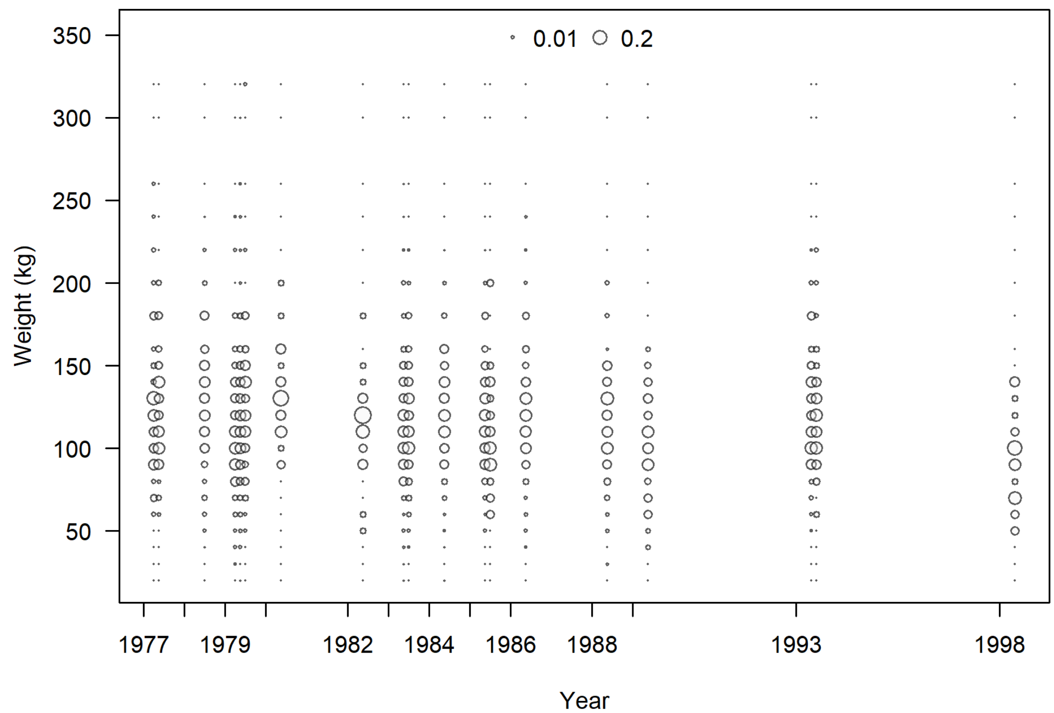


Figure 5. Quarterly generalized size composition data by fishery used in the stock assessment (see Table 3). The sizes of the circles are proportional to the number of observations. Data from F12 Other longline fleets are in eye-fork length (cm EFL, top). Data from the F4 Japanese driftnet fleet are in weight (kg, bottom).

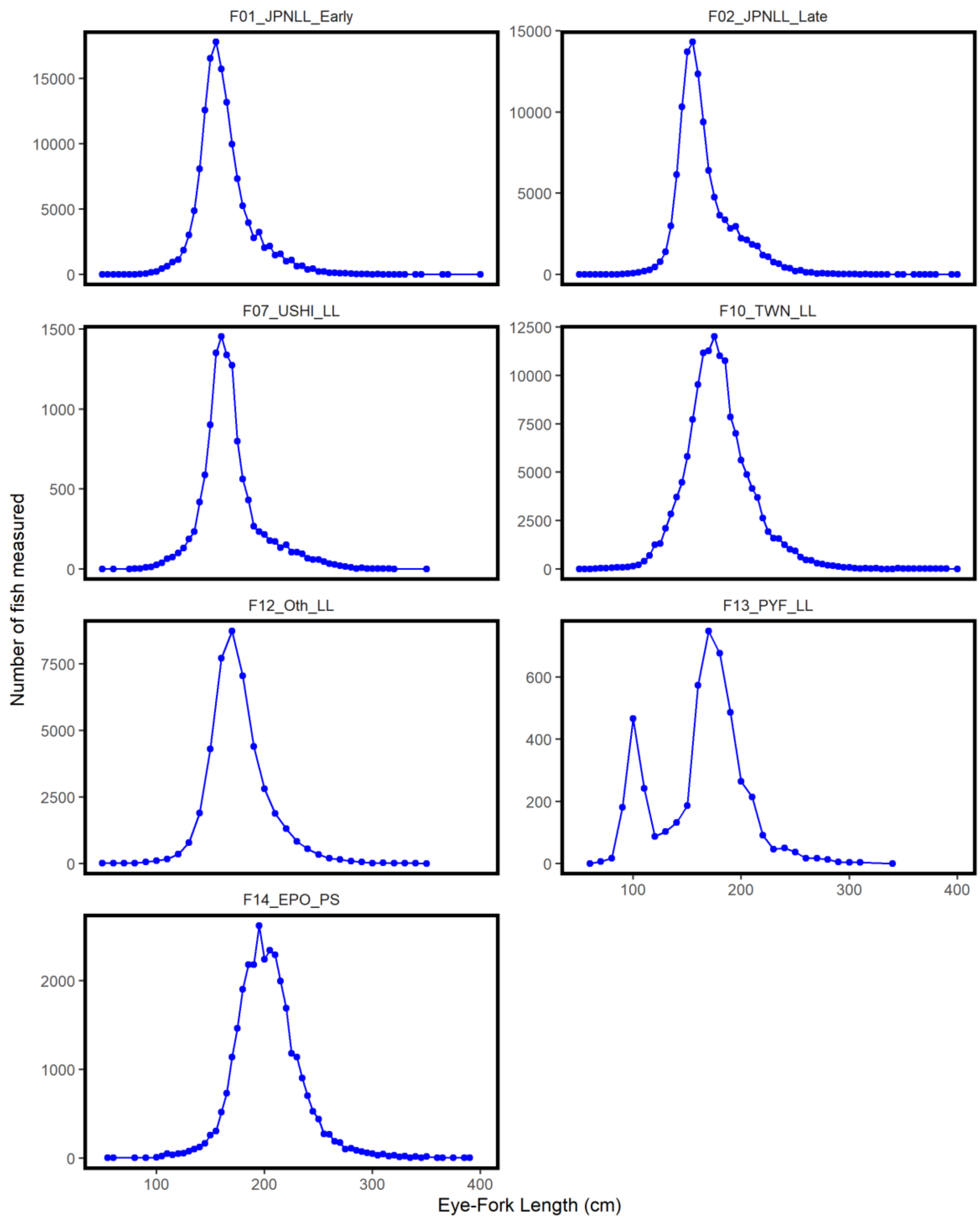


Figure 6. Aggregated length compositions available for use in the stock assessment (see Table 3 for descriptions of the composition data). Fleet 13, French Polynesia was not included in the final models. All measurements were eye-fork lengths (EFL, cm).

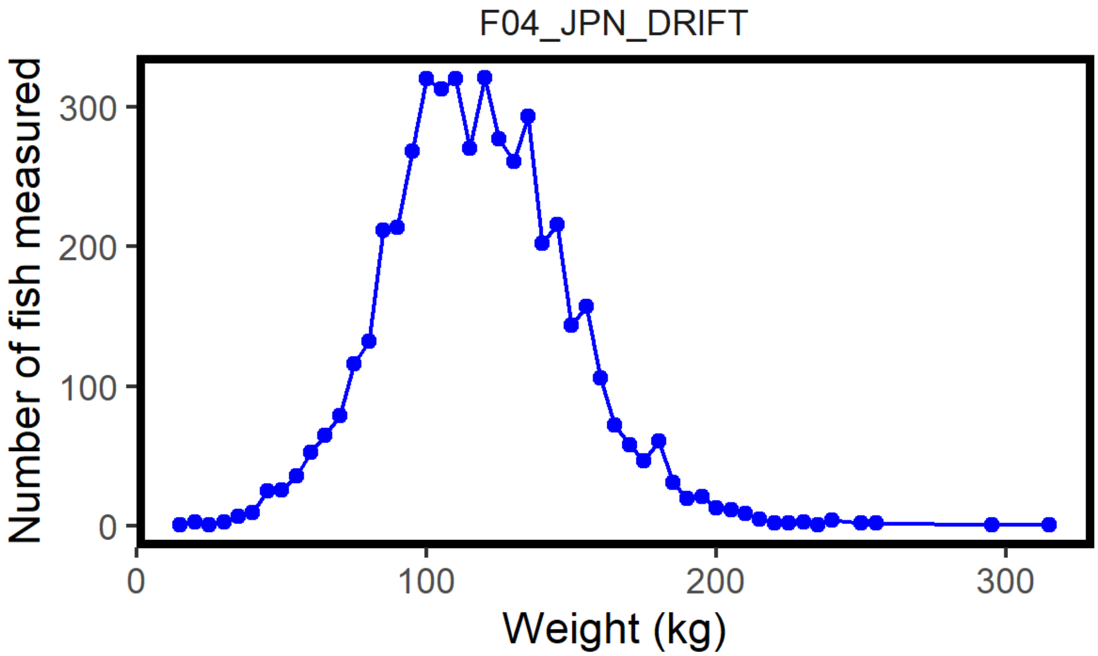


Figure 7. Aggregated weight compositions available for use in the stock assessment (see Table 3 for descriptions of the composition data). All measurements were weight (kg).

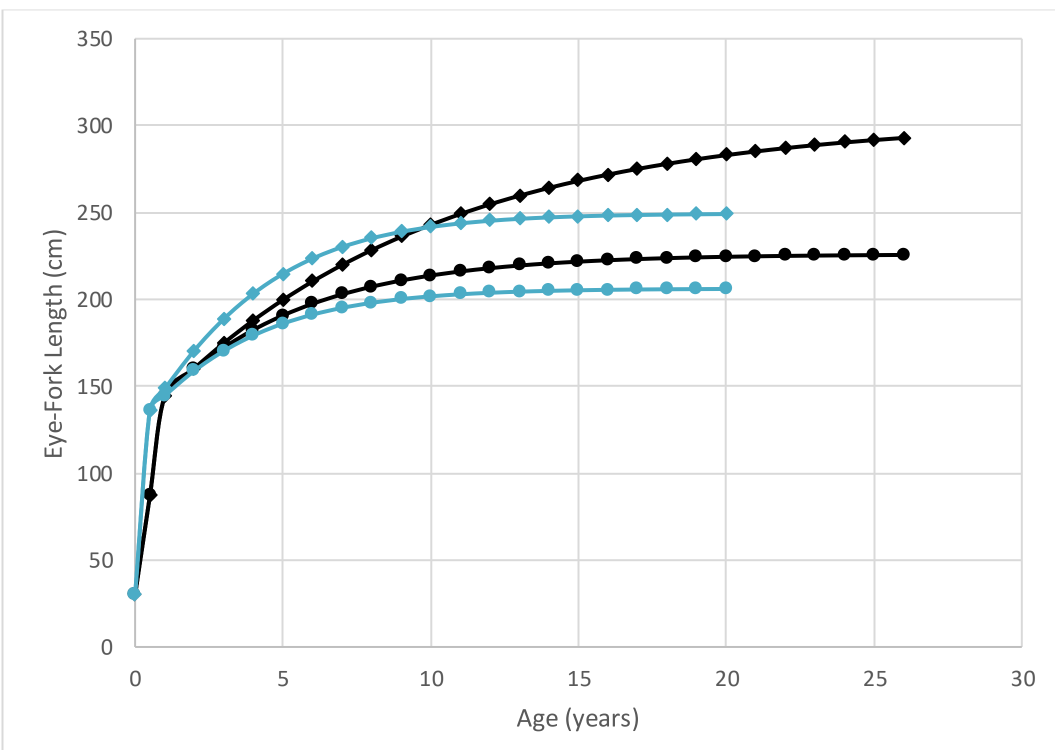


Figure 8. Length at age for Pacific blue marlin for the old growth model (von-bertalanffy curve, black lines) and the new growth model (Richards curve, blue lines) by sex (females diamonds, males circles) used in the 2021 assessment.

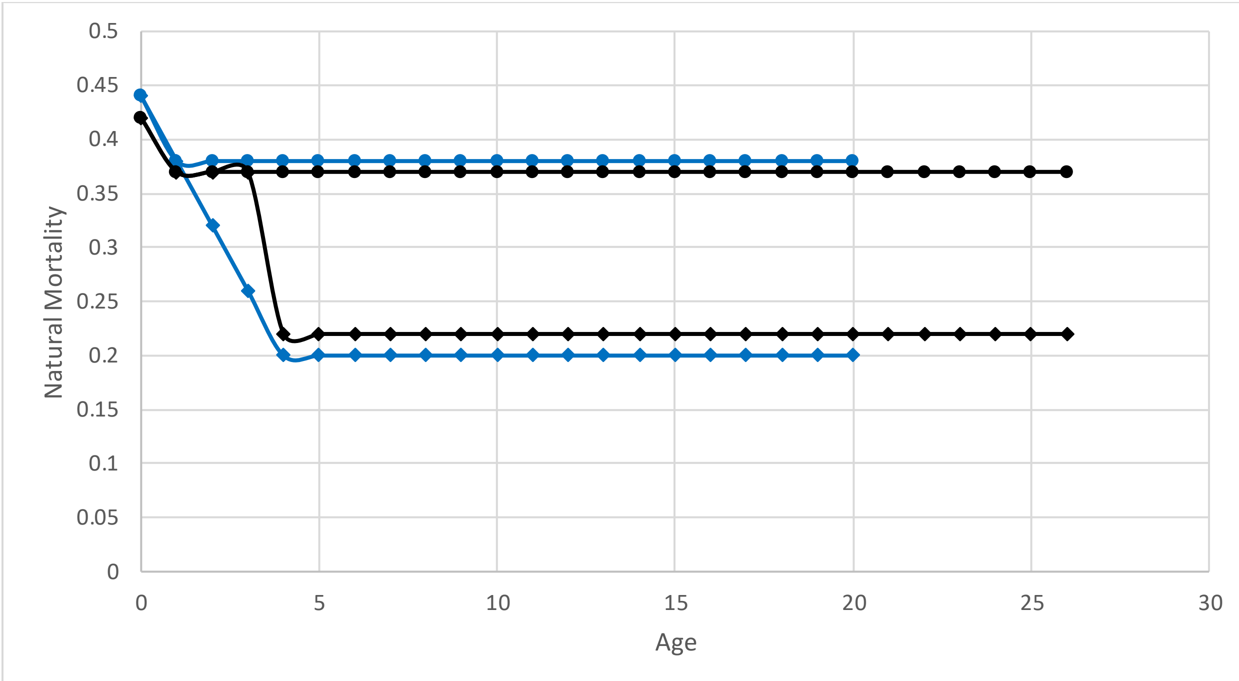


Figure 9. Natural mortality at age for Pacific blue marlin for the old growth model (black lines) and the new growth model (blue lines) by sex (females diamonds, males circles) used in the 2021 assessment.

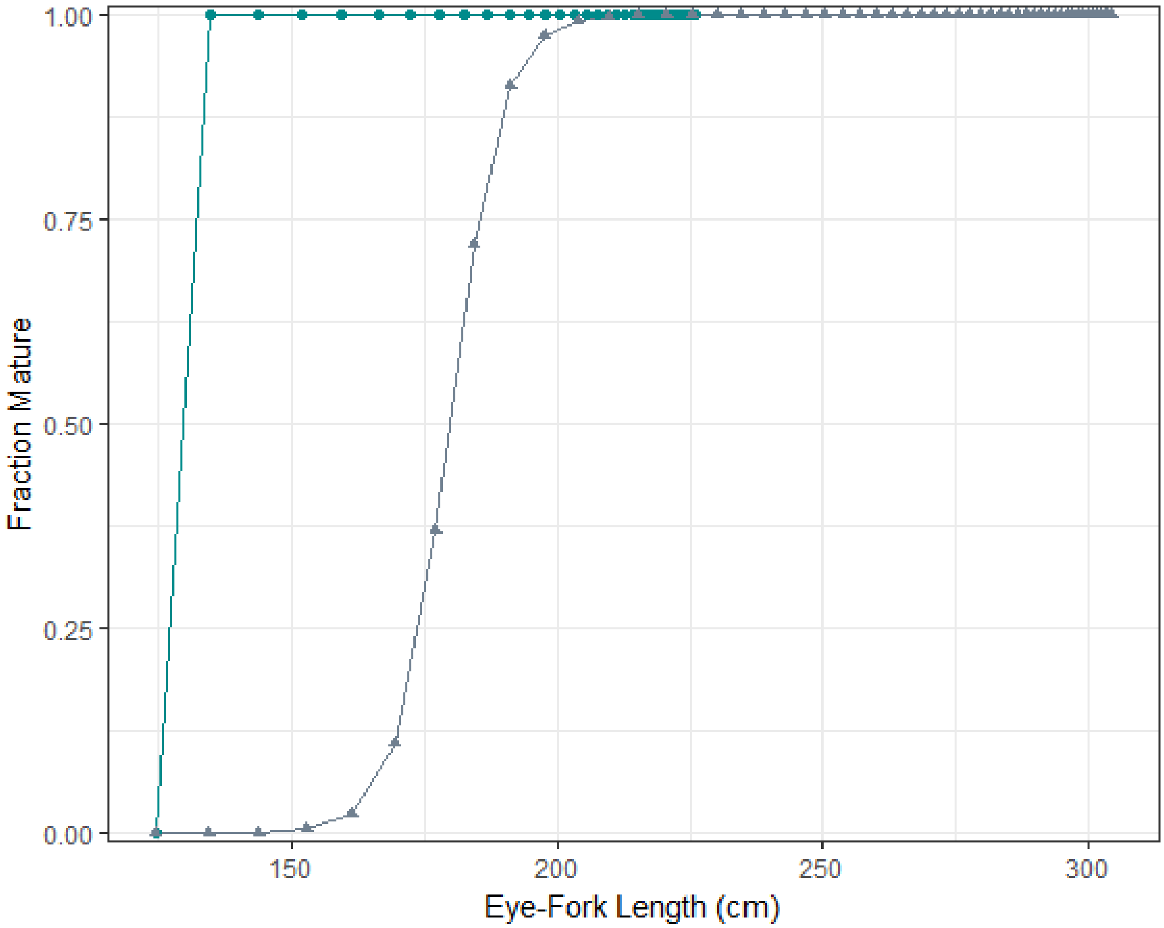


Figure 10. Probability of maturity at length for male (blue) and female (grey) Pacific blue marlin used in the 2021 assessment. Only the female maturity ogive is conisdered in Stock Synthesis models.

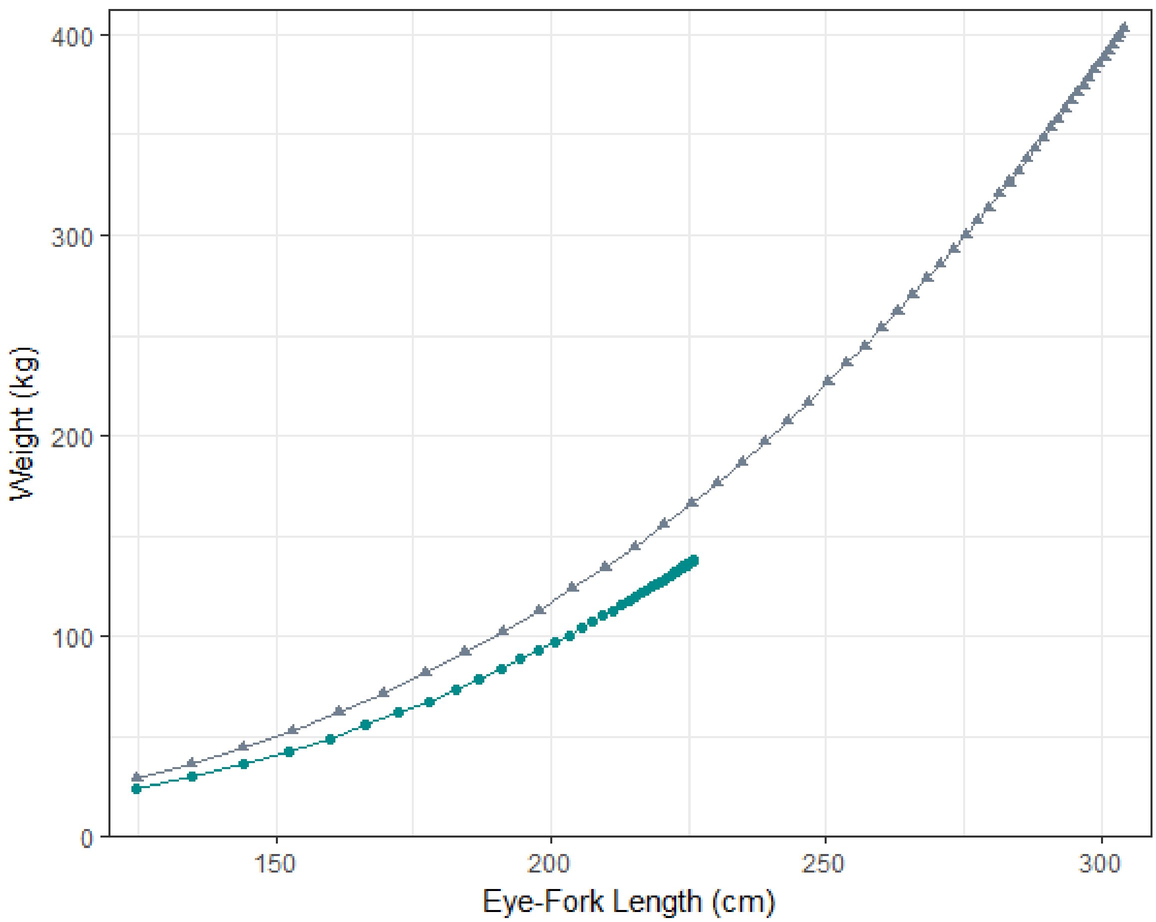


Figure 11. Weight at length curves for male (blue) and female (grey) Pacific blue marlin used in the 2021 assessment.

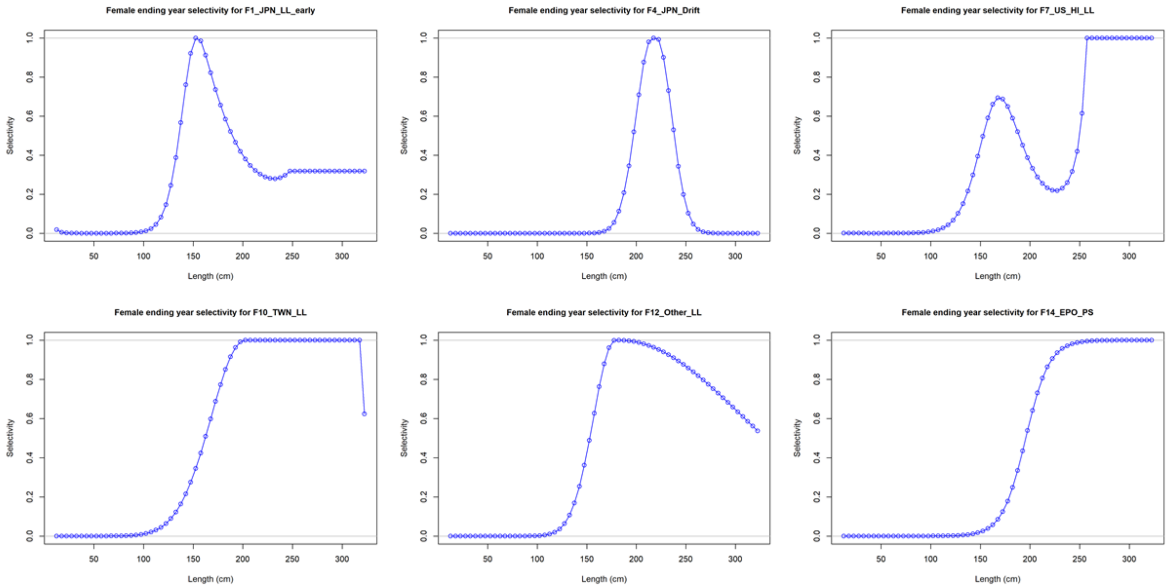


Figure 12. Estimated length-based selectivity of fisheries in the final year of the model (2019) for the 2021 Pacific blue marlin assessment.

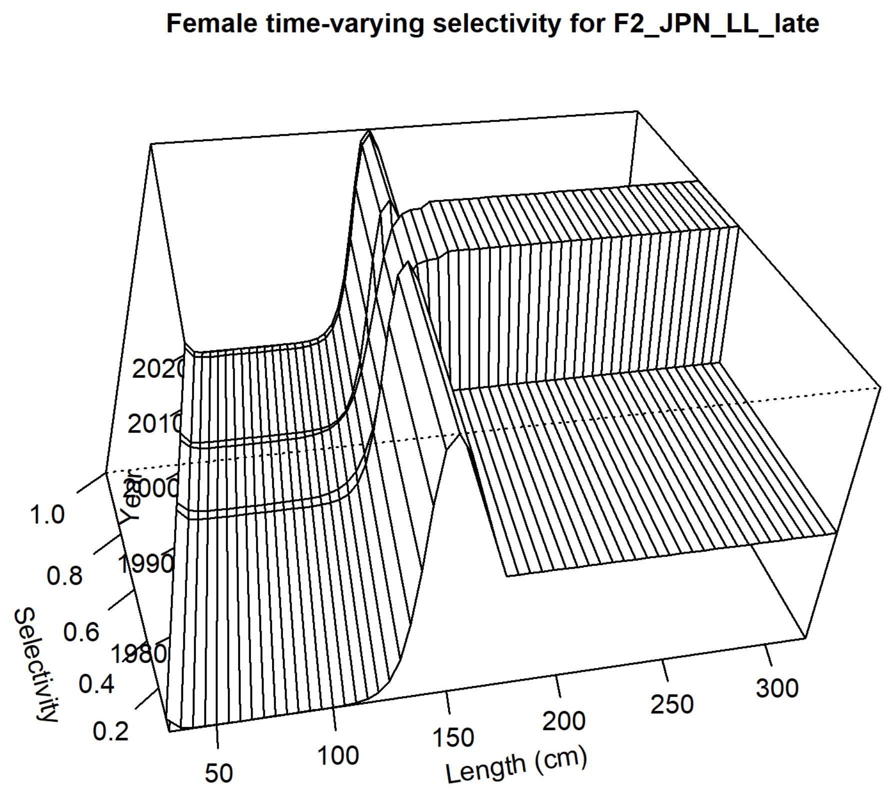


Figure 13. Estimated annual length-based selectivity of F2, Japanese longline late length composition data for the 2021 Pacific blue marlin assessment.

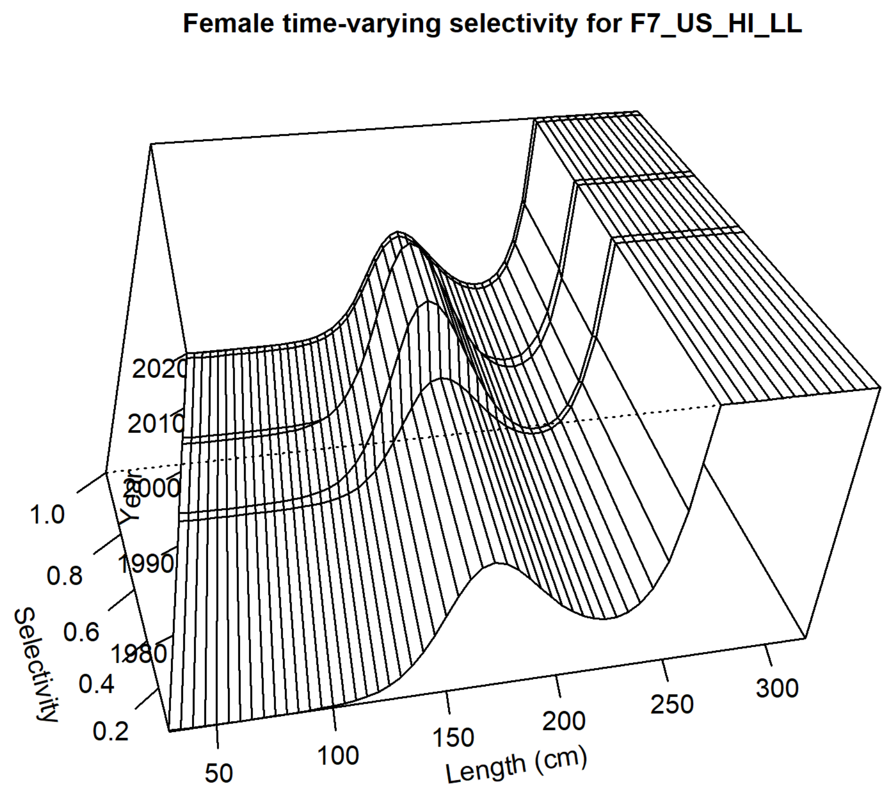


Figure 14. Estimated annual length-based selectivity of F7, US Hawaii longline length composition data for the 2021 Pacific blue marlin assessment.

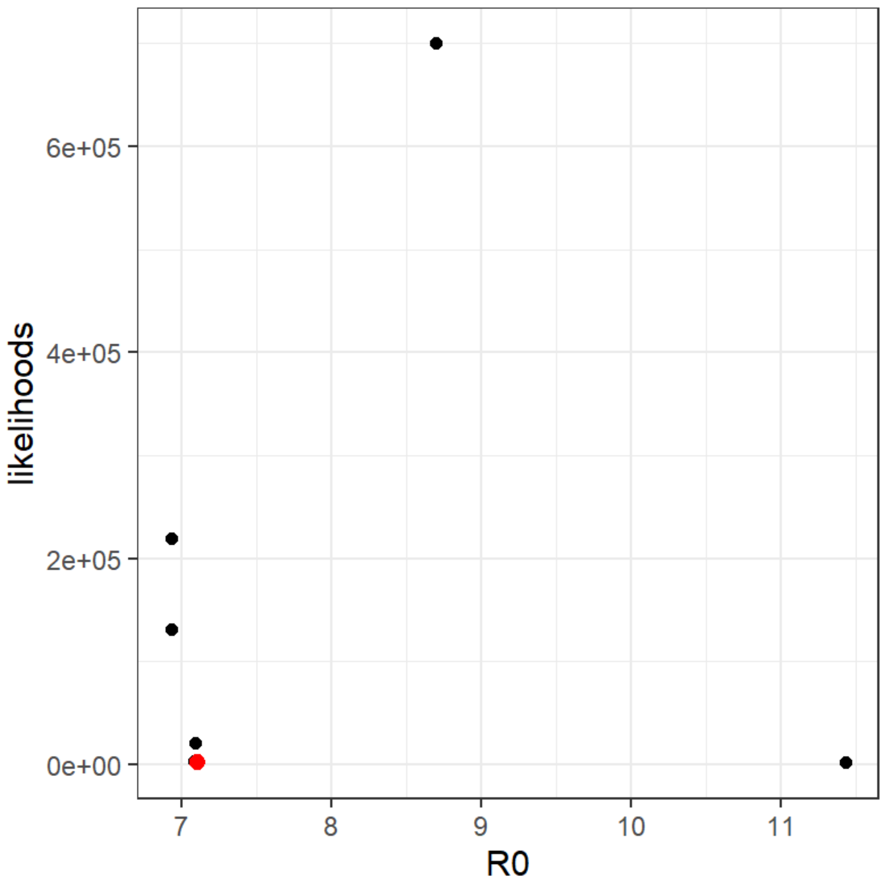


Figure 15. Results of a randomized initial parameter value diagnostic for the old growth model where 30 randomized initial conditions were used with a CV of 10% assigned to each parameter. Results are shown for the old growth model (MLE, solid red circle) and the old growth model with randomized initial parameter values (Jitter runs, solid black circle).

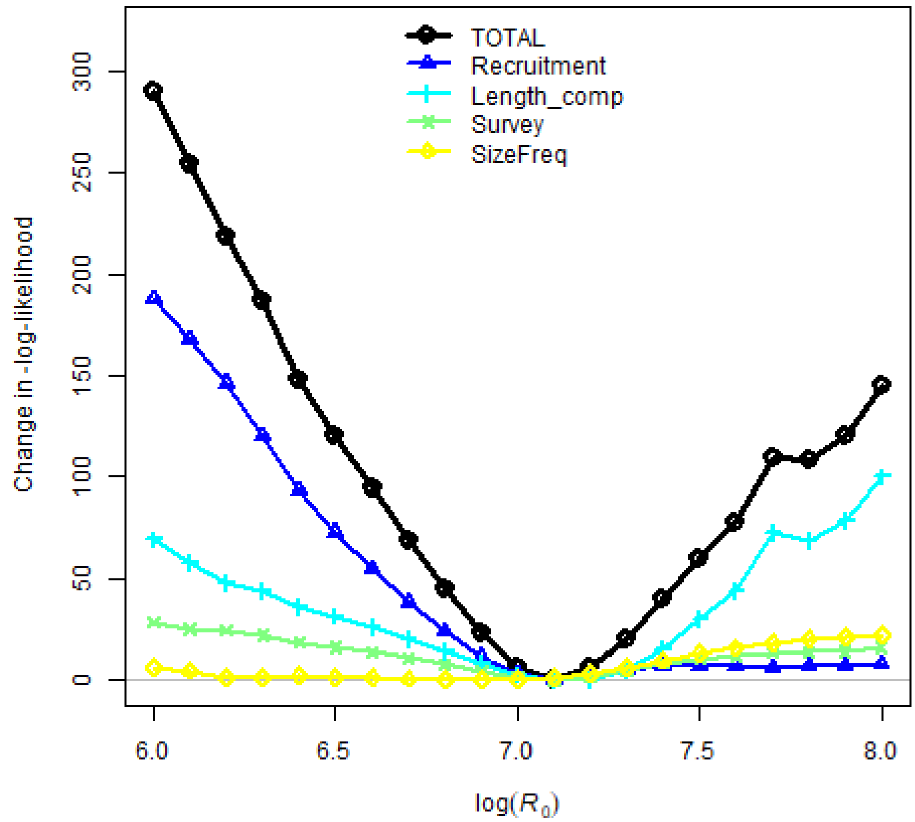


Figure 16. Profiles of the negative log-likelihoods relative to the minimum value of each component for the different likelihood components affecting the unfished recruitment parameter R_0 in log-scale (i.e., the x-axis is log(R_0)) ranging from 6.0 to 8.0 for the old growth model, where recruitment represents the likelihood component based on the deviations from the stock-recruitment curve, sizefreq represents the joint likelihood component for combined fleets based on the fish generalized size composition data, and length data represents the joint likelihood component for combined fleets based on the fish length composition data.

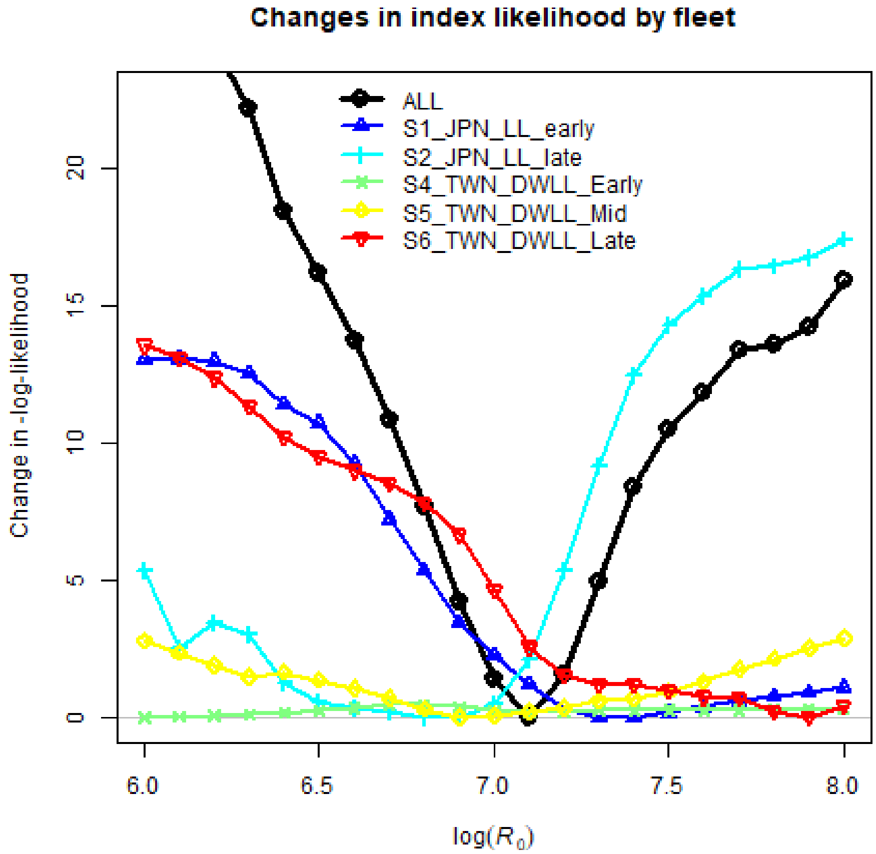


Figure 17. Profiles of the relative negative log-likelihoods by fleet-specific index likelihood components for the virgin recruitment in log-scale (log(R_0)) ranged from 6.0 to 8.0 of the old growth model. See Table 3 for descriptions of the index data. S3 was not included in the total likelihood.

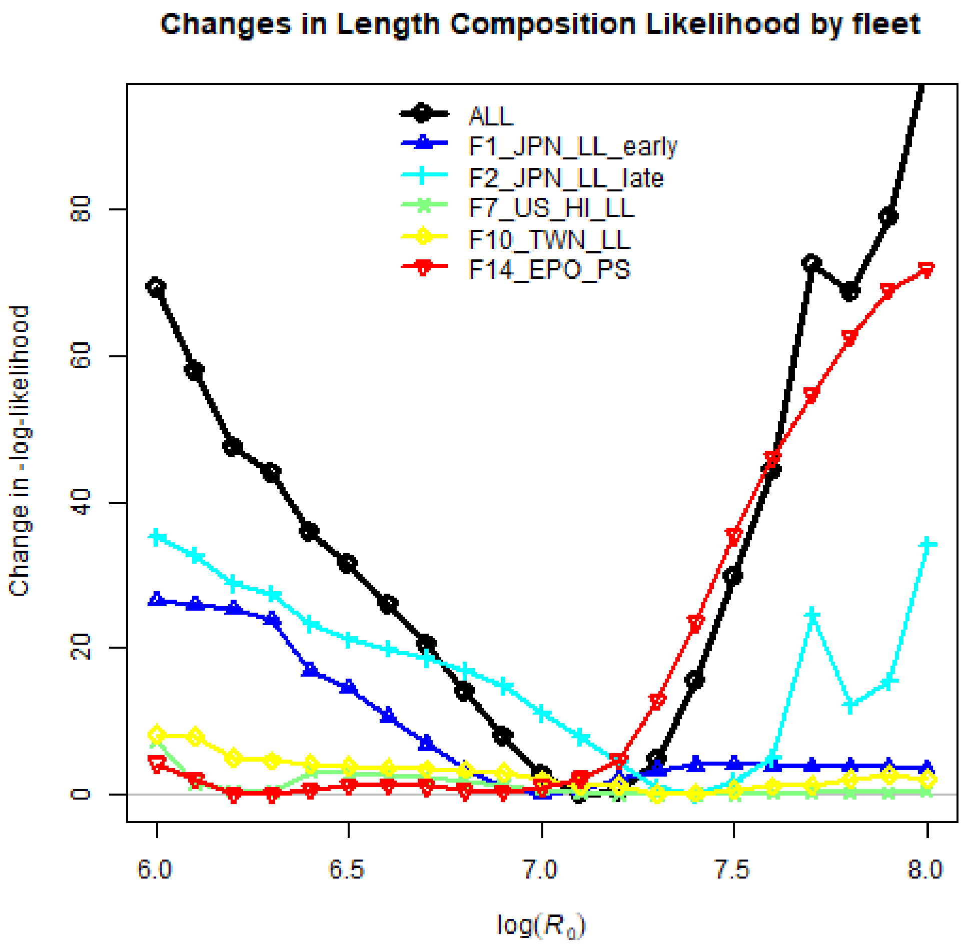


Figure 18. Profiles of the relative negative log-likelihoods by fleet-specific length composition likelihood components for the virgin recruitment in log-scale (log(R_0)) ranged from 6.0 to 8.0 of the old growth model. See Table 3 for descriptions of the length composition data.

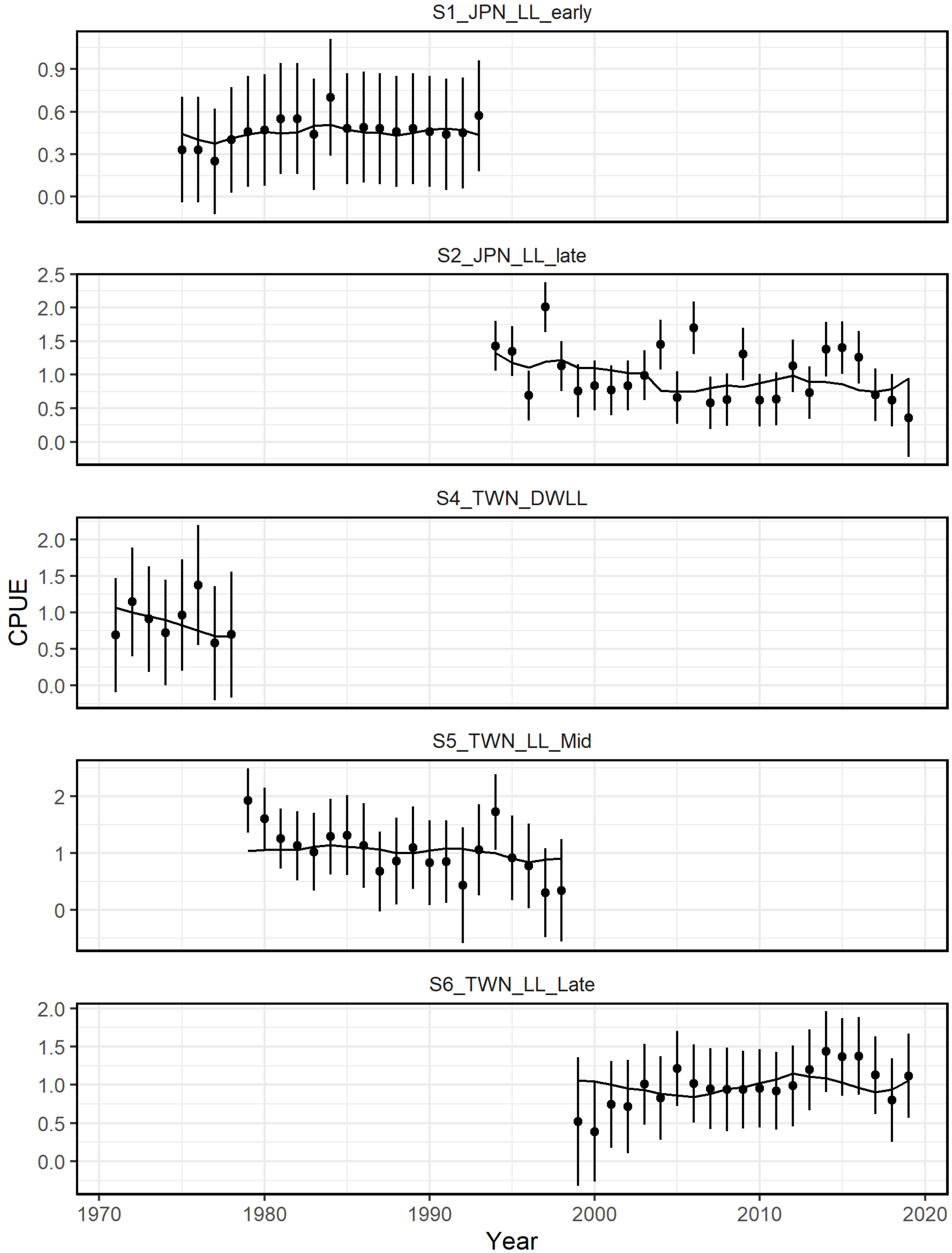


Figure 19. Fits to the standardized catch-per-unit-effort (CPUE) data sets from different fisheries for the old growth model. The line is the model predicted value and the points are observed (data) values. The vertical lines represent the estimated confidence intervals (± 1.96 standard deviations) around the CPUE values. S3 was not included in the total likelihood.

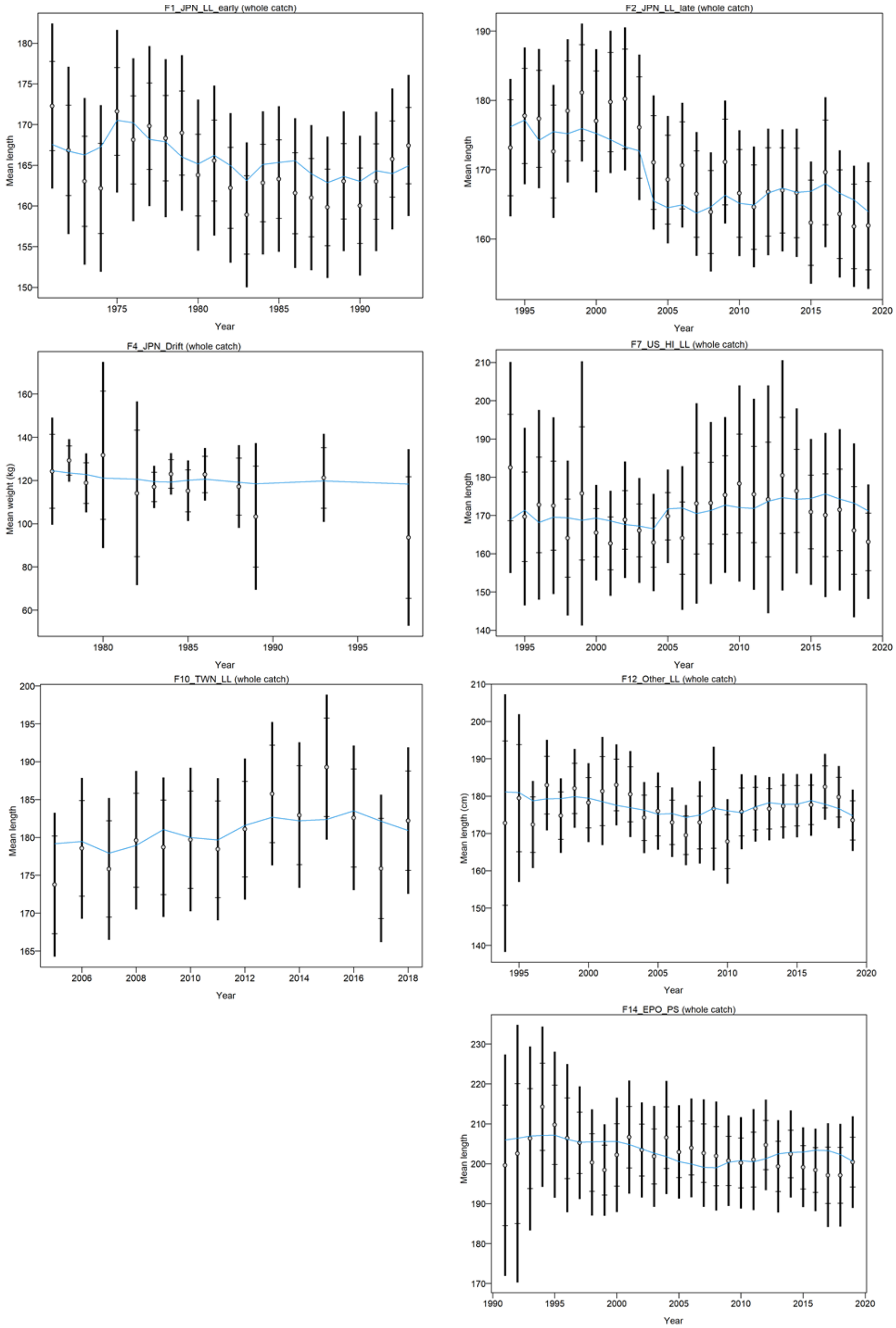


Figure 20. Old growth model fit (lines) to mean length of the composition data (points, showing the observed mean age and 95% credible limits around mean age (vertical lines)). See Table 3 for descriptions of the data. All measurements were eye-fork lengths (EFL, cm) except for F4 Japanese Driftnet, which was in weight (kg).

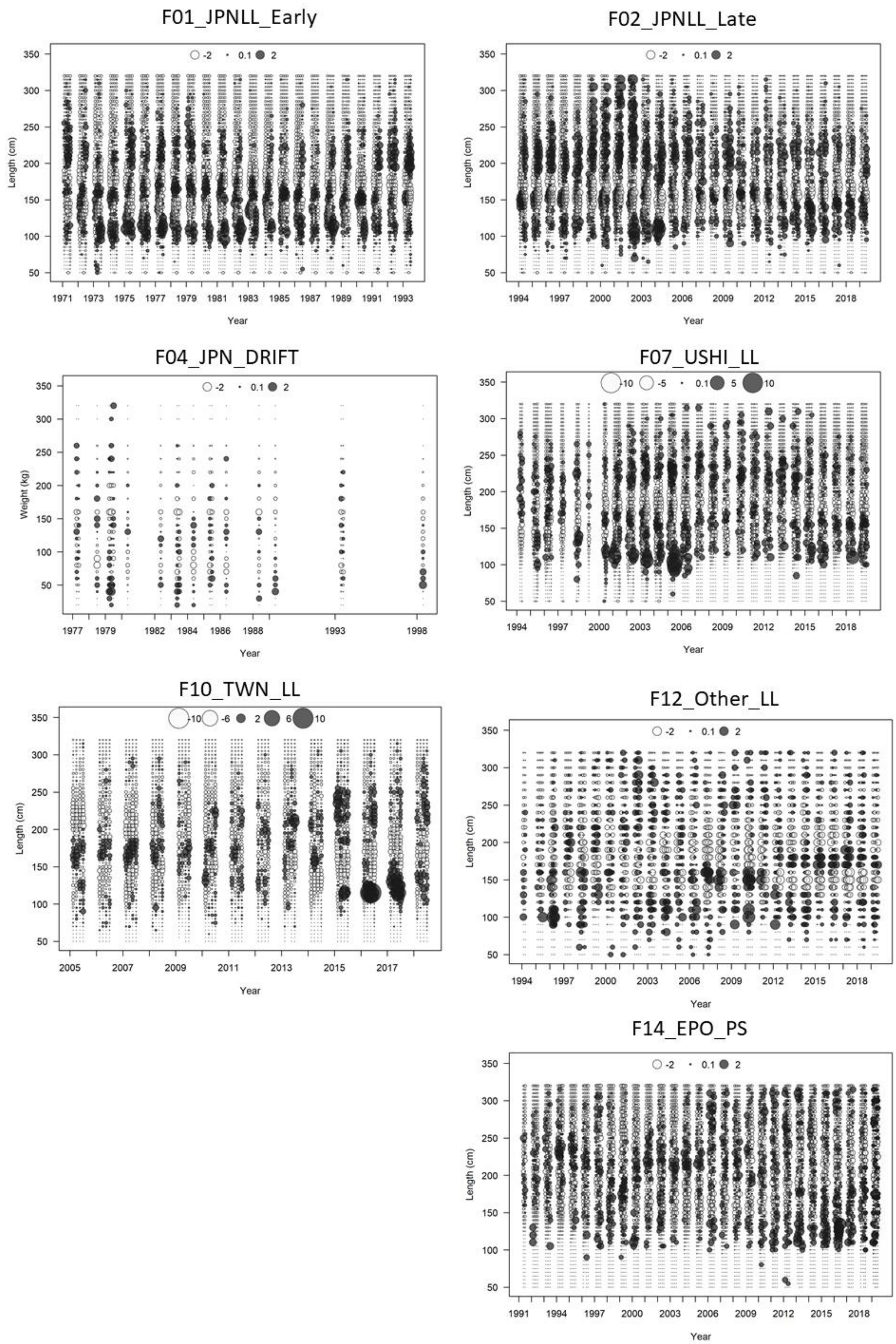


Figure 21. Pearson residual plots of model fits to the various length and weight composition data for the Pacific blue marlin fisheries used in the old growth model.

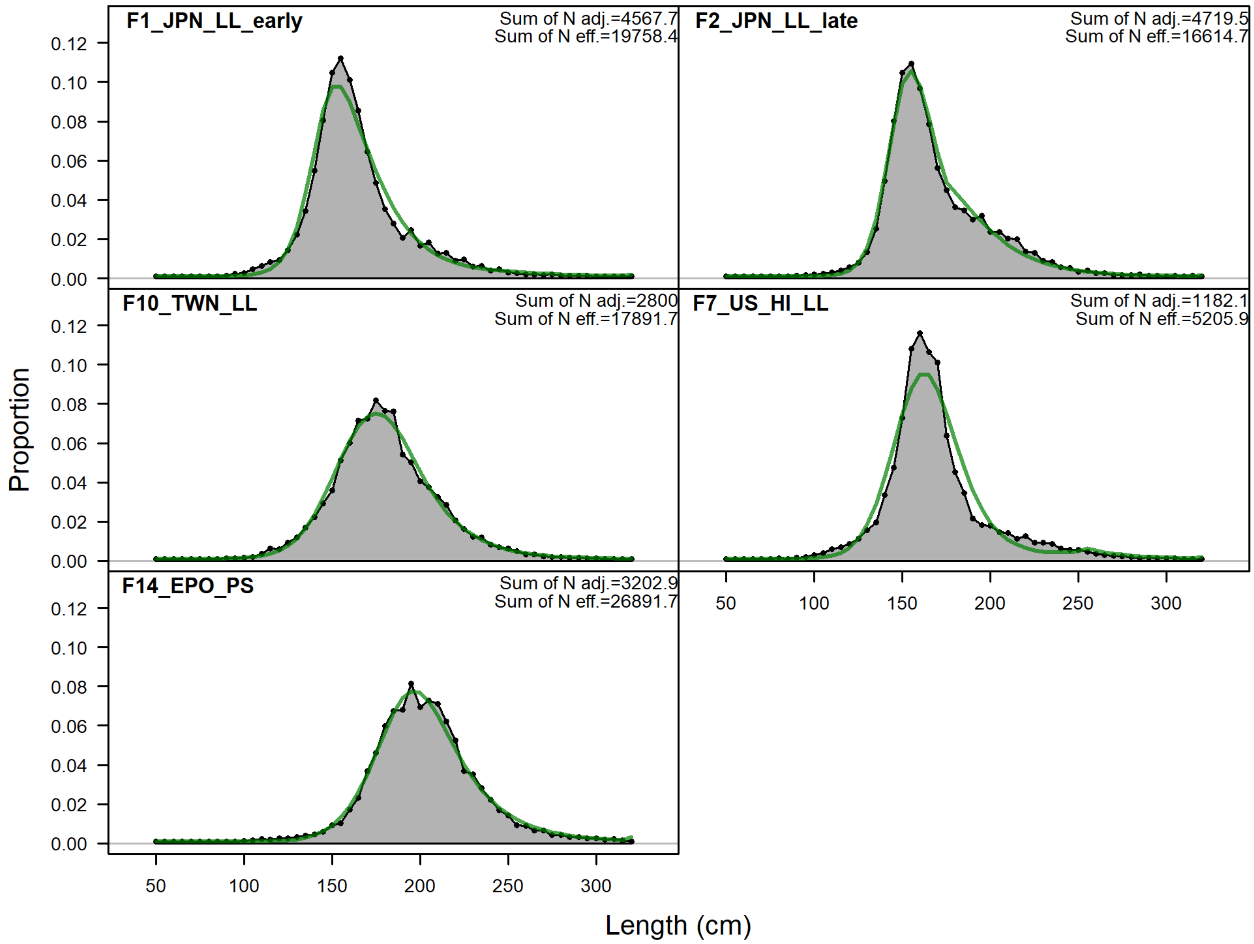


Figure 22. Comparison of observed (gray shaded area and blue dots) and model-predicted (green solid line) length compositions for fisheries used in the Pacific blue marlin old growth model. All measurements were eye-to-fork lengths (EFL, cm).

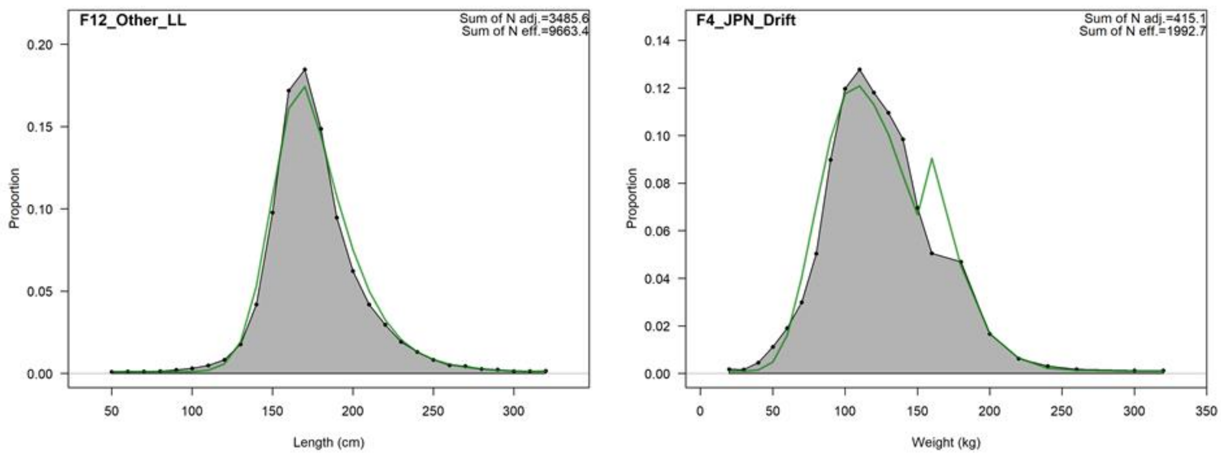


Figure 23. Comparison of observed (gray shaded area and blue dots) and model-predicted (green solid line) generalized size compositions for fisheries used in the Pacific blue marlin old growth model. F12 Other longline is measured in eye-to-fork lengths (EFL, cm), F4 Japanese Driftnet is measured in weight (kg).

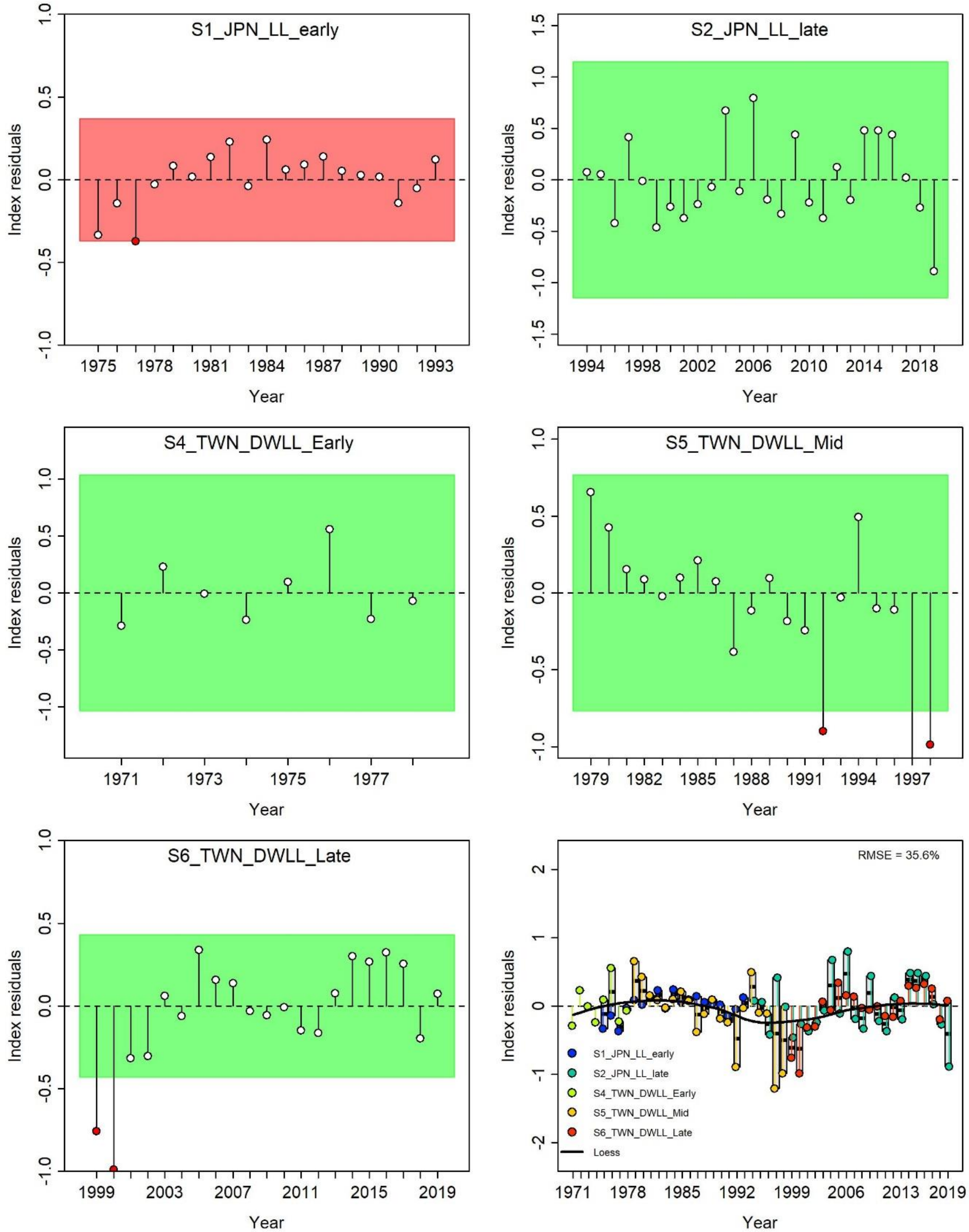


Figure 24. Runs test results for the CPUE fits included in the old growth model. Green shading indicates no evidence (p≥0.05) and red shading indicates evidence (p<0.05) to reject the hypothesis of a randomly distributed time-series of residuals. The red/green shaded areas span three residual standard deviations to either side from zero, the red points outside of the shading violate the "three-sigma limits" for that series.

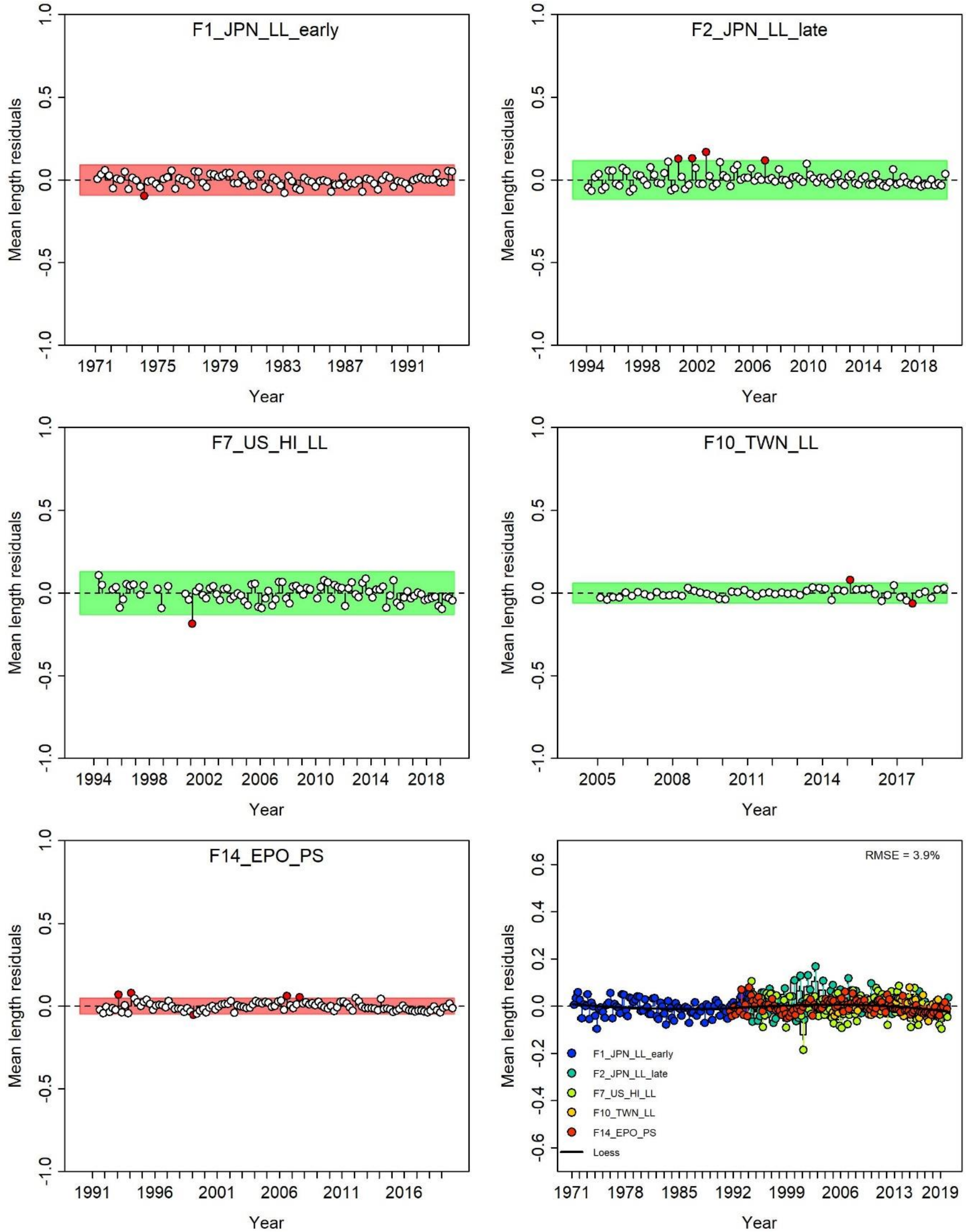


Figure 25. Runs test results for the mean lengths of size composition data included in the old growth model. Green shading indicates no evidence ($p \geq 0.05$) and red shading indicates evidence ($p < 0.05$) to reject the hypothesis of a randomly distributed time-series of residuals. The red/green shaded areas span three residual standard deviations to either side from zero, the red points outside of the shading violate the "three-sigma limits" for that series.

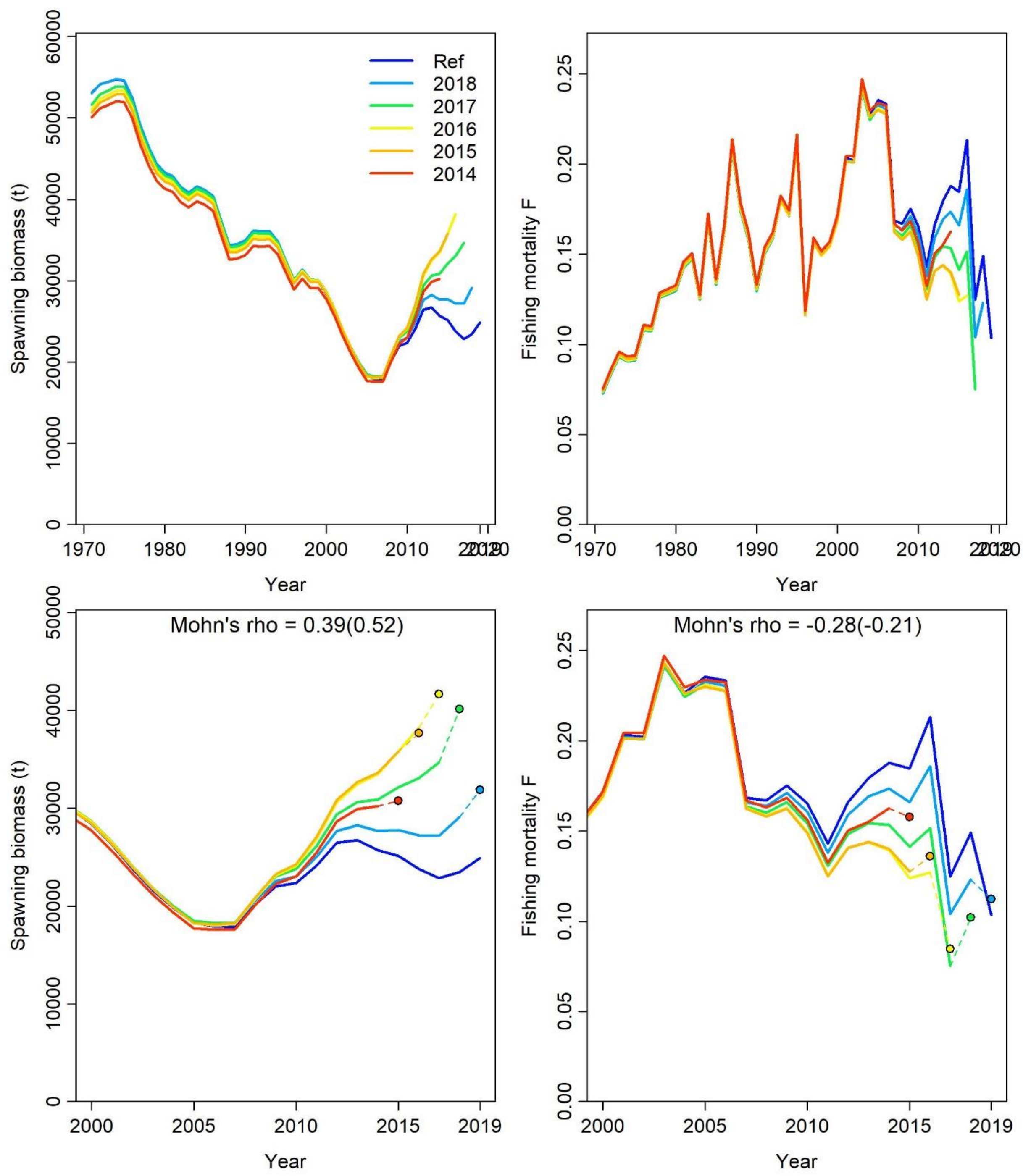


Figure 26. Retrospective analysis of spawning biomass (left) and fishing mortality (right) consisting of 5 reruns of the base case model each fitted with one more year of data removed from the old growth model. The top panels are the entire time series (1971-2019), the bottom panels are the time series since 2000 for visibility.

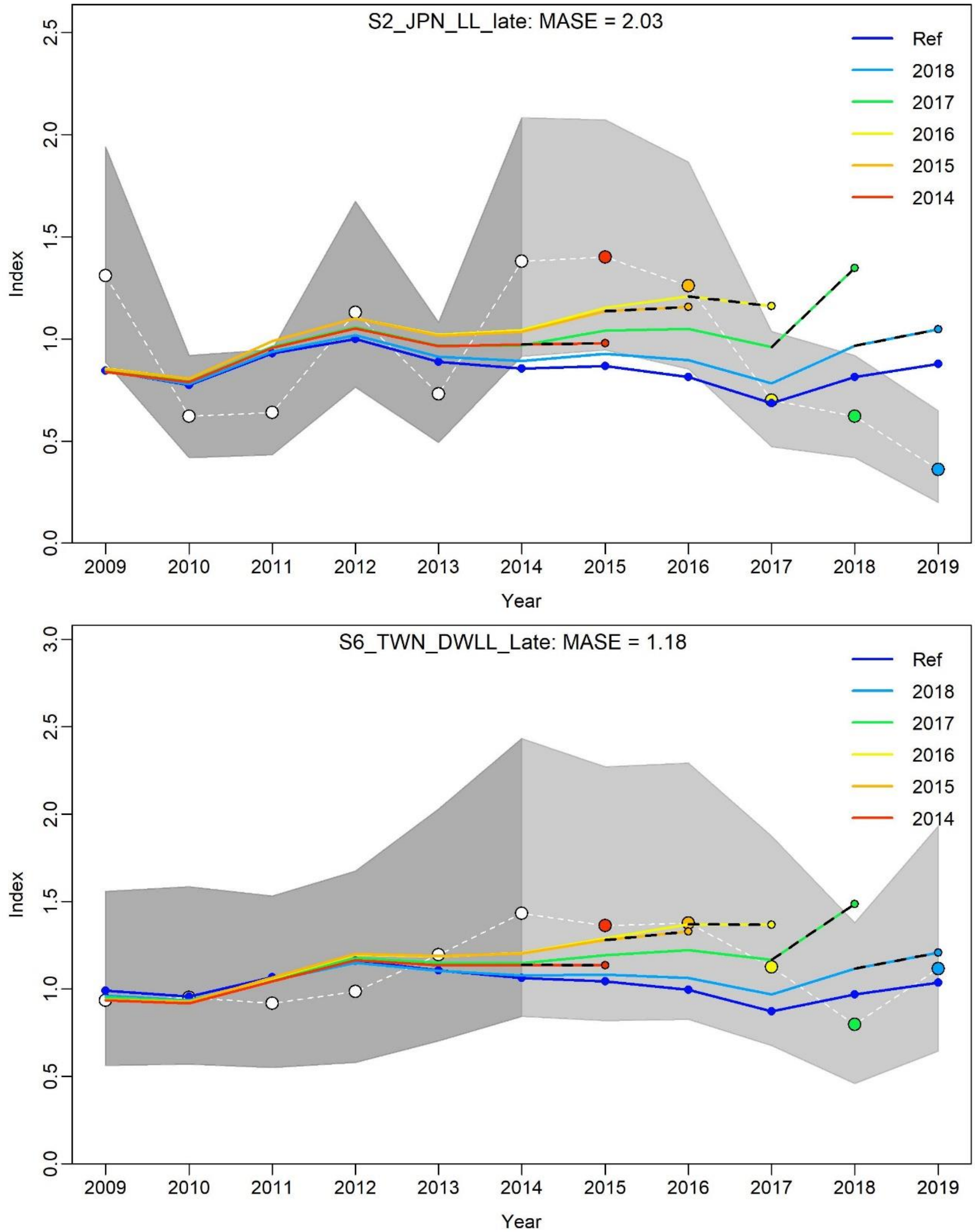


Figure 27. Hind casting cross-validation (HCxval) results for Japanese longline late (top) and Chinese Taipei longline late (bottom) CPUE fits, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the old growth model. The 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.

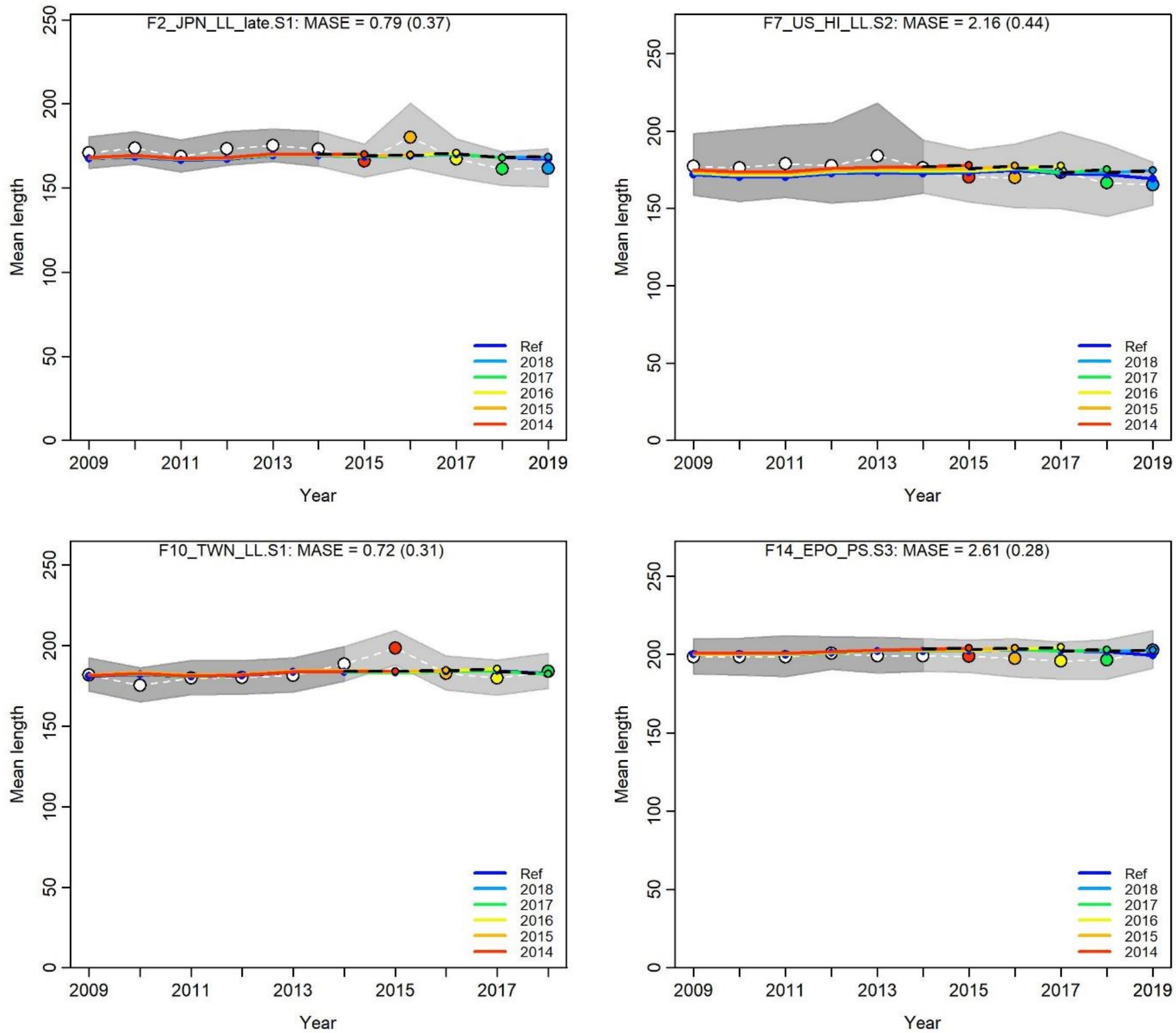


Figure 28. Hind casting cross-validation (HCxval) results for Japanese longline late (top left), US Hawaii longline (top right), Chinese Taipei longline (bottom left), and EPO purse seine (bottom right) size composition mean lengths, showing observed (large points with dashed line), fitted (solid lines), 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.

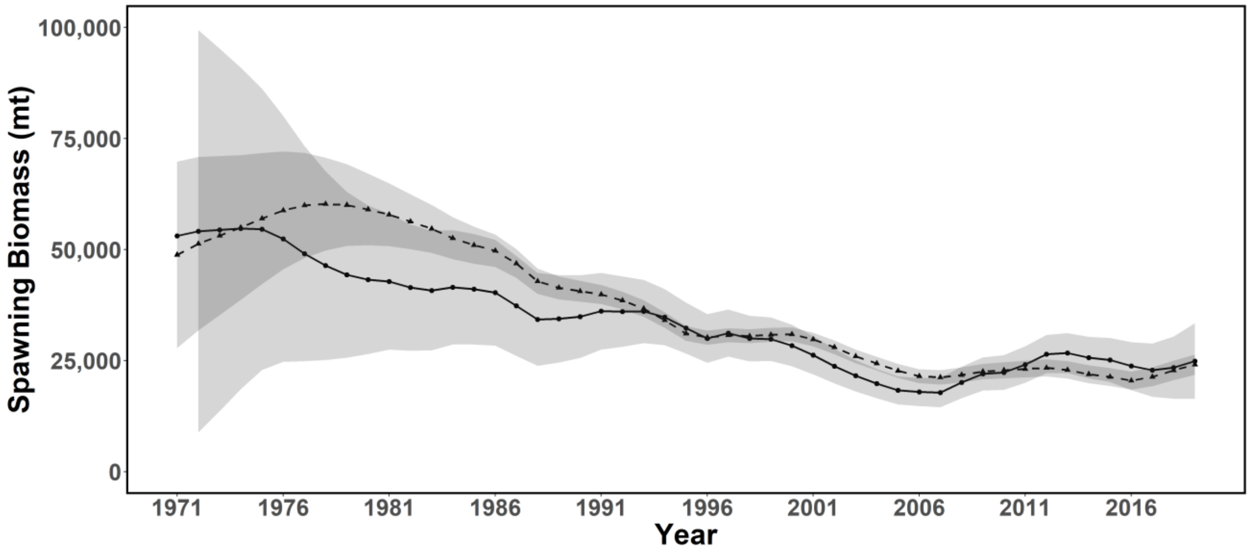


Figure 29. Age structured production model diagnostic for the old growth model. Spawning stock biomass estimates from the full model (black circles, solid line; grey shading indicates 95% confidence interval) and ASPM model diagnostic (black triangles, dashed line; grey shading indicates 95% confidence interval).

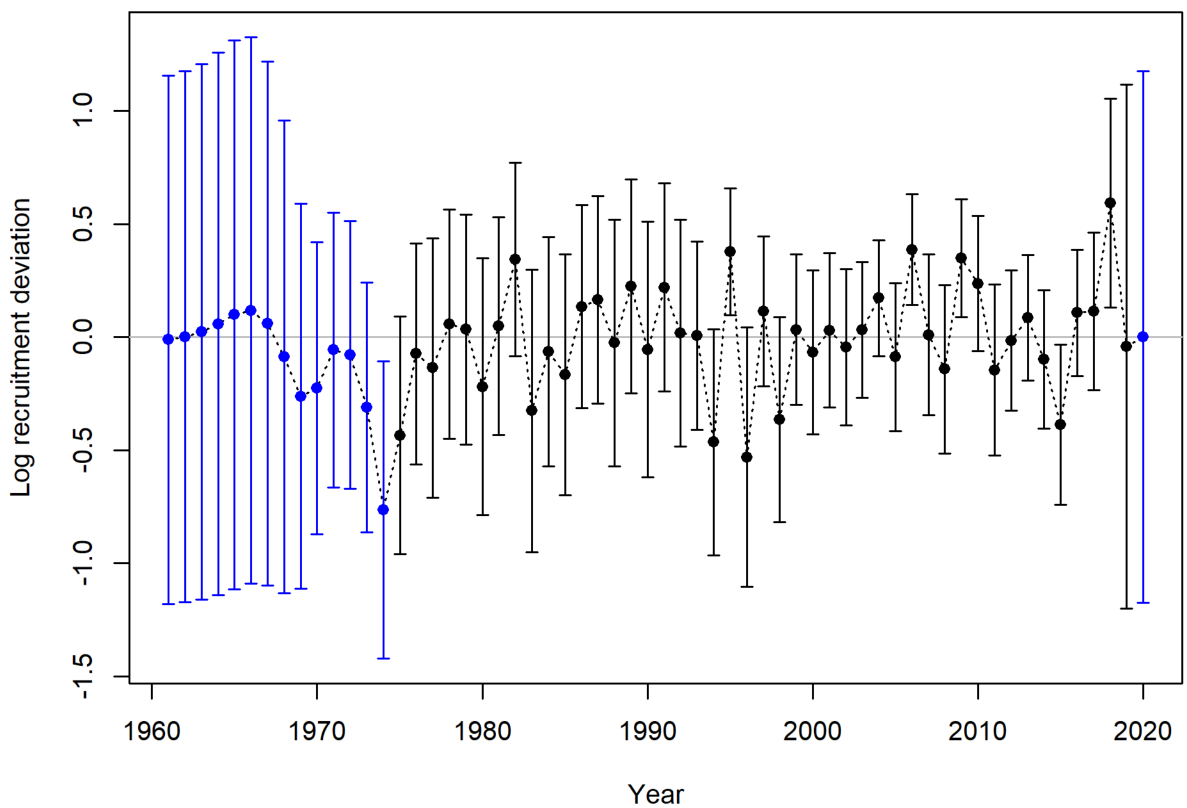


Figure 30. Recruitment deviations (log scale) for the stock synthesis old growth model with sigmaR = 0.6. Blue indicates early recruitment deviations from 1960-1971, and forecast recruitment in 2020.

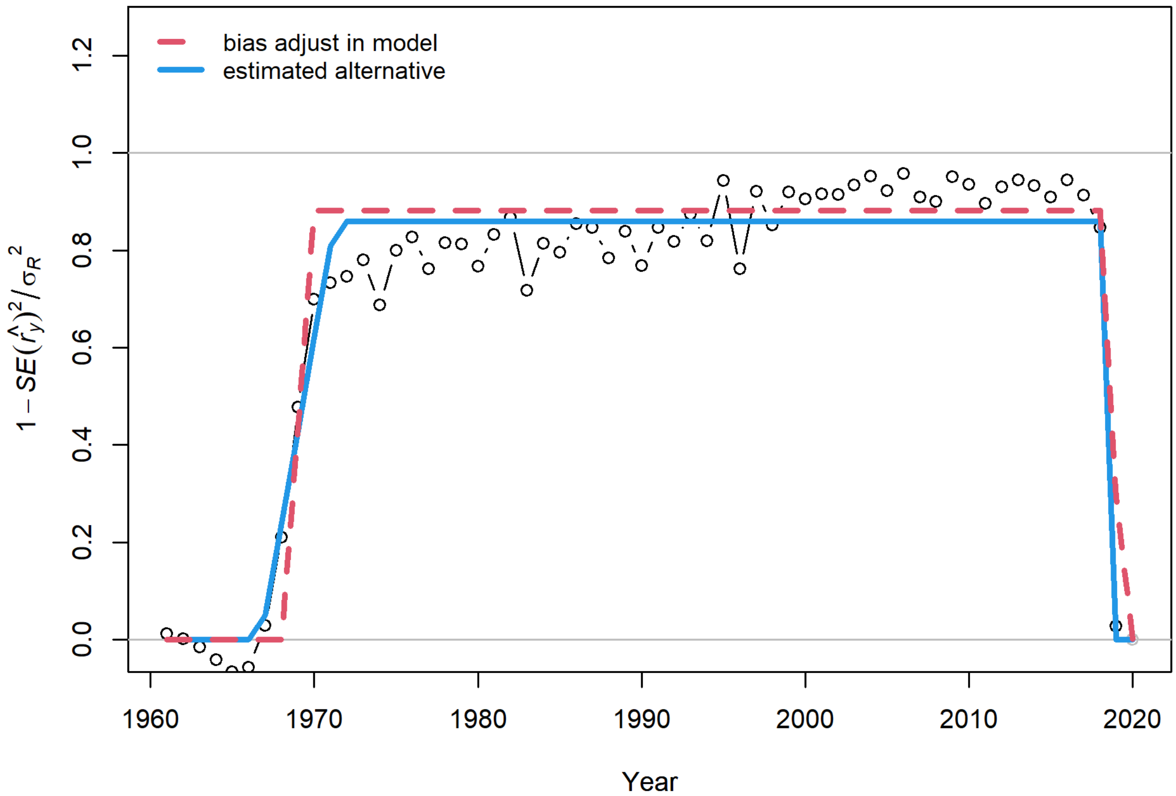


Figure 31. Estimated bias adjustment used to correct recruitment deviations in the old growth model. Red dashed line indicates the adjustment used in the model, blue indicates a re-estimated alternative.

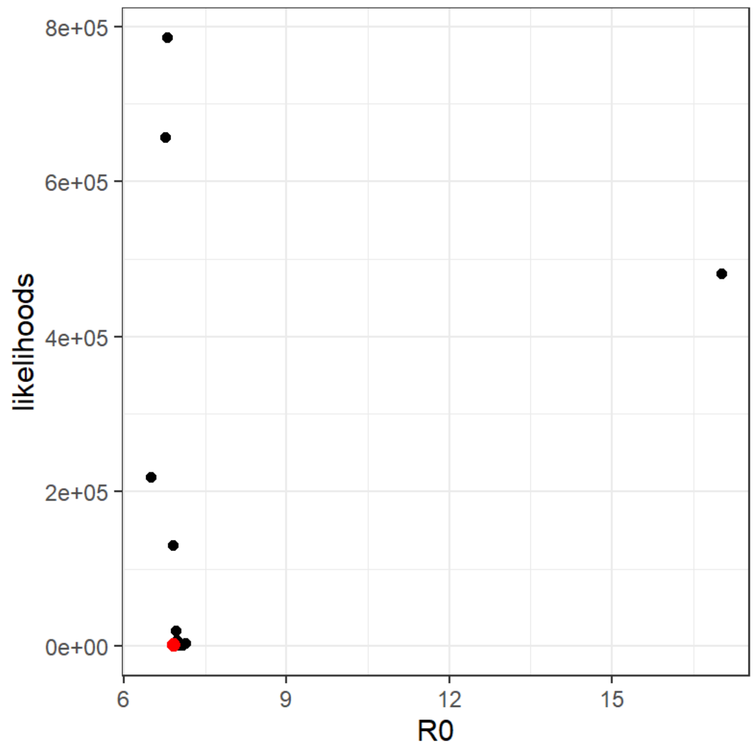


Figure 32. Results of a randomized initial parameter value diagnostic for the new growth model where 30 randomized initial conditions were used with a CV of 10% assigned to each parameter. Results are shown for the new growth model (MLE, solid red circle) and the new growth model with randomized initial parameter values (Jitter runs, solid black circle).

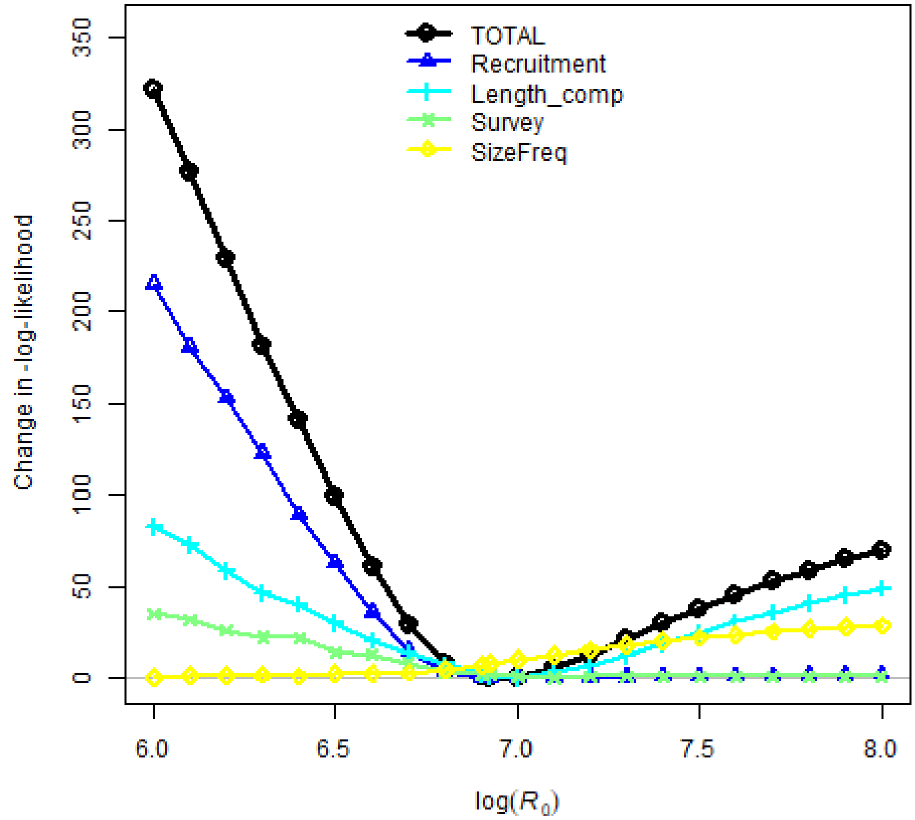


Figure 33. Profiles of the negative log-likelihoods relative to the minimum value of each component for the different likelihood components affecting the unfished recruitment parameter R_0 in log-scale (i.e., the x-axis is log(R_0)) ranging from 6.0 to 8.0 for the new growth model, where recruitment represents the likelihood component based on the deviations from the stock-recruitment curve, sizefreq represents the joint likelihood component for combined fleets based on the fish generalized size composition data, and length data represents the joint likelihood component for combined fleets based on the fish length composition data.

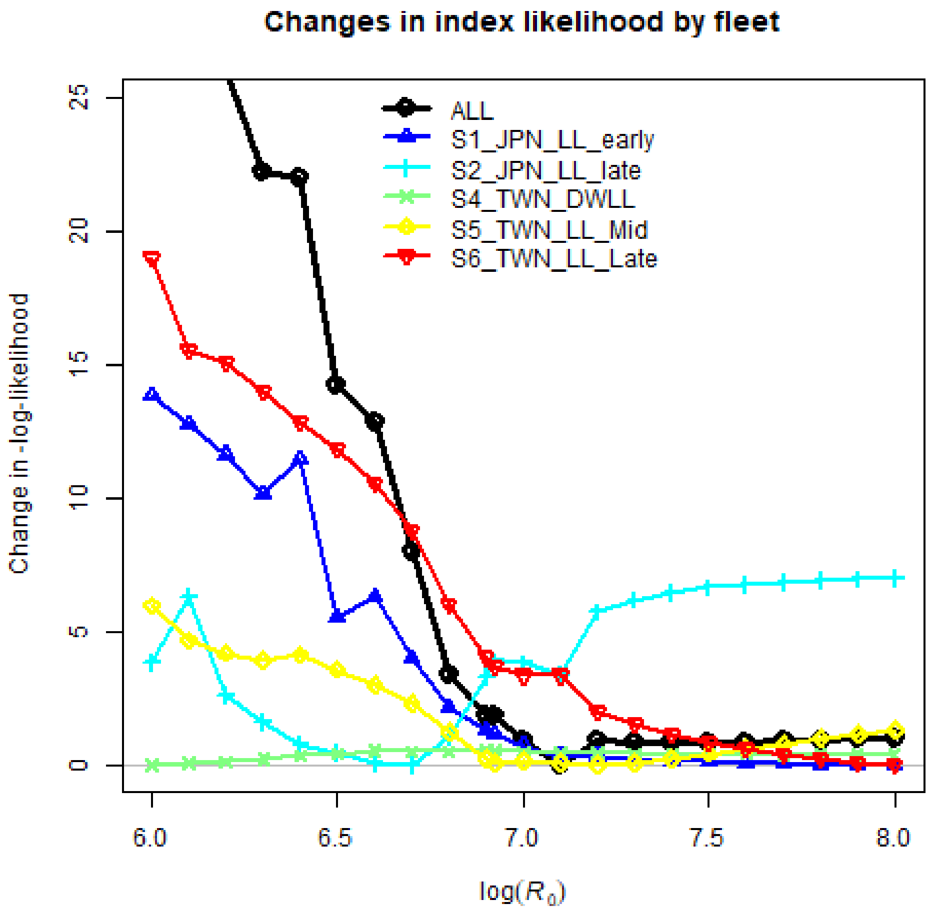


Figure 34. Profiles of the relative negative log-likelihoods by fleet-specific index likelihood components for the virgin recruitment in log-scale (log(R_0)) ranged from 6.0 to 8.0 of the new growth model. See Table 3 for descriptions of the index data. S3 was not included in the total likelihood.

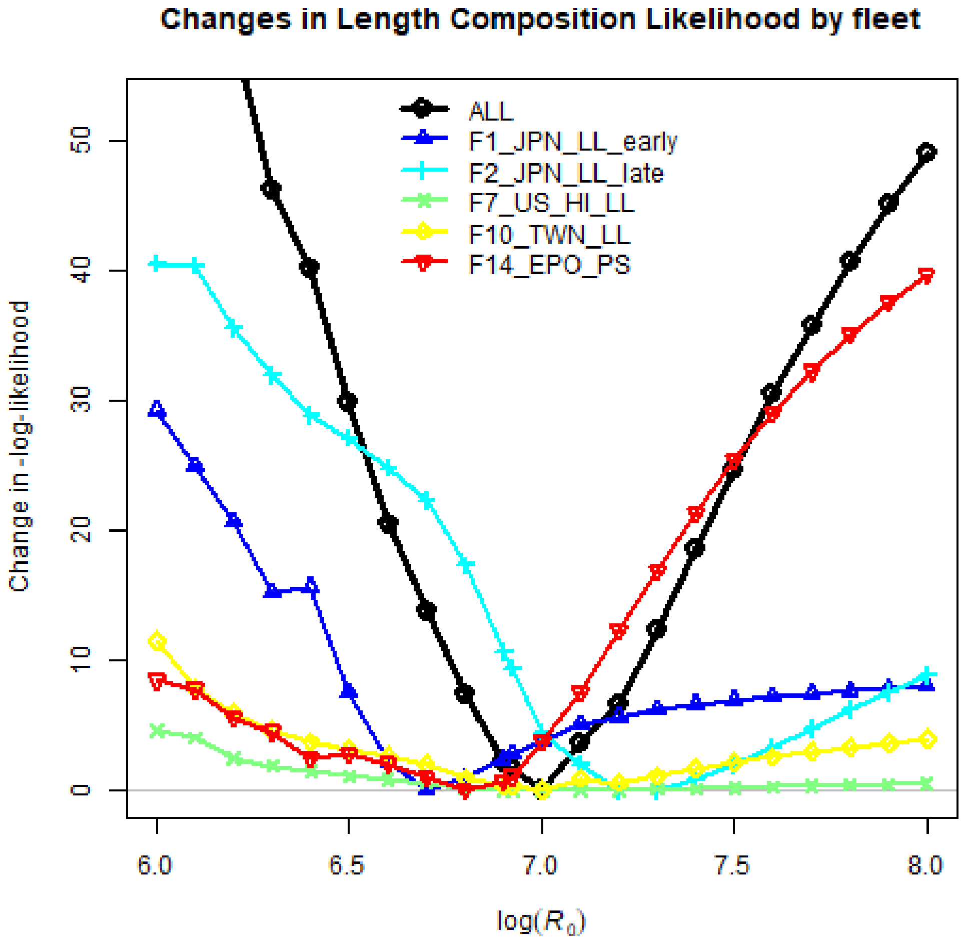


Figure 35. Profiles of the relative negative log-likelihoods by fleet-specific length composition likelihood components for the virgin recruitment in log-scale ($\log(R_0)$) ranged from 6.0 to 8.0 of the new growth model. See Table 3 for descriptions of the length composition data.

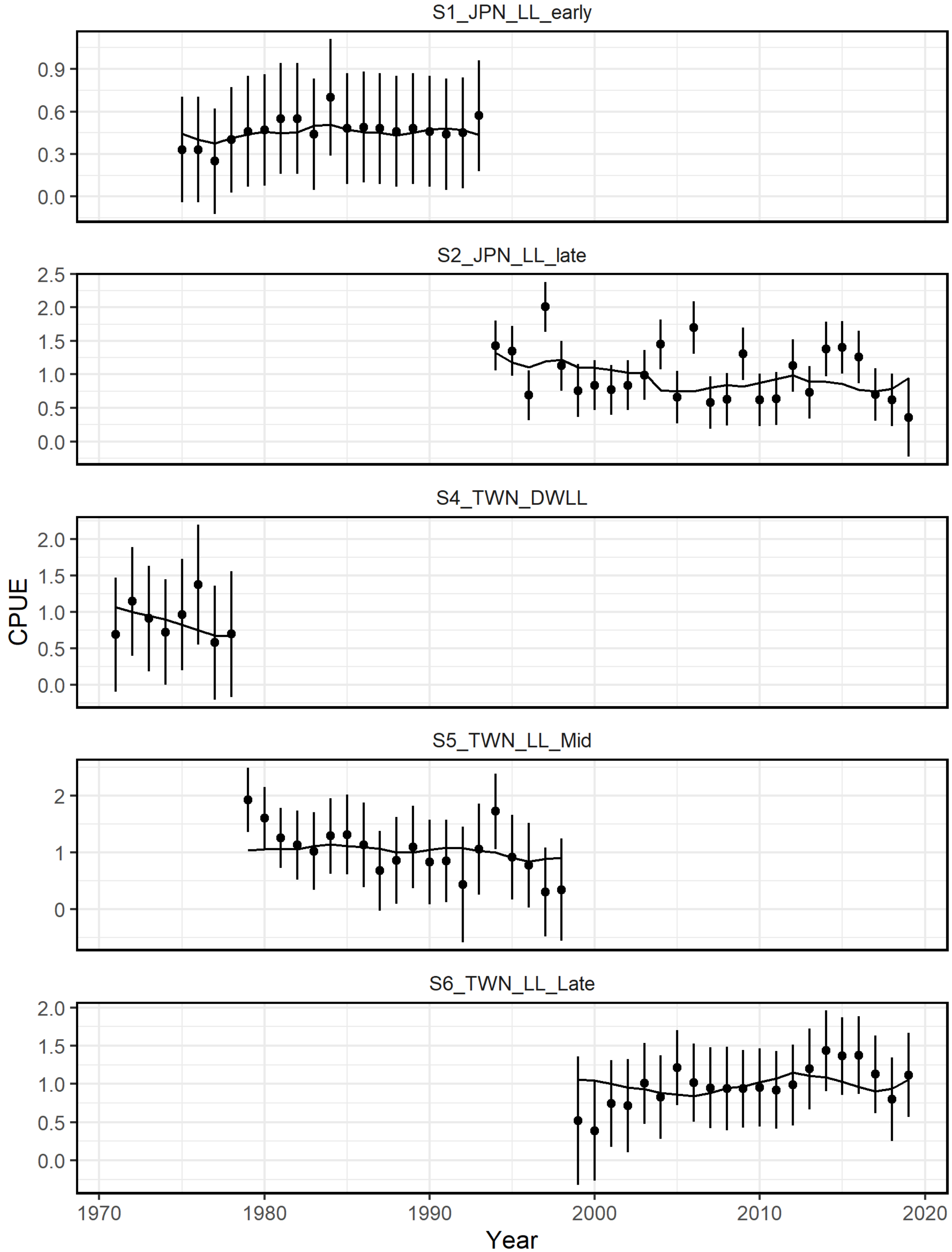


Figure 36. Fits to the standardized catch-per-unit-effort (CPUE) data sets from different fisheries for the new growth model. The line is the model predicted value and the points are observed (data) values. The vertical lines represent the estimated confidence intervals (± 1.96 standard deviations) around the CPUE values. S3 was not included in the total likelihood.

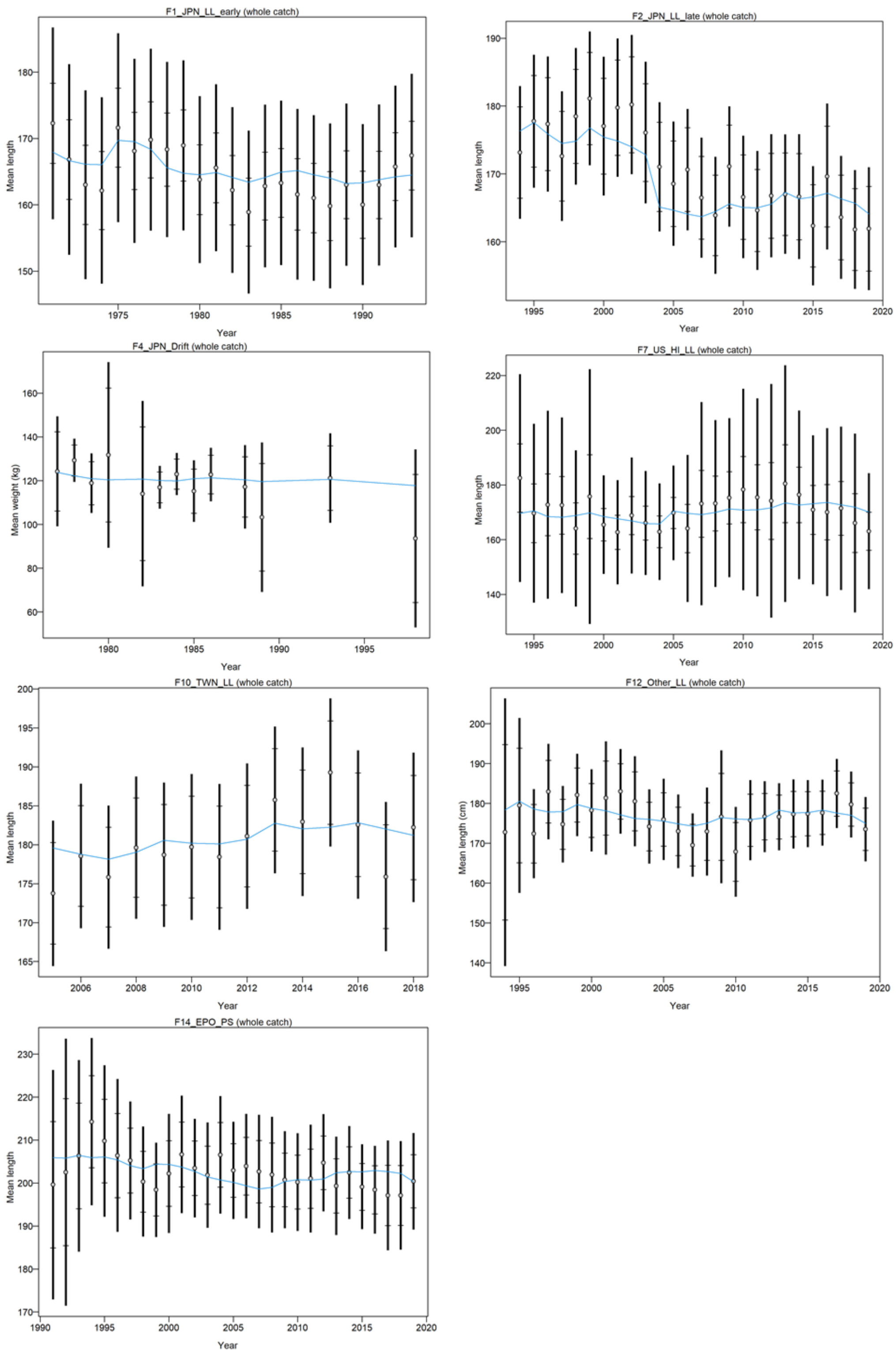


Figure 37. New growth model fit (lines) to mean length of the composition data (points, showing the observed mean age and 95% credible limits around mean age (vertical lines)). See Table 3 for descriptions of the data. All measurements were eye-fork lengths (EFL, cm) except for F4 Japanese Driftnet, which was in weight (kg).

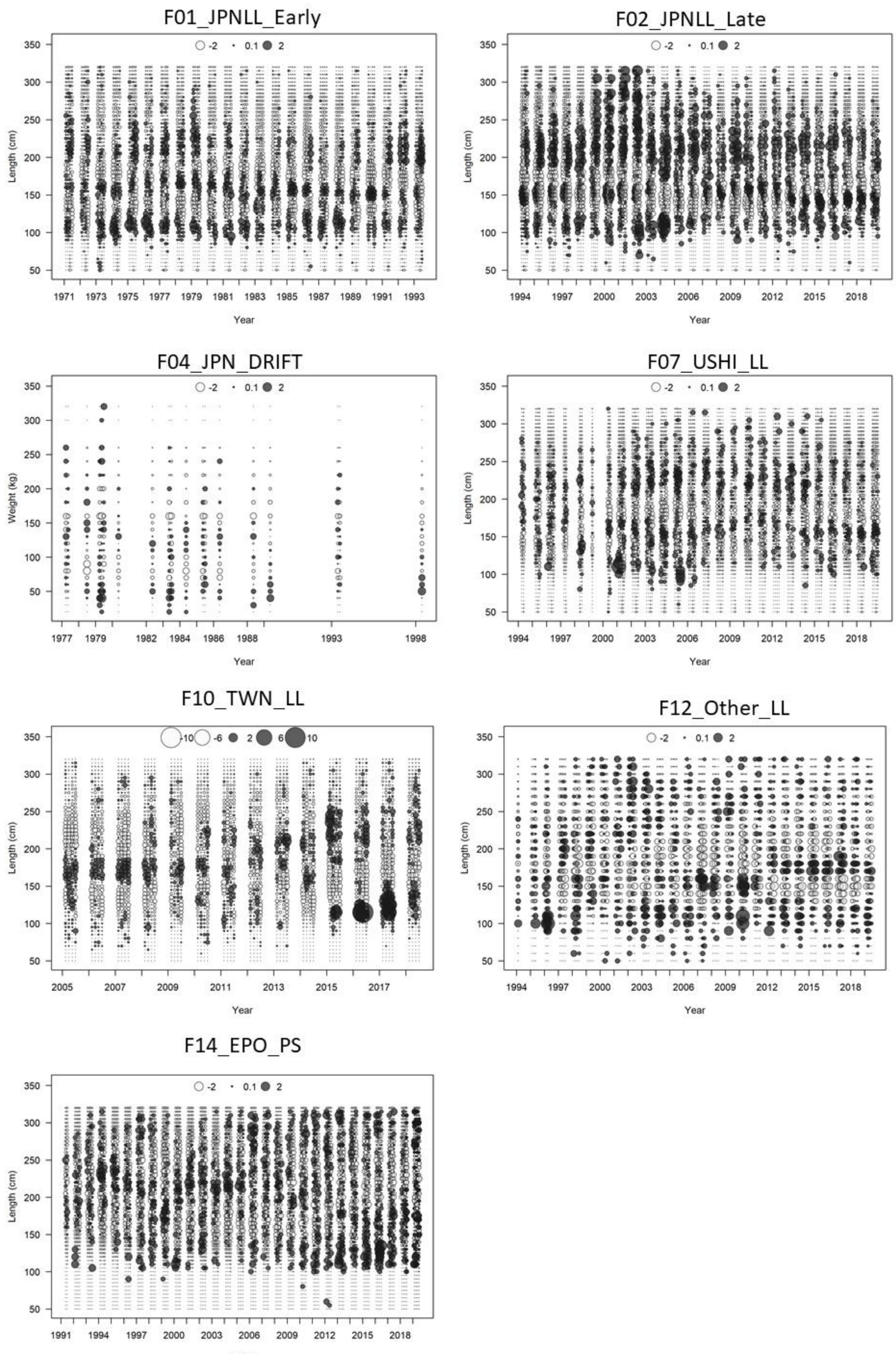


Figure 38. Pearson residual plots of model fits to the various length and weight composition data for the Pacific blue marlin fisheries used in the new growth model.

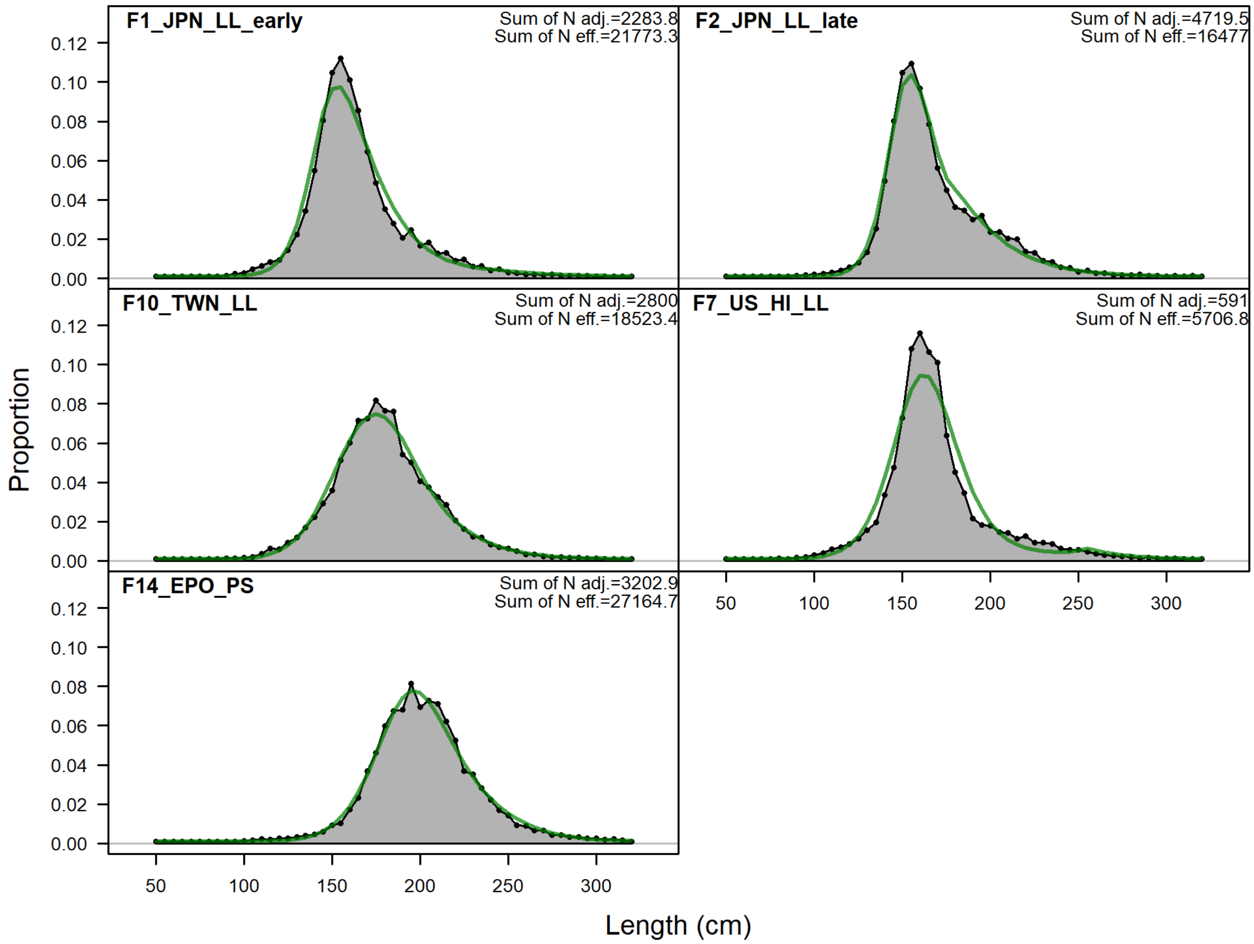


Figure 39. Comparison of observed (gray shaded area and blue dots) and model-predicted (green solid line) length compositions for fisheries used in the Pacific blue marlin new growth model. All measurements were eye-to-fork lengths (EFL, cm).

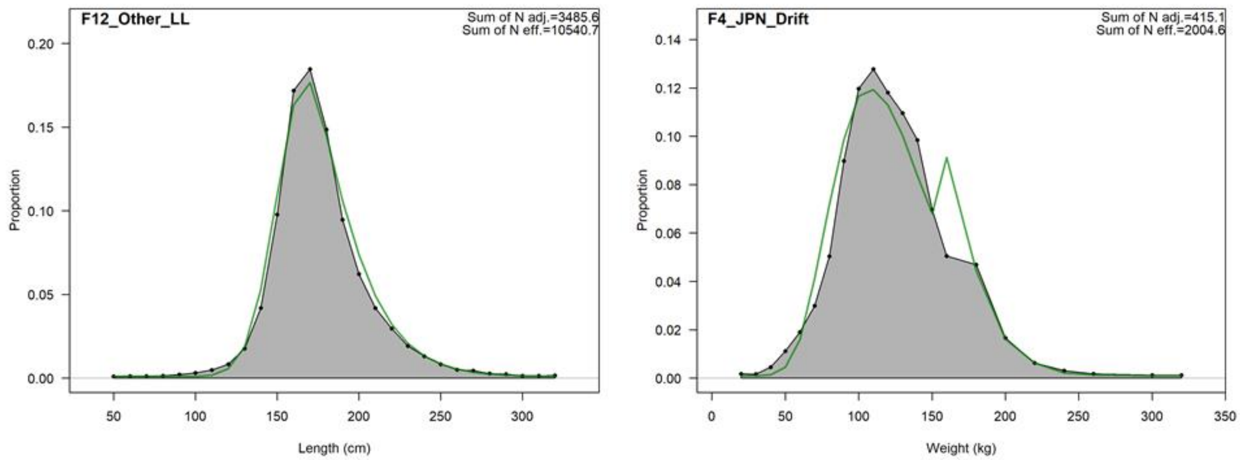


Figure 40. Comparison of observed (gray shaded area and blue dots) and model-predicted (green solid line) generalized size compositions for fisheries used in the Pacific blue marlin new growth model. F12 Other longline is measured in eye-to-fork lengths (EFL, cm), F4 Japanese Driftnet is measured in weight (kg).

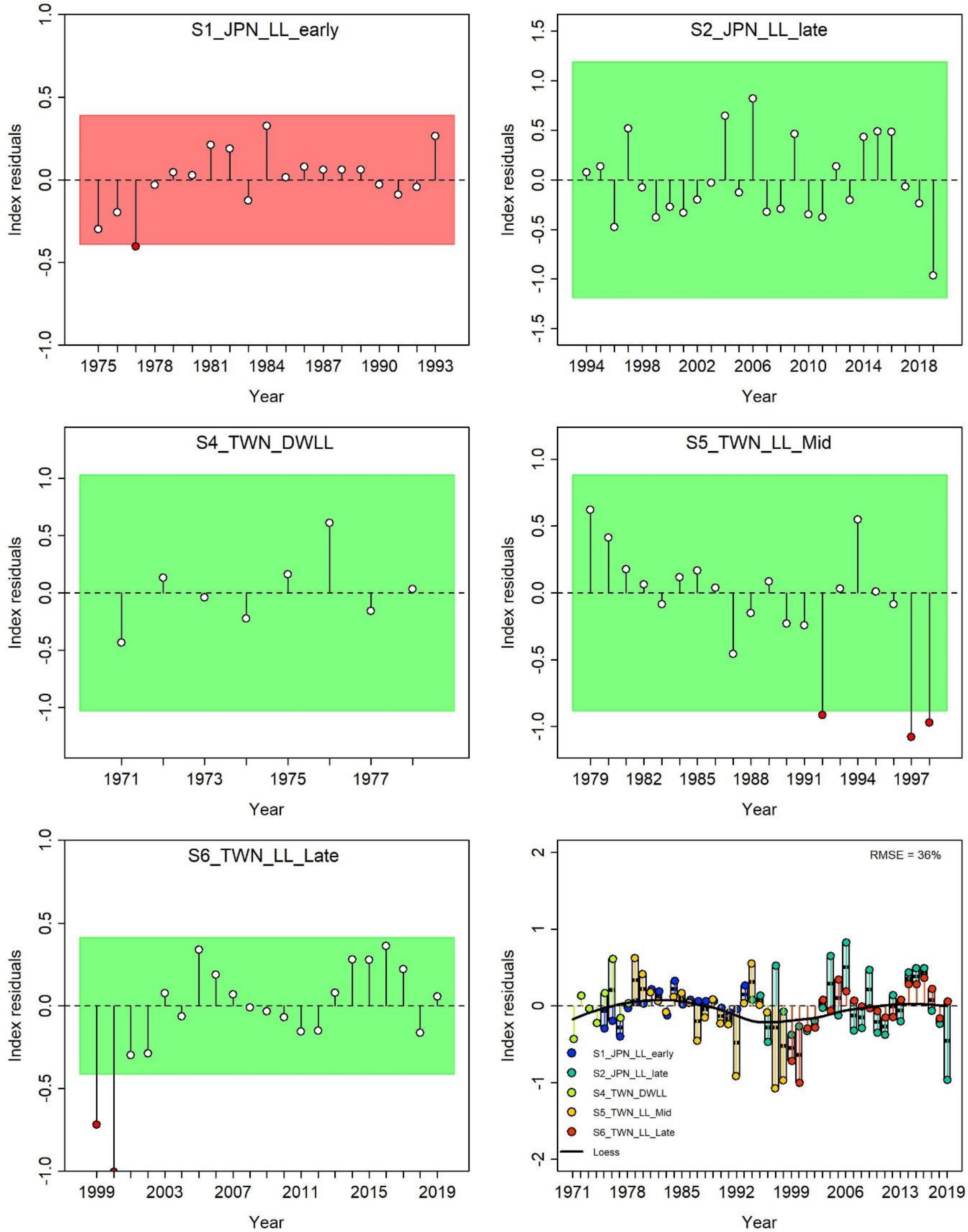


Figure 41. Runs test results for the CPUE fits included in the new growth model. Green shading indicates no evidence ($p \geq 0.05$) and red shading indicates evidence ($p < 0.05$) to reject the hypothesis of a randomly distributed time-series of residuals. The red/green shaded areas span three residual standard deviations to either side from zero, the red points outside of the shading violate the "three-sigma limits" for that series.

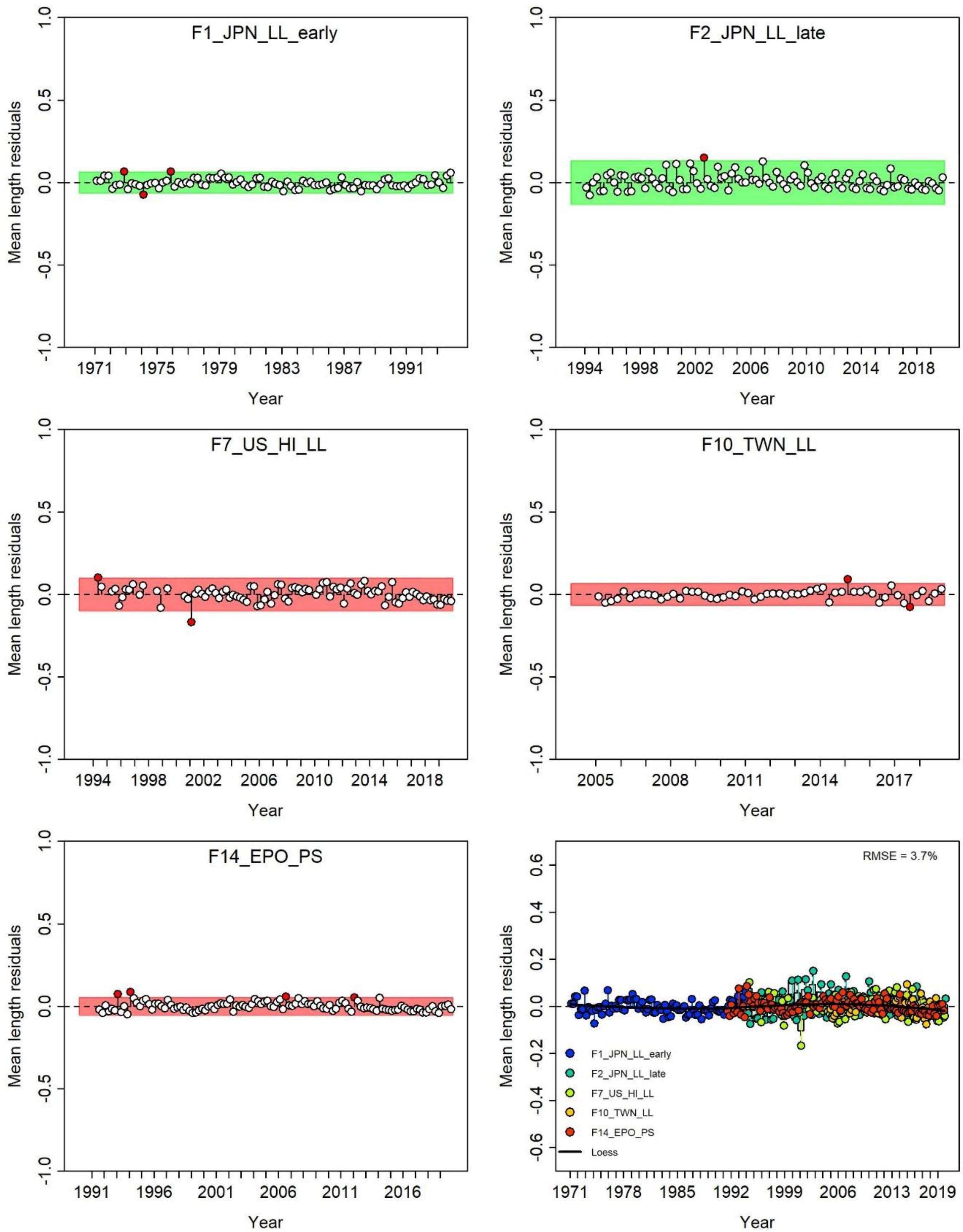


Figure 42. Runs test results for the mean lengths of size composition data included in the new growth model. Green shading indicates no evidence ($p \geq 0.05$) and red shading indicates evidence ($p < 0.05$) to reject the hypothesis of a randomly distributed time-series of residuals. The red/green shaded areas span three residual standard deviations to either side from zero, the red points outside of the shading violate the “three-sigma limits” for that series.

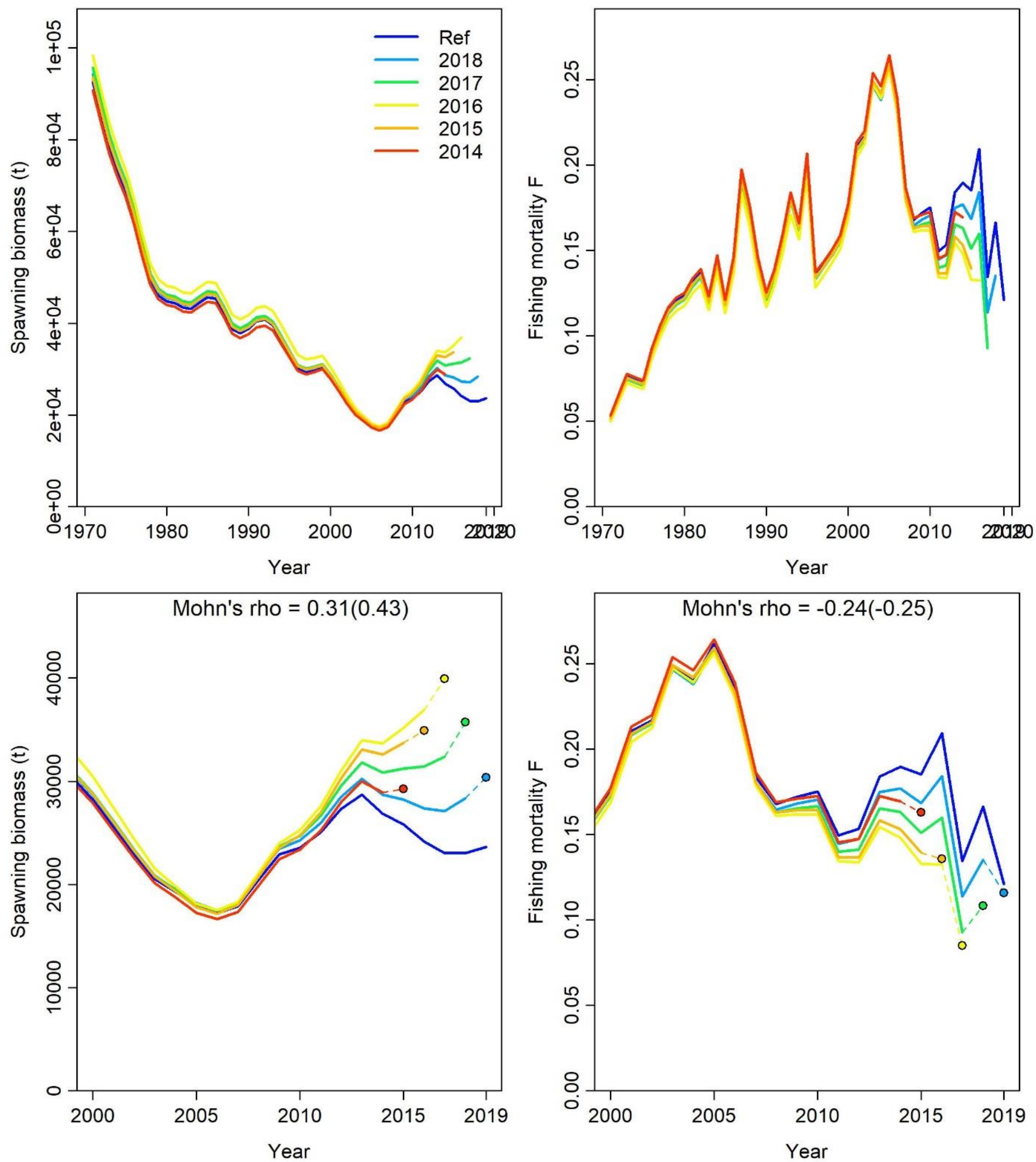


Figure 43. Retrospective analysis of spawning biomass (left) and fishing mortality (right) consisting of 5 reruns of the base case model each fitted with one more year of data removed from the new growth model. The top panels are the entire time series (1971-2019), the bottom panels are the time series since 2000 for visibility.

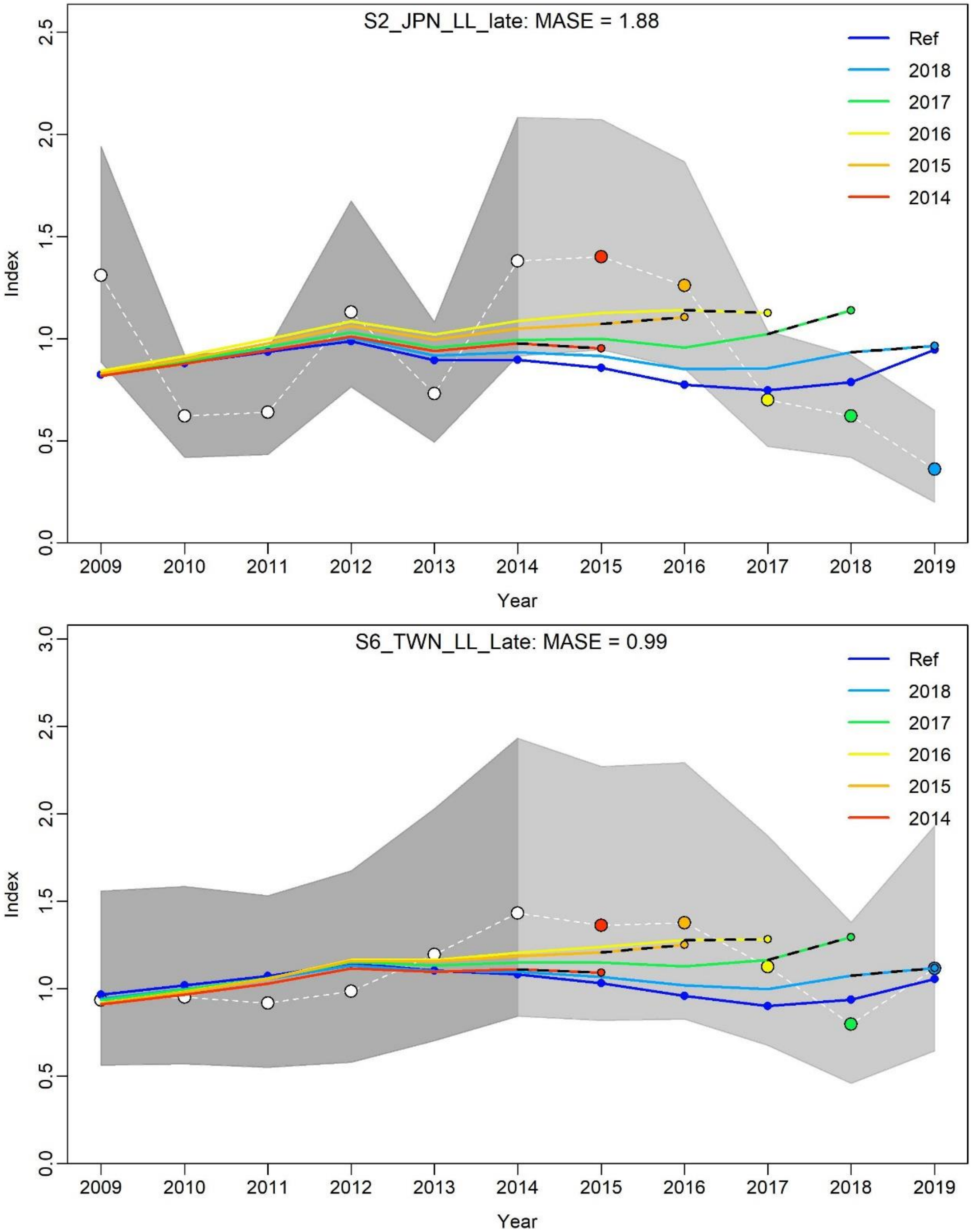


Figure 44. Hind casting cross-validation (HCxval) results for Japanese longline late (top) and Chinese Taipei longline late (bottom) CPUE fits, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the new growth model. The 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.

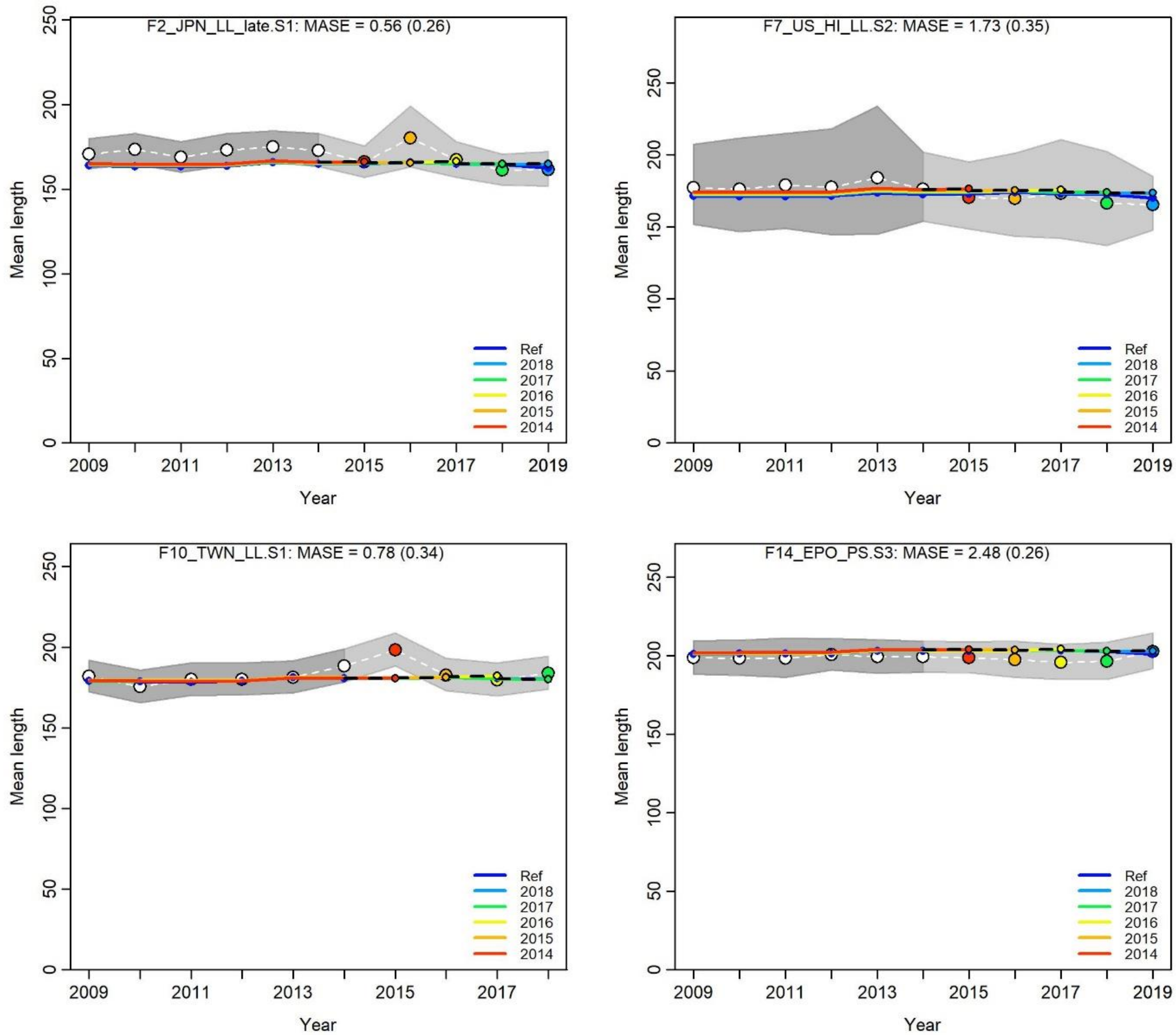


Figure 45. Hind casting cross-validation (HCxval) results for Japanese longline late (top left), US Hawaii longline (top right), Chinese Taipei longline (bottom left), and EPO purse seine (bottom right) size composition mean lengths, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the new 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.

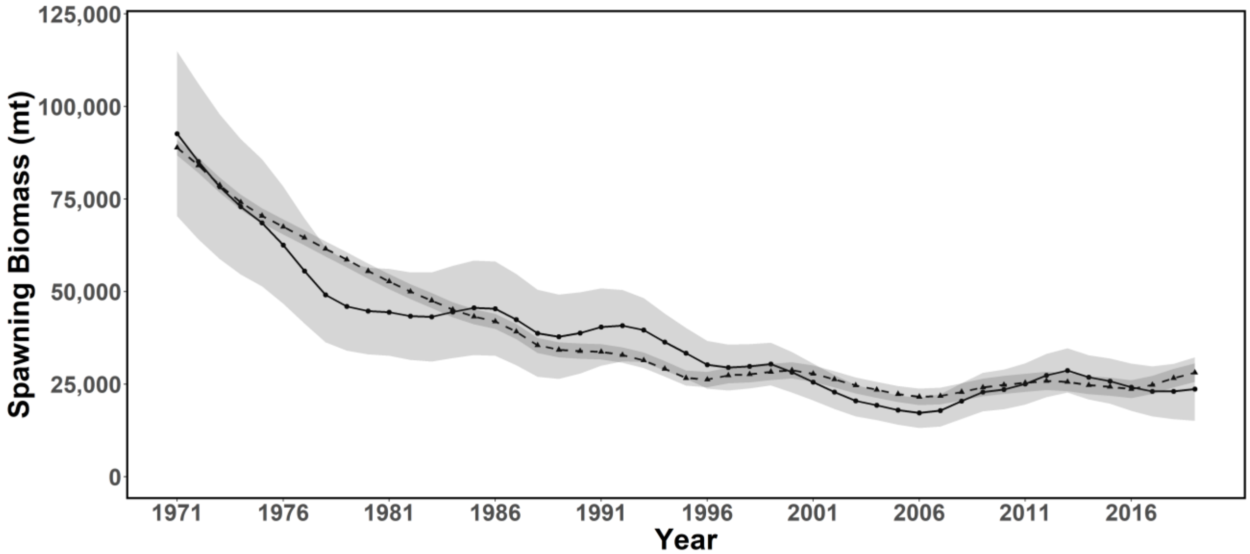


Figure 46. Age structured production model diagnostic for the new growth model. Spawning stock biomass estimates from the full model (black circles, solid line; grey shading indicates 95% confidence interval) and ASPM model diagnostic (black triangles, dashed line; grey shading indicates 95% confidence interval).

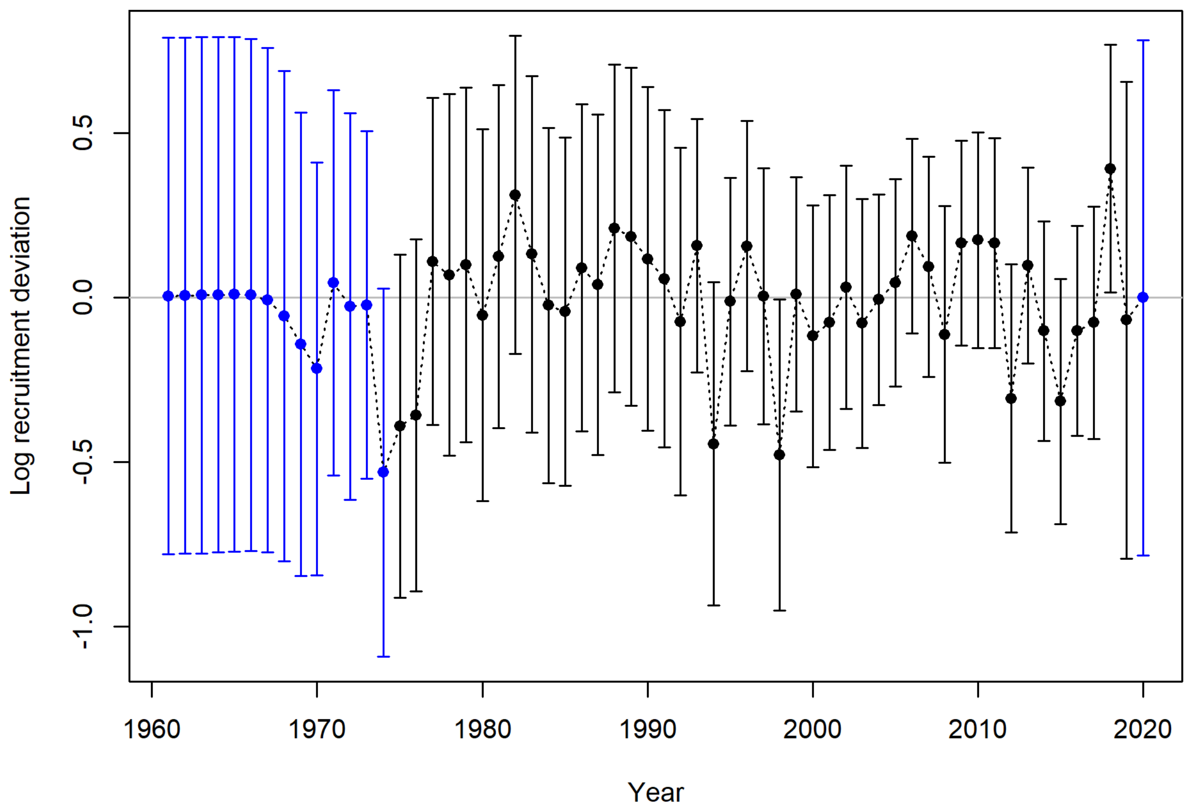


Figure 47. Recruitment deviations (log scale) for the new growth model with sigmaR = 0.4. Blue indicates early recruitment deviations from 1960-1971, and forecast recruitment in 2020

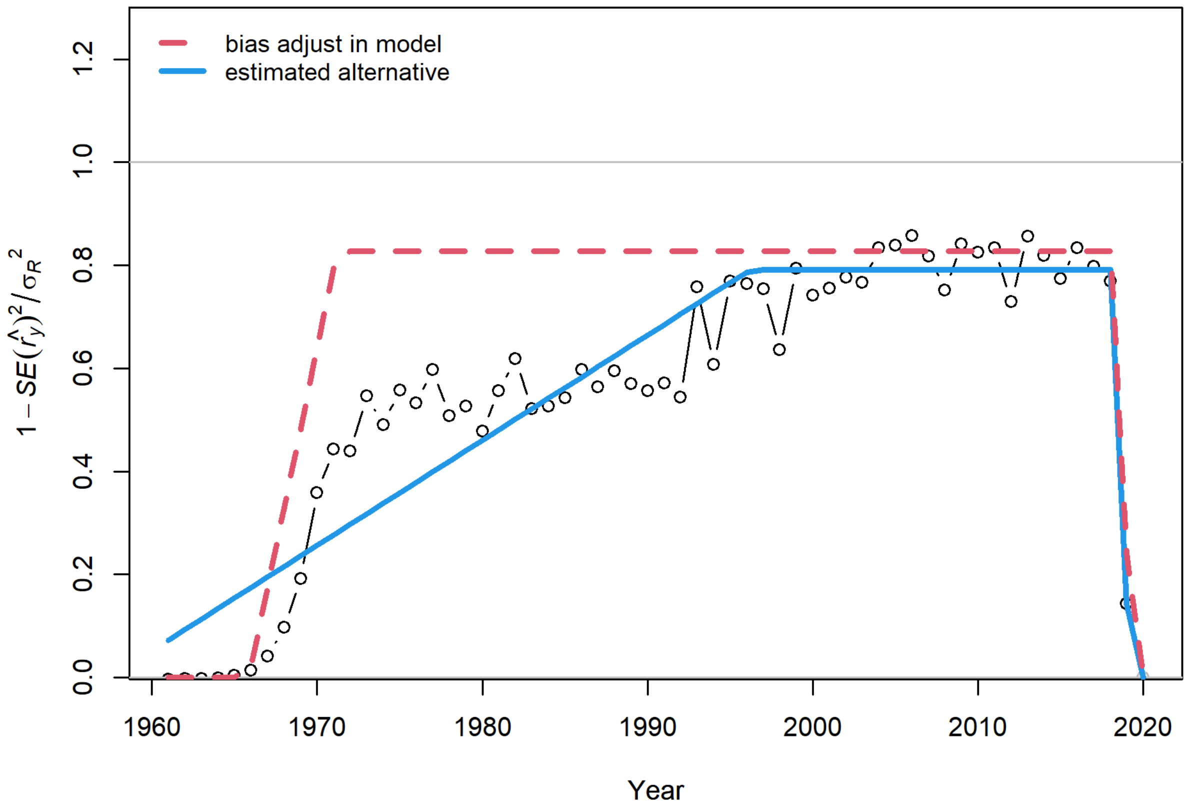


Figure 48. Estimated bias adjustment used to correct recruitment deviations in the new growth model. Red dashed line indicates the adjustment used in the model, blue indicates a re-estimated alternative.

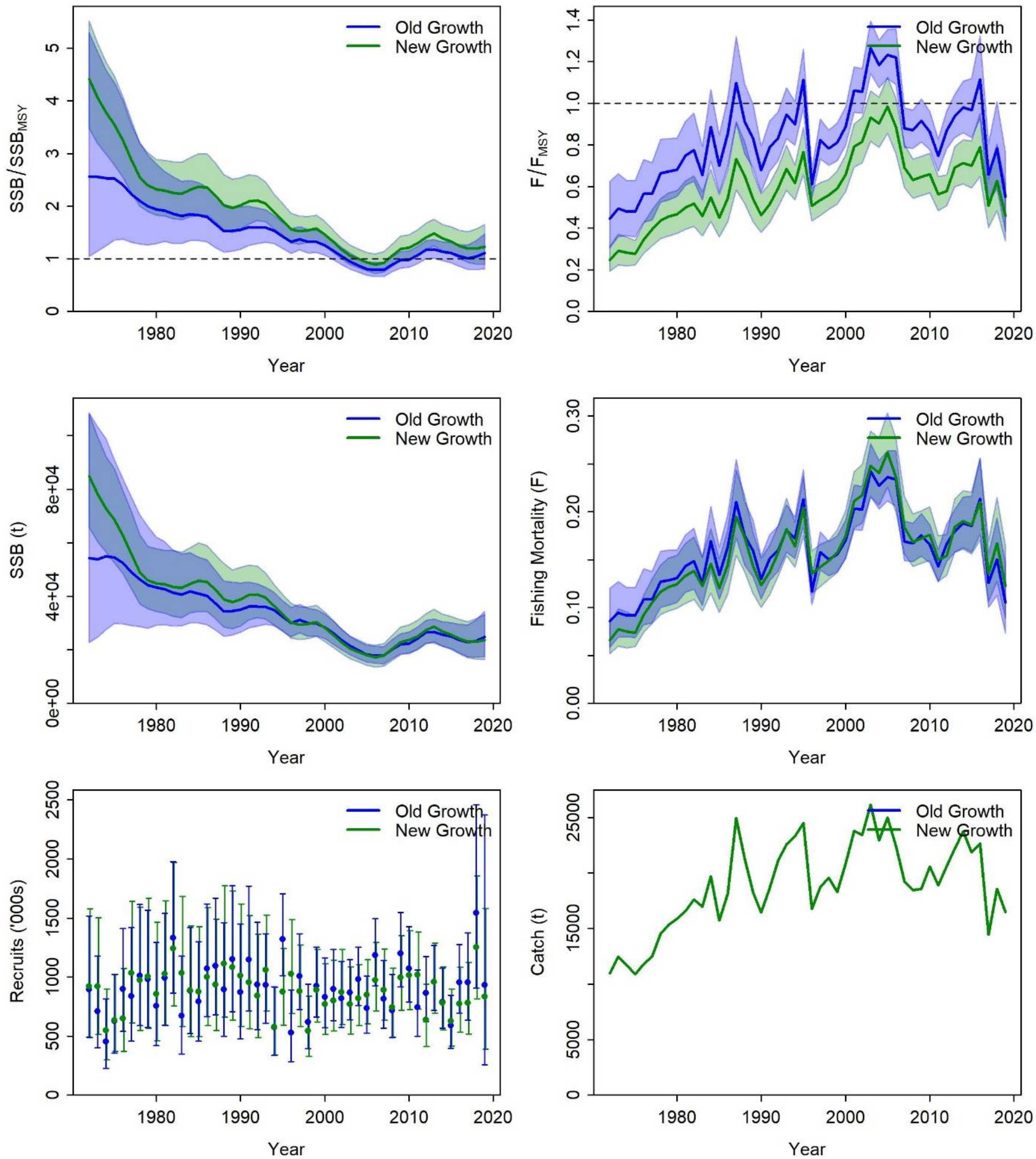


Figure 49. Time-series of SSB/SSB$_{MSY}$ (top left), F/F$_{MSY}$ (top right), spawning stock biomass (center left), instantaneous fishing mortality (center right), recruitment (thousands of age-0 fish, bottom left), and catch (bottom right) for the old growth (blue) and new growth (green) models. Shading indicates 95% confidence intervals, lines indicate median value.

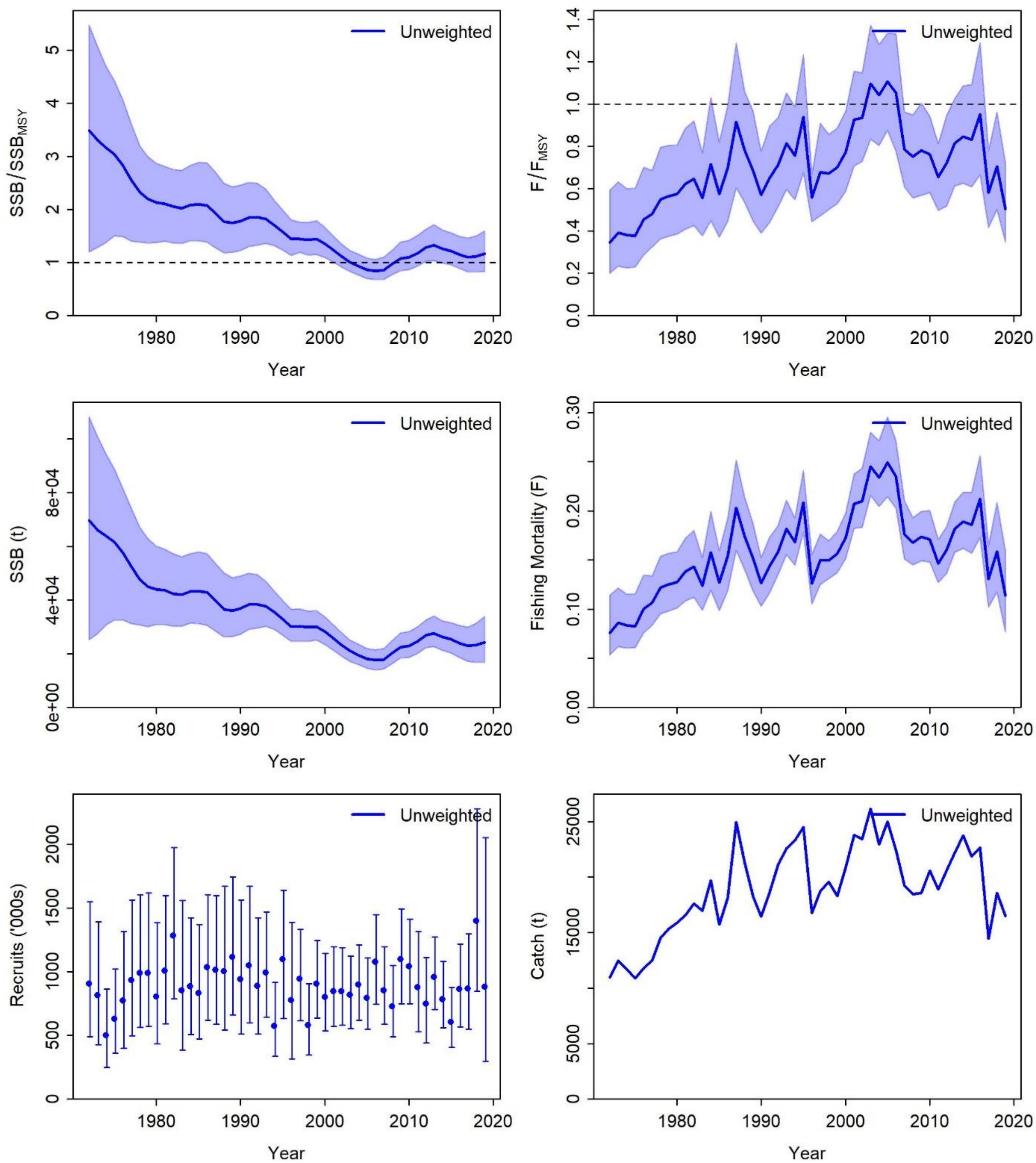


Figure 50. Time-series of SSB/SSB$_{MSY}$ (top left), F/F$_{MSY}$ (top right), spawning stock biomass (center left), instantaneous fishing mortality (center right), recruitment (thousands of age-0 fish, bottom left), and catch (bottom right) for the ensemble model with equal weightings (i.e. unweighted models) based upon 10,000 draws each from a multivariate lognormal distribution for each growth model. Shading indicates 95% confidence intervals, lines indicate median value.

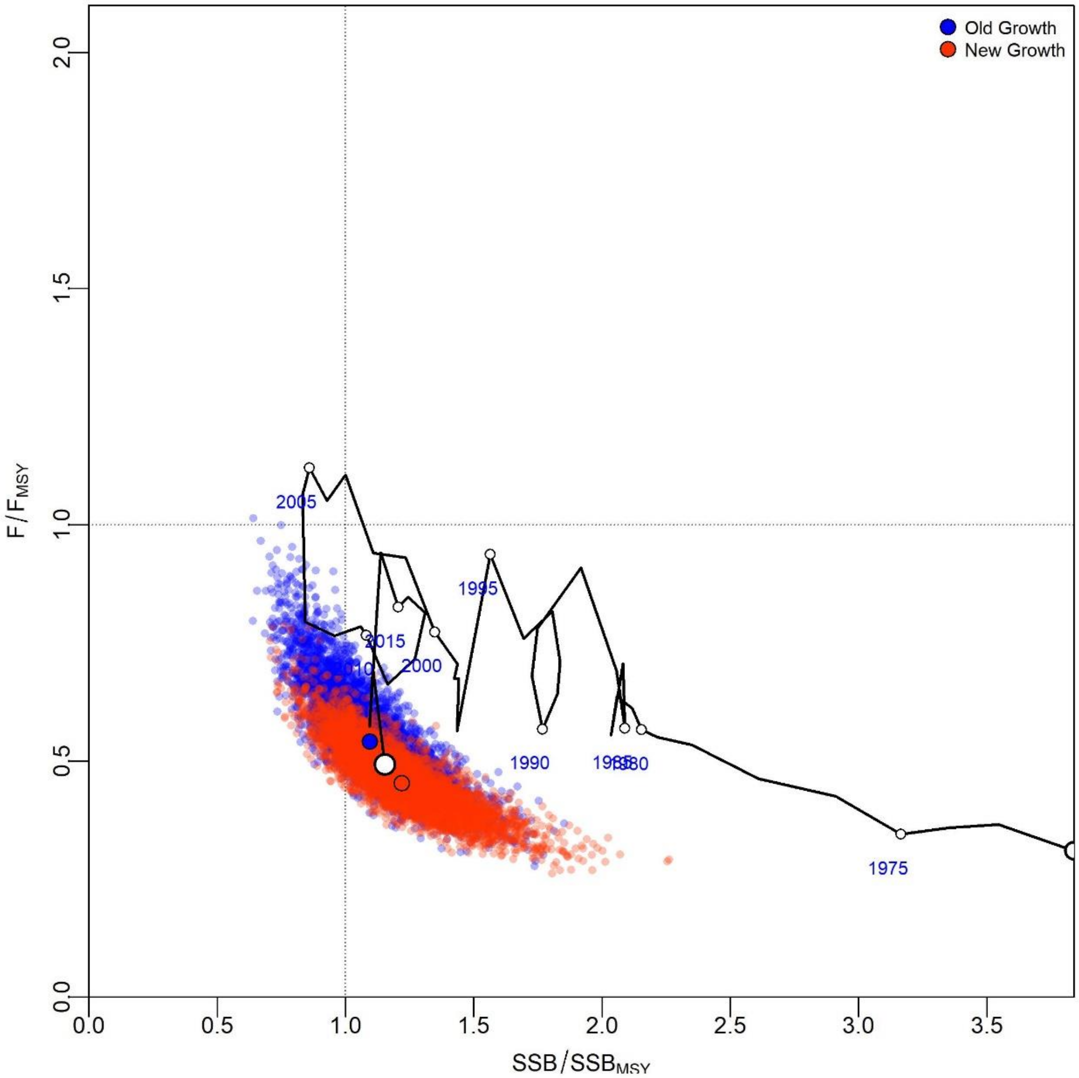


Figure 51. Kobe plot of the trends in estimates of relative fishing mortality (average of age 1-10) and spawning stock biomass of Pacific blue marlin (*Makaira nigricans)* during 1971-2019. The solid line and labeled points represent the ensemble model status estimates per year. The large white dot indicates the terminal year (2019) status from the ensemble model; the large red dot is the 2019 stock status from the new growth model; the large blue dot indicates the 2019 stock status from the old growth model. The cloud of points represents the 10,000 draws from the multivariate lognormal distribution for the old growth (blue) and new growth (red) models to visualize the joint uncertainty from the two models.

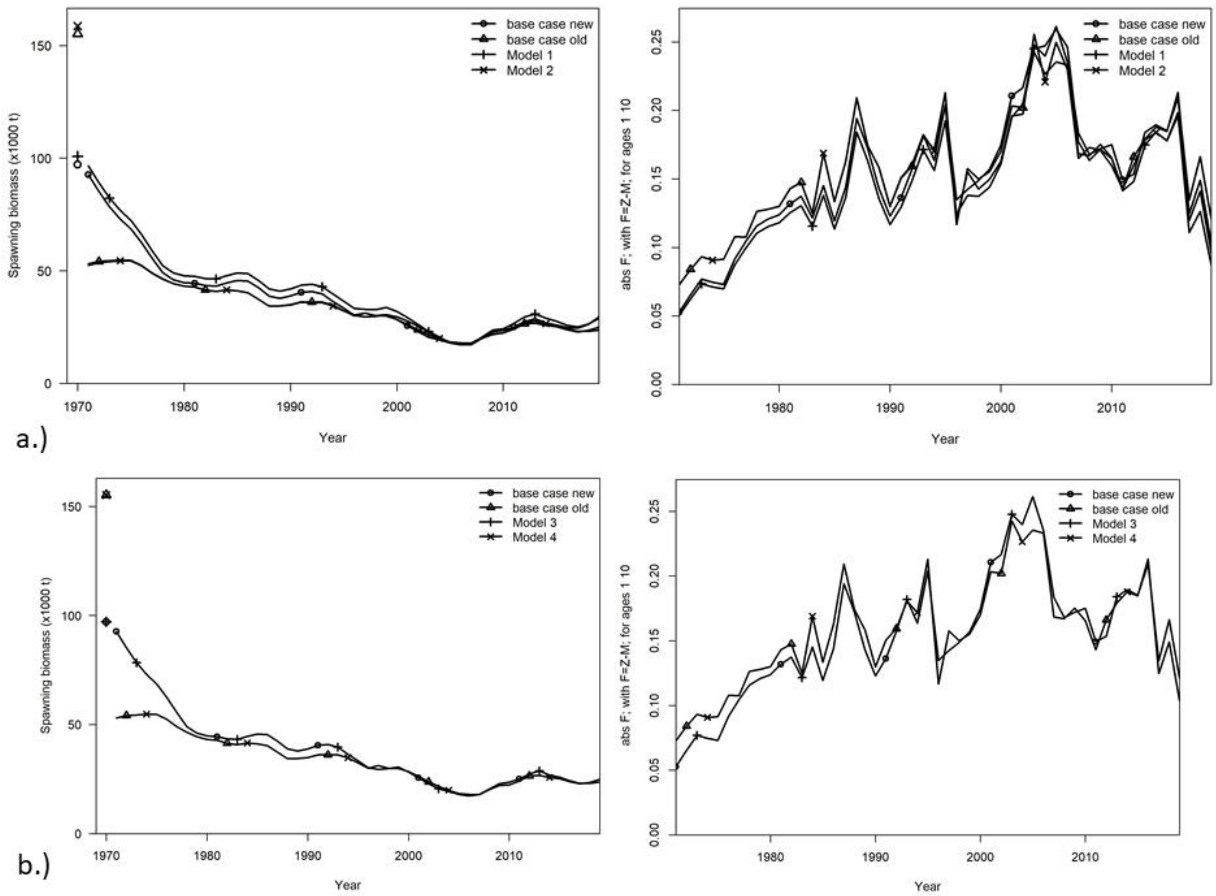
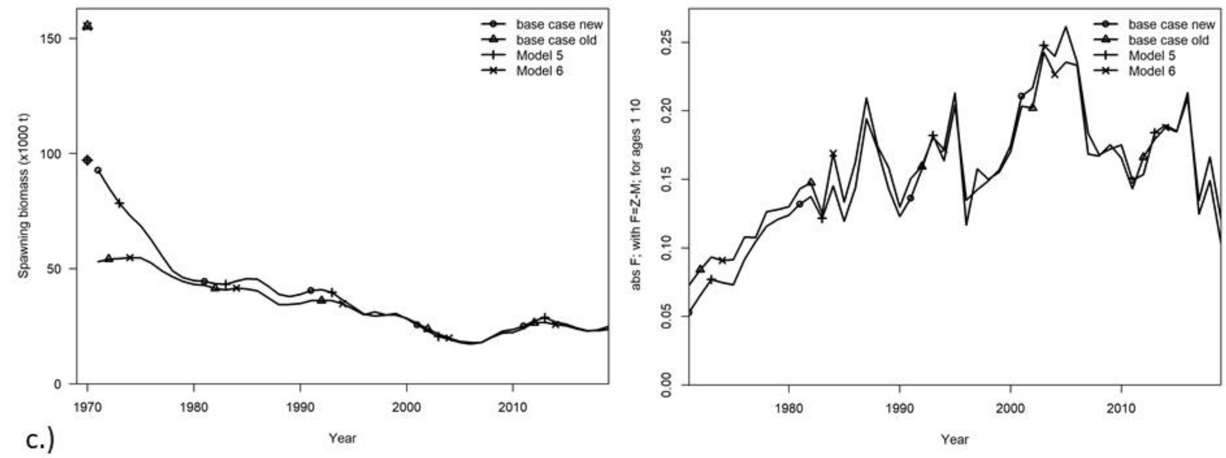


Figure 52. Trajectories of spawning stock biomass and fishing mortality from 30 sensitivity analyses listed in Table 16, compared to the old growth and new growth models: (a) Runs 1 and 2 drops the Japanese longline late CPUE index (S2); (a) Runs 3 and 4 drops the Chinese Taipei longline late CPUE index (S6); (c) Runs 5 and 6 down-weight all length composition data to 0.5; (d) Runs 7 through 10 use alternative natural mortality; (e) Runs 11 through 14 use alternative growth L_{inf} parameters; (f) Runs 15 through 18 use alternative growth curve K parameters; (g) Runs 19 through 22 use alternative maturity ogives; (h) runs 23 through 28 use alternative steepness parameters; (i) Runs 29 and 30 fix initial equilibrium catches.



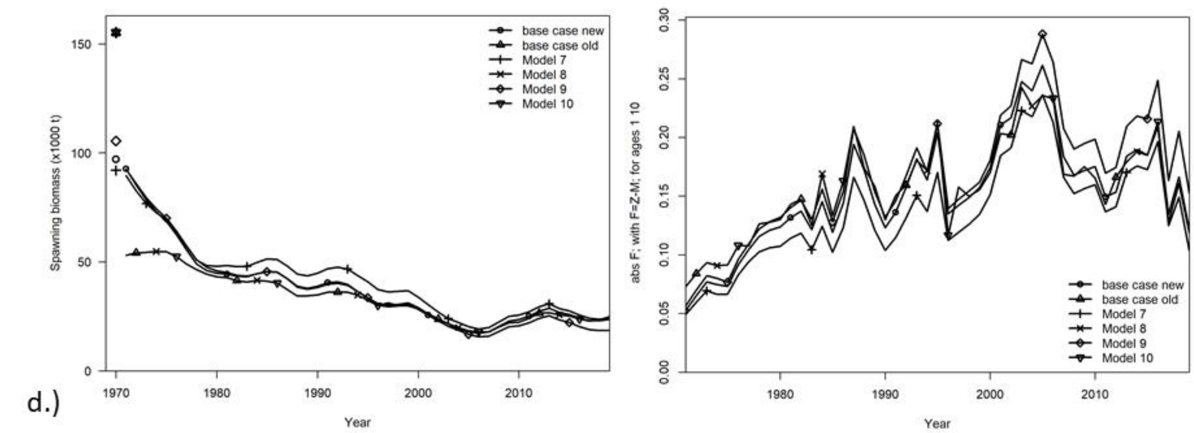
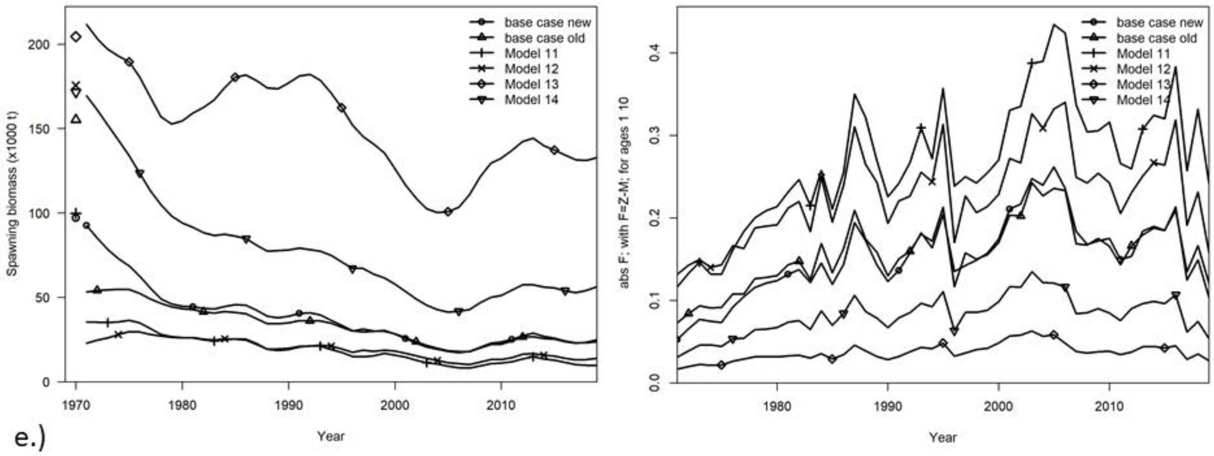
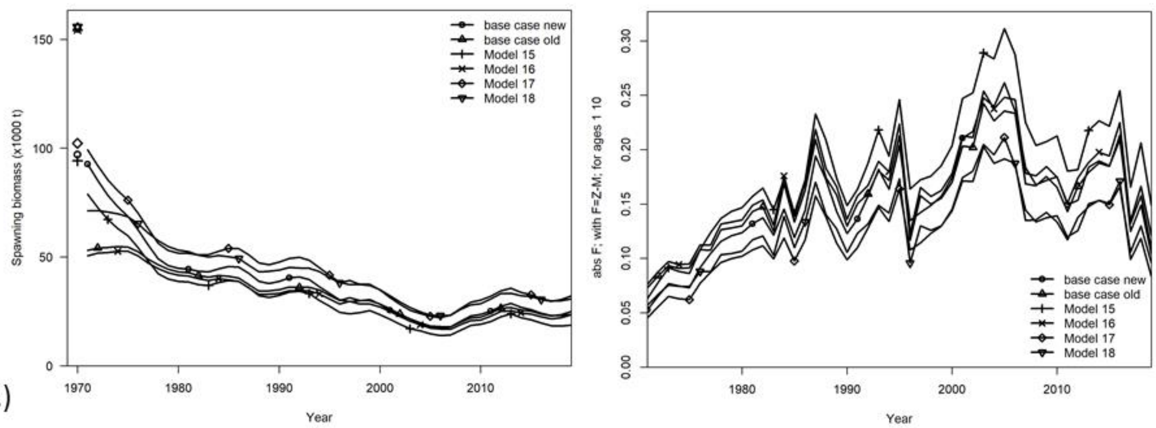


Figure 52. Continued



e.)



f.)

Figure 52. Continued

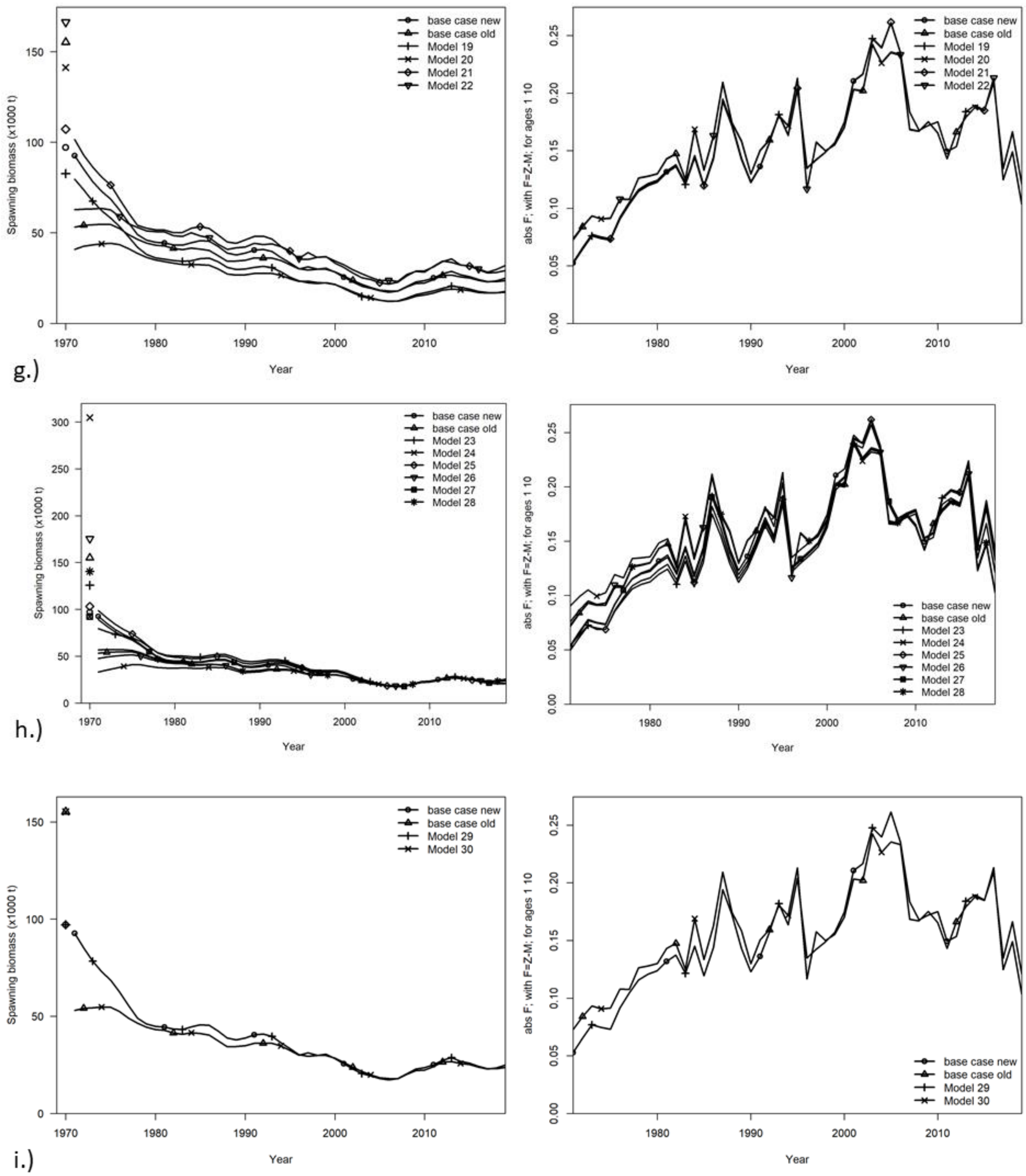


Figure 52. Continued

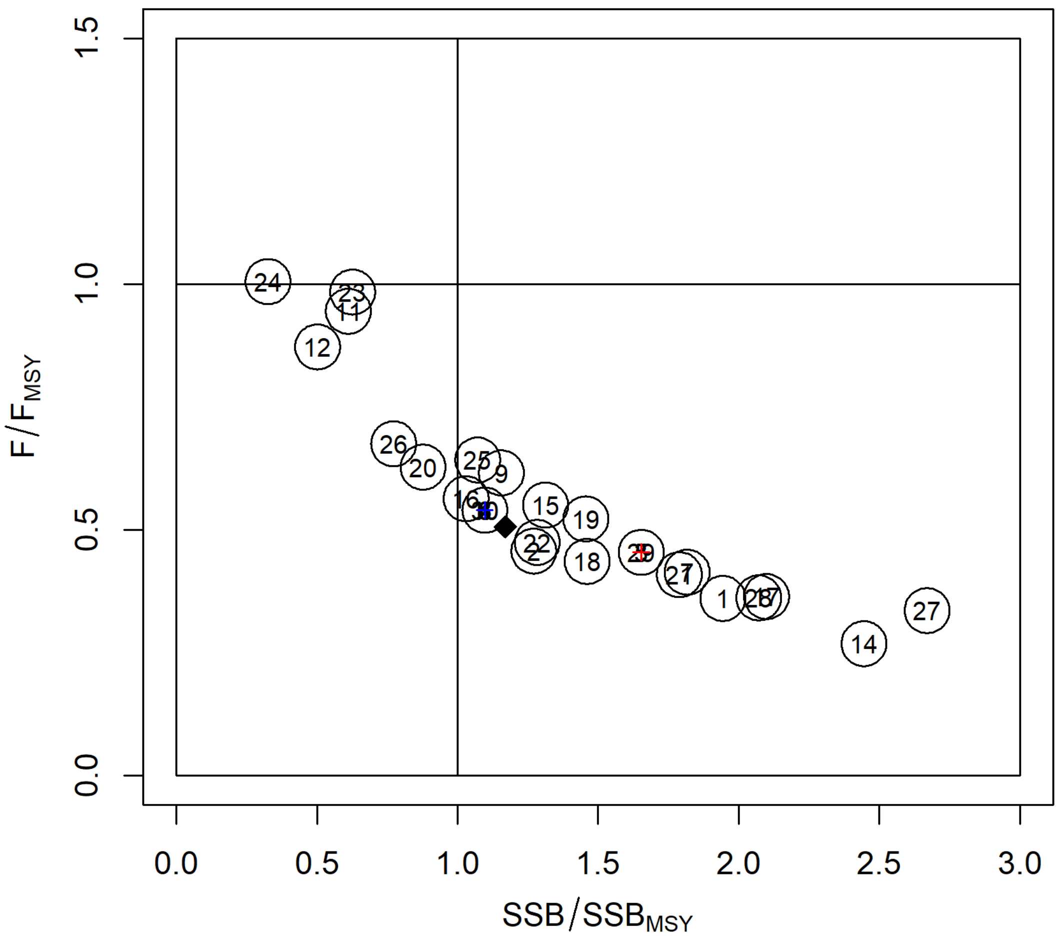


Figure 53. Kobe plot showing the terminal-year stock status for the ensemble model (black diamond), new growth model (red x), old growth model (blue x), and the sensitivity analyses as indicated by the run numbers. For the list of sensitivity runs, please see Table 16.

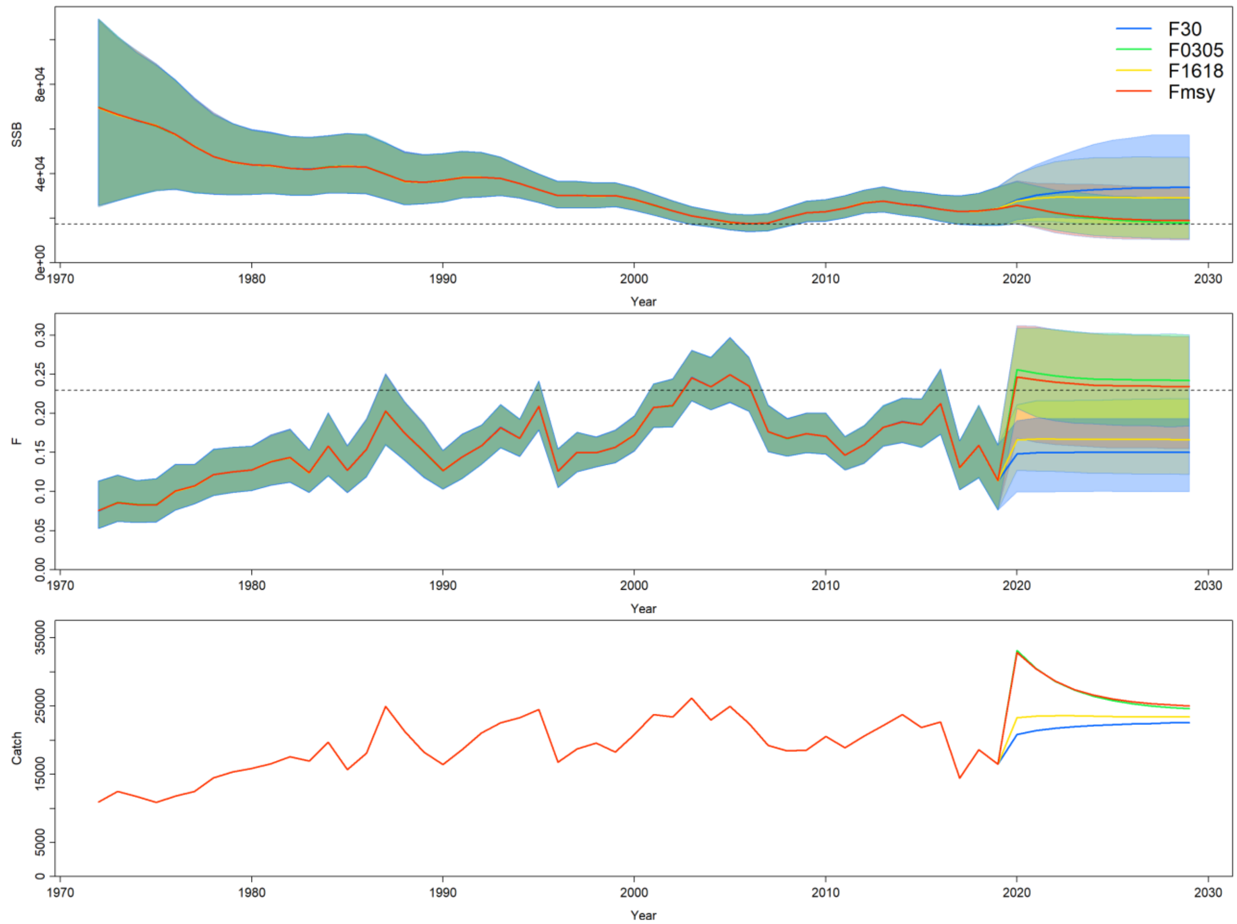


Figure 54. Historical and projected trajectories of spawning biomass and total catch from the Pacific blue marlin ensemble models based upon the four F scenarios: projected spawning biomass, dotted line indicates SSB_{MSY}, shading indicates 95% confidence intervals (top); projected instantaneous fishing mortality (ages 1-10 $year^{-1}$), dotted line indicates F_{MSY}, shading indicates 95% confidence intervals (center); and projected catch (mt. bottom). Green indicates scenario 1, $F_{2003\text{-}2005}$; red indicates scenario 2, F_{MSY}; yellow indicates scenario 3, $F_{2016\text{-}2018}$; and blue indicates scenario 4, $F_{30\%}$. The list of projection scenarios can be found in Table 17.

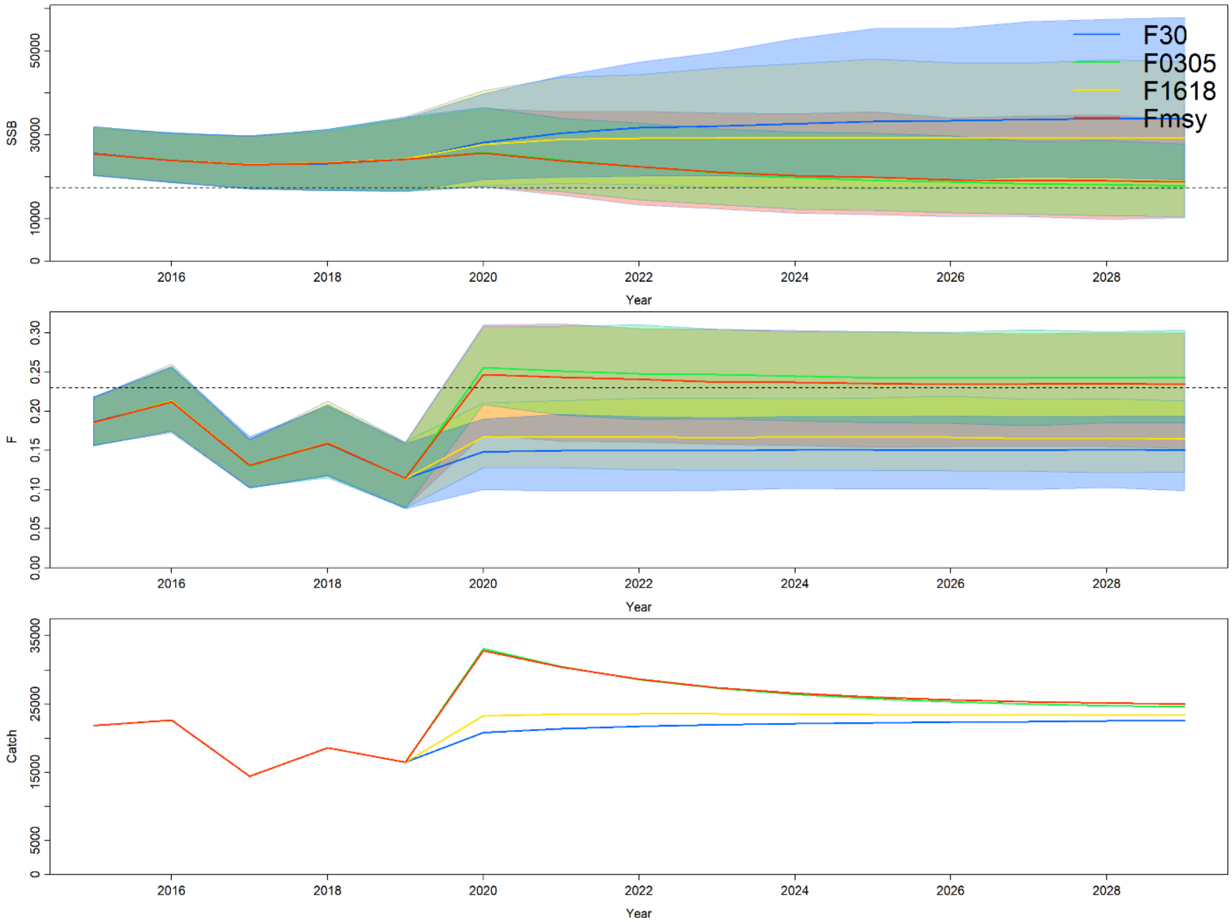


Figure 55. Historical and projected trajectories of spawning biomass and total catch from the Pacific blue marlin ensemble model since 2015 based upon the four F scenarios: projected spawning biomass, dotted line indicates SSB_{MSY}, shading indicates 95% confidence intervals (top); projected instantaneous fishing mortality (ages 1-10 $year^{-1}$), dotted line indicates F_{MSY}, shading indicates 95% confidence intervals (center); and projected catch (mt. bottom). Green indicates scenario 1, $F_{2003\text{-}2005}$; red indicates scenario 2, F_{MSY}; yellow indicates scenario 3, $F_{2016\text{-}2018}$; and blue indicates scenario 4, $F_{30\%}$. The list of projection scenarios can be found in Table 17.