Stock Assessment Update for Blue Marlin (*Makaira nigricans*) in the Pacific Ocean through 2014

Yi-Jay Chang[1], Brian Langseth[2], Hirotaka Ijima[3], and Mikihiko Kai[3]

[1]Joint Institute for Marine and Atmospheric Research, University of Hawaii
NOAA Fisheries, 1845 Wasp Blvd., Honolulu, HI, 96818

[2]NOAA Fisheries, National Marine Fisheries Service,
Pacific Islands Fisheries Science Center,
1845 Wasp Blvd., Honolulu, HI, 96818

[3]National Research Institute of Far Seas Fisheries
5-7-1 Orido, Shimizu
Shizuoka, Japan 424-8633

Abstract

We present an update of the stock assessment for the Pacific blue marlin (*Makaira nigricans*) stock conducted in 2013 by the ISC Billfish Working Group (BILLWG). The assessment update consisted of running a Stock Synthesis model with newly available catch, abundance index, and length and size composition data for 1971-2014. We used the same model structure and parameters as were used in the base case run from the 2013 stock assessment. The results indicated that biomass (age 1 and older) for the Pacific blue marlin stock fluctuated around 120,000 metric tons from 1971 until 1984, thereafter exhibited a long-term decline to the lowest level of 69,720 metric tons in 2009, and then increased to around 78,000 metric tons for the last three years (2012-2014). Estimated fishing mortality gradually increased from the early 1970s to the mid-2000s, peaked at 0.38 year^{-1} in 2005 in response to higher catches, and declined to 0.28 year^{-1} in the most recent years (2012-2014). Compared to MSY-based reference points, the current spawning biomass (average for 2012-2014) was 23% above SSB_{MSY} and the

current fishing mortality (average for ages 2 and older in 2012-2014) was 14% below F_{MSY}. The base case model indicated that the Pacific blue marlin stock was not overfished and was not subject to overfishing relative to MSY-based reference points. The aim of this working paper is to provide the basic update assessment model and its result to the BILLWG. Further in-depth exploration of various data sets and different alternative model scenarios will be discussed at the 2016 BILLWG assessment meeting.

Introduction

The Billfish Working Group (BILLWG) of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) completed a benchmark stock assessment for blue marlin (*Makaira nigricans*) in the Pacific Ocean in 2013 (ISC, 2013). The 2013 assessment included data from 1971-2011, and showed a long term decline in blue marlin biomass. Spawning stock biomass (SSB) was 24,990 metric tons in 2011 or 129% of SSB at maximum sustainable yield (SSB_{MSY}), and fishing mortality (F) on the stock (average on ages 2 and older) was F = 0.26 during 2009-2011 or 81% of *F* at maximum sustainable yield (F_{MSY}). Overall trends in SSB and recruitment indicated a long-term decline in SSB and suggested a fluctuating pattern without trend for recruitment. Kobe plots indicated that the Pacific blue marlin SSB decreased to the MSY level in the mid-2000s, and since then has increased slightly. The base case assessment model indicated that the Pacific blue marlin stock was not overfished and was not subject to overfishing relative to MSY-based reference points. There is a three year cycle for assessments in the BILLWG, so an update assessment of the 2013 blue marlin benchmark was scheduled for 2016.

This stock assessment report describes the updated 2016 stock assessment for blue marlin in the Pacific Ocean. The most up to date catch, catch-per-unit-effort (CPUE), and composition data from 1971-2014 were provided by individual ISC countries, the Western and Central Pacific Fisheries Commission (WCPFC), the Inter-American Tropical Tuna Commission (IATTC), and the Secretariat of the Pacific Community (SPC). The 2016 assessment used the same modeling platform (Stock synthesis, SS) and version (3.24f) as were used in the 2013 assessment.

Furthermore, similar model structure and parameters were used in comparison to the 2013 assessment.

Materials and methods

Spatial and Temporal stratification

The geographic area encompassed in the assessment for blue marlin was the entire Pacific Ocean. Three types of data were used: fishery-specific catches, relative abundance indices, and length and size measurements. These data were compiled for 1971-2014, and catch, and length and size compositions were compiled by quarter. Available data, sources of data, and temporal coverage of the datasets used in the updated stock assessment were summarized in Figure 1. Details are presented below.

Definition of fisheries

As in the 2013 assessment, 16 fisheries were defined on the basis of country, gear type, location, and time period were considered relatively heterogeneous fishing units. These fisheries consisted of eight country-specific longline fisheries (JPNEarlyLL, JPNLateLL, JPNCLL, HWLL, ASLL, TWNLL, OthLL, and PYFLL), one driftnet fishery (JPNDRIFT), one bait fishery (JPNBait), two purse seine fisheries (EPOPS and WCPFCPS), two troll and handline fisheries (HWOth and EPOOth), and two mixed gear fisheries (JPNOth and TWNOth). Detailed descriptions of the defined fisheries were presented in Table 1.

Catch

Catch was input into the model seasonally (i.e., by calendar year and quarter) from 1971 to 2014 for the 16 individual fisheries. Catch was recorded and reported in numbers (thousands of fish) for the American Samoa longline fishery (ASLL) and the Eastern Pacific Ocean purse seine fishery (EPOPS), and in weight (metric tons) for all other fisheries. Because 2011 catch data were incomplete for the last assessment, updated catch data from 2011-2014 for all fisheries except JPNEarlyLL were used for the assessment update. In addition, updated time series of catch prior to 2011 were used for OthLL, WCPFCPS, JPNDrift, JPNOth, ASLL, JPNCLL, PYFLL, and the EPOPS fisheries.

Three countries (i.e., Japan, Taiwan, and the USA) provided updated national catch data (Ijima and Shiozaki, 2016; Nan-Jay Su personal communication, Jan 15, 2016; Ito 2016). Logbook catch data for 2014 from JPNCLL, JPNDRIFT, JPNBait, and JPNOth were incomplete, so yearbook catches from 2013 were used as 2014. Blue marlin catches for all other fishing countries were collected from WCPFC and IATTC category I & II data (Chang et al. 2016). Overall, use of the updated catch data led to a small increase of 1.6% in reported blue marlin catch prior to 2011 compared to the 2013 assessment. Individual differences in catch between this update and the 2013 assessment were shown in Appendix Figure A1.

The updated time-series of blue marlin catches by fishery from 1971-2014 was shown in Figure 2. The historical maximum and minimum annual blue marlin catches were 25,588 metric tons in 2003 and 9,160 metric tons in 1971, respectively. The JPNEarlyLL fishery took most of the blue marlin catch during the early assessment period, but declined after 1995 (as JPNLateLL). Since reaching a maximum in 2003, catches declined and with the exception of 2010, were similar between 2007-2011. After 2011, catches increased to current levels. The average catch of blue marlin in the Pacific Ocean was about 19,663 metric tons during the update period (2011-2014) with the TWNOth and OthLL fisheries taking 27% and 39%, respectively.

Abundance indices

Relative abundance indices of CPUE were available for this assessment and are shown in Figure 3 and Table 2. All indices were updated except for S1_JPNEarlyLL (1975-1993). Set-by-set logbook data was used in the CPUE standardization for Japanese distant water and offshore longline fisheries (Kai et al., 2016). A deterministic habitat-based standardization model (HBS; Hinton and Nakano 1996) using the same data filtering and assumptions as the CPUE standardization for the last 2013 assessment was used (Kanaiwa et al., 2013).

Operational data collected in the deep-set sector of the Hawaiian longline fishery by fishery observers in 1995-2014 were used for CPUE standardization of S3_HW_LL (Carvalho et al. 2016). Similar patterns of the standardized CPUE indices were produced by the delta-lognormal and zero-inflated negative binomial models. The same approach used in the last assessment (the zero-inflated negative binomial generalized linear model (GLM)) was used to develop the relative abundance index for S3_HW_LL.

Data aggregated by 5°×5° grids, with quarters, latitude, longitude, and year information from 1967 to 2014, and those with hooks per basket (HPB) information for 2000-2014 were standardized using GLM for the Taiwanese distant-water longline fishery (Su et al., 2016). The CPUE standardization models were conducted based on three periods, 1967-1978 (S4_TWNLL), 1979-1999 (S5_TWNLL) and 2000-2014 (S6_TWNLL), due to the changes in the fishery such as targeting. Given the timeframe of the model was limited to 1971-2014, the early years (1967-1970) of the CPUE time series for S4_TWNLL were removed.

Visual inspection of all indices grouped by fishery type showed a stable trend over time with the exception of an increasing trend of S1_JPNEarlyLL (1975-1984), a large decreasing trend of S3_HWLL, and a minor decreasing trend of S5_TWNLL (Figure 3). Updated CPUE indices on a relative scale were compared to the indices used in the 2013 assessment (Appendix Figure A2). In general, the updated CPUE indices showed a consistent trend to the previous CPUE indices, although the updated CPUE of S2_JPNLateLL and S3_HWLL showed higher variability. The updated S4_TWNLL and S6_TWNLL were less variable compared to the previous indices used in the 2013 assessment.

Correlations among CPUE indices were analyzed in the 2013 assessment. Similarly, correlations among the updated CPUE indices were also examined (Appendix Table A1). Pearson correlation coefficients (ρ) were interpreted as measuring the association among pairs of CPUE series.

Patterns in correlations among CPUE indices for the update assessment were similar to those in the last assessment. S1_JPNEarlyLL and S4_TWNLL (n=4) and

S1_JPNEarlyLL and S5_TWNLL (n=15) showed a consistent trend (ρ ranged from 0.11 to 0.38). S2_JPNLateLL and S3_HWLL (n=20, ρ = 0.24), S2_JPNLateLL and S5_TWNLL (n=6, ρ = 0.23), and S2_JPNLateLL and S6_TWNLL (n=15, ρ = 0.22) were also positively correlated. However, negative correlations were found between the S3_HWLL and S5_TWNLL (n=5, ρ = -0.14) and S3_HWLL and S6_TWNLL (n=15, ρ = -0.24). Based on the graphical inspection of relative CPUEs and the correlation analysis, the updated data supported the use of a similar base case (i.e., S1, S2, S4, S5, and S6 were fitted and contributed to the total likelihood) to the one for the 2013 assessment.

Composition data

Quarterly length or weight composition data from 1971–2014 for eight fisheries were used in the update assessment, and were summarized in Table 3. Updated length frequency data were available for six fisheries, and weight (size) frequency data for one. An updated time-series of length composition data for TWNLL was not available, so composition data from the last assessment were used. Since not all samples were known by sex, all compositions were assumed to be for a single gender.

As was done in the previous assessment, length frequency data were compiled using 5-cm length bins from 80 to 320 cm for JPNEarlyLL (F1), JPNLateLL (F2), HWLL (F7), TWNLL (F10), and EPOPS (F14), and using 10-cm bins from 80 to 320 cm for OthLL (F12) and PYFLL (F13). Weight frequency data for JPNDRIFT (F4) were compiled using varying binning structure from 10 to 300 kg according to the allometric length-weight relationship by using 10-cm bins from 80 to 320 cm. OthLL, PYFLL, and JPNDRIFT were inputted as generalized-size composition data in SS. 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.

There were some differences between the updated and previously used compositional data, as shown in Figure A3. The differences in mean length or size between the updated and the previous dataset were generally less than 5%, with the exception of smaller mean for JPNDRIFT in all years and for OthLL in 2011. Despite the differences, the new composition data were agreed upon at the BILLWG data workshop as the best available scientific information for the 2016 stock assessment.

Figure 4 shows the updated quarterly length and size compositions. Most of the fisheries exhibited consistent, clear seasonal cycles in their composition data. There were some variations in the distributions within a fishery; e.g., JPNLateLL in 2003, HWLL after 2000, EPOPS before 1992, and OthLL before 1997. The PYFLL size distributions also varied considerably between 1996-2002 and 2003-2014.

There was also considerable variation in both the length and size distributions and modal positions among fisheries (Figure 5). Length distributions for JPNEarlyLL, JPNLateLL, and HWLL were generally skewed to lengths less than 200 cm EFL and typically exhibited a single mode near 150 cm EFL. Length distributions for TWNLL, and size distributions for EPOPS, JPNDRIFT, and OthLL were less skewed. The TWNLL and OthLL exhibited a single mode near 160 cm EFL, and the JPNDRIFT had a mode around 100 kg. The EPOPS exhibited a single mode at around 200 cm EFL, meaning that this fleet caught larger blue marlin. Two modes were observed for PYFLL, one near 100 cm EFL and the second near 180 cm EFL.

Model Description

This stock assessment update for blue marlin was conducted using the same stock assessment model (SS, version 3.24f; Methot and Wetzel, 2013) as used previously. The model structure and parameters were similar to the base case run used in the 2013 stock assessment. Biological and demographic assumptions and fishery dynamics are summarized in Table 4 and Table 5, respectively.

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 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 approximated as 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.

The log(SE) of each candidate index was first estimated by the statistical model used to standardize the index in the various BILLWG working papers (Table 2). Input CPUE values and the reported log(SE) for all indices are shown in Tables 6 and 7, respectively.

The reported log(SE) for the abundance indices only capture observation error within the standardization model and do not reflect process error inherent between the unobserved vulnerable population and the observed abundance indices. Following the previous assessment, a minimum average log(SE) for indices of 0.14 was assumed for each series. Series with average log(SE) < 0.14 were scaled to log(SE) = 0.14 through the addition of a constant. Series with average log(SE) > 0.14 were input as given.

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.

To obtain random samples from the population, approximations of the amount of clustering were taken from an analysis of the relationship with number of trips sampled in the HWLL fleet which found around 10 fish per trip for marlin (Courtney, unpublished). Thus for all longline fisheries (F1, F2, F7, F10, F12, F13), sample size was assumed to be number of fish measured/10. For JPNDRIFT and EPOPS (F4, F14), sample size was assumed to be the number of fish measured. The minimum quarterly sample size was fixed at 2.5 (i.e. 25 samples/10) for all longline fisheries and was fixed at 25 for JPNDRIFT and EPOPS, so as to remove unrepresentative observations. Length or size composition records with effective sample sizes > 50 were set to 50 for all fisheries.

Data weighting

Index data were prioritized in the previous assessment. To maintain consistency with the previous assessment, index data were also 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). In the last assessment, samples sizes of the composition data were 50 for F1, F2, F4, F10, and F14 after following the procedures for stage 1 weighting described in the 'Data observation models' section. These samples therefore exhibited little within-fishery variability. In order to retain the relative among-sample variability when fitting the models, a single iteration of the model was made. The effective sample sizes estimated in this tuning fit were then re-scaled by a scalar (i.e., stage 2 weighting).

The value of the scalar used in the last assessment was not reproducible, and so for this update assessment, we used a similar stage-1 weighting scheme for the length or size composition data of fleets F1, F2, F4, F10, and F14, but a different stage-2 scalar. The process used to calculate the stage-2 scalar for fleets F1, F2, F4, F10, and F14 in this update assessment was to:

1) estimate the effective sample size for compositional data using a single iteration of SS3;
2) replace input sample size of each fleet with the estimated effective sample size relative to its mean, and re-scale to have a mean value of 30, which was based on the values in Table 5.3 from the last assessment (ISC, 2013); and
3) if the new input sample size > 50, set the sample size to 50.

Goodness-of-fit to abundance data

For each abundance index, the standard deviation of the normalized (or standardized) residuals (SDNR) was used to examine the goodness-of-fit (Francis, 2011). For an abundance data set to be fitted well, the SDNR should be less than $\left[\chi^2_{0.95,\,m-1}/(m-1)\right]^{0.5}$ where $\chi^2_{0.95,\,m-1}$ is the 95th percentile of a χ^2 distribution with m–1 degrees of freedom. Various residuals plots, including the observed and expected abundances, were also examined to assess goodness-of-fit.

Future Projection

As were done in the previous 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 simply based on the spawner-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' (2012-2014) selectivity patterns and relative fishing mortality rates were used in population future projection. The projection started in 2015 and continued through 2024 under four different harvest scenarios:

1. High F Scenario: Select the 3-year time period with the highest average F (age 2+) and apply this fishing mortality rate to the stock estimates beginning in 2015;
2. F_{MSY} Scenario: Apply the estimate of the F_{MSY} fishing mortality rate to the stock estimates beginning in 2015;
3. Status Quo F Scenario: This will be the average F (age 2+) during 2012-2014 ($F_{2012\text{-}2014}$);
4. Low F Scenario: Apply an $F_{30\%}$ fishing mortality rate to the stock estimates beginning in 2015.

Results

Base case model

Our exploration of the updated data supported the use of a similar base case to the one for the 2013 assessment. Although there were some variations in indices used in the update assessment compared to the 2013 assessment (i.e., S2_JPNLateLL), the correlation analyses supported the choice to utilize the same abundance indices in this update assessment (i.e., exclude S3_HWLL from the total likelihood; Table A1).

The proposed weighting method for the composition data produced similar input values and variation among year compared to the previous weighting method (Figure A4). The initial mean input sample sizes, mean estimated sample sizes, and re-scaled mean estimated sample sizes were shown in Table 8. The proposed weighting method produced relatively smaller sample sizes compared to the initial N and estimated N, thus down-weighting the composition data. The mean effective sample sizes for F1, F2, F4, F10, and F14 scaled down the initial N by factors between 0.55 and 0.6 (with mean sample sizes ranging from 24.6 and 29.27), with the greatest effect being on JPNEarlyLL and JPNDRIFT.

Recruitment variability (σ_R, the standard deviation of log-recruitment) was iteratively rescaled in the final model to match the expected variability and set to 0.28 based on the RMSE of the recruitment deviations. This followed the same approach as was used in the 2013 assessment, but resulted in a different value than what was used in the 2013 assessment, which was 0.32.

Model convergence

All estimated parameters in the base case model were within the set bounds, and the final gradient of the model was 4.15869e-005, which indicated that the model had converged onto a local or global minimum. Results from 30 model runs with different initial starting values for estimated parameters using the internal routine in SS suggested a global minimum was obtained (i.e., there was no evidence of a lack of convergence to a global minimum) (Figure 6). In addition, the log(R0) values were similar from runs with total negative log-likelihoods similar to the base case model.

Model diagnostics

Figure 7.1 presents the results of the likelihood profiling on log(R0) for each data component. Detailed information on changes in negative log-likelihoods among the various fisheries' data was shown in Figure 7.2 and Tables 9 and 10.

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). In general, the changes in negative log-likelihoods of abundance indices were small over the range of R0 (Figure 7.1).

S1_JPNEarlyLL (max 20.72) and S2_JPNLateLL (max 19.37) showed the largest change in likelihood across values of R0 among abundance indices (Table 9). Change in likelihood was also high for S3_HWLL, but S3_HWLL was not included in the model likelihood for the base case. The MLE estimate for log(R0) matched a local minimum between 6.5 and 7.0 in the fleet combined likelihood profile for index data. The likelihood profile of individual fleets was similar to the overall MLE estimate for S4_TWNLL and S5_TWNLL, was similar to the fleet combined likelihood pattern for S2_JPNLateLL and S6_TWNLL, and was different than the MLE estimate for S1_JPNEarlyLL (Figure 7.2). S1_JPNEarlyLL may provide conflicting information compared to other fleets' indices based on its lower fleet specific MLE estimate (Table 9).

In general, the changes in log-likelihoods among eight composition data were small over a range of log(R0) values except for the JPNEarlyLL and JPNLateLL (Table 10). The maximum changes in negative log-likelihoods for F1_JPNEarlyLL and F2_JPNLateLL are 75.21 and 41.80, respectively. Five of eight fleets had minimum relative negative log-likelihoods that occurred between 6.7-6.9.

This implies that length data (F1, F2, F7, F10, and F14) are informative in the fitting process. The MLE estimate also matched well with the likelihood profile of individual fleets except F7_HWLL (Figure 7.2). This implies F7_HWLL may provide conflicting information compared to other fleets' length composition. The MLE estimate did not match the fleet combined likelihood profile for generalized-size data very well. A similar pattern was found in the likelihood profile of individual fleet's generalized-size data, with the exception of F13_PYFLL. Generalized-size data for F12_OthLL and F4_JPNDRIFT may provide conflicting information compared to the length composition data from other fleets.

The magnitude of change in the log-likelihoods for the index were similar to length composition and generalized-size composition data within the log(R0) range of 6.6-7.0, and were within 5 units of likelihood at the MLE estimate of log(R0) (6.88; Figure 7.3). Minor conflicts in the shape of the likelihood profiles between index, length composition and generalized-size composition data were observed. The likelihood profile analysis suggested that the generalized-size composition data indicated a smaller log(R0) value than the index and length composition data, and therefore was possibly uninformative with respect to population scale in the base case assessment model. There was greater agreement between the length composition data and the abundance indices for the maximum likelihood estimate of log(R0) within the range of 6.6-7.0 based on log(R0) likelihood profiles, but less agreement with the generalized-size composition. In other words, the generalized-size composition data did not stop the model from fitting abundance data for the base case model.

Residual analysis of abundance Indices

Goodness-of-fit diagnostics were presented in Table 11, and plots of predicted and observed CPUE by fishery for the base case model were shown in Figure 8. As in the last stock assessment, the root-mean-square-error (RSME) was used as a goodness-of-fit diagnostic, with relatively low RMSE values (i.e., RMSE < 0.2) being indicative of a good fit. As in the 2013 assessment, the model fit all abundance indices that were incorporated into the total likelihood well, with RMSE < 0.2. Although not included in the likelihood of the fitted models, index HWLL (S3) was included in the model to allow comparison of the fitted and observed trends.

Although the input log(SE) of S4_TWNLL and S5_TWNLL in the update assessment (0.14 and 0.14) were smaller than the 2013 assessment (0.64 and 0.45), the input log(SE) were comparable with the RMSE of residuals for the base case. Similar uncertainty between input log(SE) and the RMSE of residuals were found in other indices in the base case model. This suggested that the input log(SE) were appropriate for observation error.

The fits to abundance indices were generally within the 95 percent CIs. The residuals pattern of the assessment update was similar to the 2013 assessment (Figure 8). There was a trend of negative residuals in the early time period (1975-1977) and of positive residuals in the late time period (1984-1993) in S1_JPNEarlyLL for both assessments (Figure 8).

In contrast to the 2013 assessment, the model fit the S5_TWNLL well. There was a trend of negative residuals in 1995-1999 for the 2013 assessment, but this was not observed in the update assessment. The RMSE of residuals also showed an improved fit, 0.12 for the update assessment compared to 0.21 for the 2013 assessment. The improved performance was most likely caused by a slightly lower variability in CPUE values later (1995−1999) in the S5_TWNLL time-series for the update assessment. Although not included in the total likelihood, and therefore not fit, showing diagnostics for HWLL (S3) revealed that HWLL was inconsistent with fits to other indices.

The SDNR of the CPUE fit was used as another goodness-of-fit diagnostic (Table 11). The SDNR diagnostics also indicated that the update model did not fit S2_JPNLateLL (1.28 > 1.25) well compared to the 2013 assessment (1.16 < 1.27). It should be noted the number of observations were different for S2_JPNLateLL between two assessments (18 and 21).

Residuals analysis of composition data

Comparisons between the observed and expected mean values of composition data from Francis (2011) were used for model diagnostics. Figure 9 shows the 95% credible intervals for mean value for the five length composition data sets and the three generalized-size composition data sets. The reweighted model fit passed through almost all of the credible intervals (Figure 9), although there was a poor fit between the observed and predicted mean values for the EPOPS in 1990, OthLL in 1993 and 2010, and PYFLL in 1997, 2002 and 2009. The results suggested that our stage-2 weighting approach accounted for correlations analogous to recommended methods from Francis (2011).

Model misfit of composition data was found in four fisheries, JPNEarlyLL (F1), JPNLateLL (F2), HWLL (F7), and PYFLL (F13) (Figure 10). Patterns of positive residuals occurred around 100 cm EFL during 1971-1977 and above 200 cm EFL during 1971-1979 for JPNEarlyLL, around 150 cm EFL during 1994-2014 for JPNLateLL, and below 160 cm EFL during 2000-2006 and above 200 cm EFL during 2002-2014 for HWLL. Negative residuals occurred around 135 cm EFL during 1971-1982 and 1984-1993 for JPNEarlyLL, around 130 and 170 cm EFL during 1994-2014 for JPNLateLL, and below 150 cm EFL during 2007-2014 for HWLL. Outliers (extreme positive residuals) were found in 1997, 2002 and 2005 for PYFLL.

Assuming standardized residuals were normally distributed, 95% of the measurements would fall within 2 standard deviations of the mean. JPNLateLL, HWLL, EPOPS, OthLL, PYFLL, and JPNDRIFT were found with 0.1%, 0.3%, 0.1%, 0.8%, 2.2%, and 0.3% of their Pearson residuals greater than 2 or smaller than -2, indicating appropriate distributional assumptions (Figure 10). Nonetheless, the observations with extreme standardized residuals might need further investigation**.**

The model fit the length modes in composition data aggregated by fishery fairly well using the input effective sample sizes (Figure 11). The precision of the model predictions was greater than that of the observations, and indirectly related to effective sample size. Estimated effective sample size was used for the goodness-of-fit diagnostics for the composition data in the 2013 assessment. In this updated stock assessment, the effective sample sizes as derived from our stage-2 weighting process were slightly smaller than the input effective sample sizes used in the 2013 assessment (Table 12).

Estimation of selectivity

The same selectivity configurations were used in this update stock assessment as were used for the 2013 assessment. The results of the estimated selectivity patterns were consistent with the assumed selectivity patterns (Figure 12). There was a significant change for JPNDRIFT with higher selectivity for the smaller fish and lower selectivity for the larger fish (i.e., the selectivity curve shifted left). There was also a minor change in selectivity during the second time block for PYFLL and the selectivity for EPOPS. There was lower selectivity for fish around 120-170 cm EFL for PYFLL in 2003-2014 and higher selectivity for fish greater than 250 cm EFL for EPOPS.

Stock assessment results

Estimates of population biomass (age 1 and older at the beginning of the year) declined from a high of 135,623 in 1971 until 1977, increased to 124,812 metric tons in 1984, decreased again to the lowest level of 69,720 metric tons in 2009, and increased to around 78,000 metric tons during the final three years (2012−2014) (Figure 13a). Compared to the 2013 stock assessment, the population biomass estimates were higher in 1971-1990, slightly lower in 1991-1993, 1997-1998, and 2010-2011.

SSB also exhibited a decline during 1971−1979, was stable during 1980-1986, declined to the lowest level of 20,972 metric tons in 2006, and increased to 24,809 in 2014 (Figure 13b). The time-series of SSB at the beginning of the spawning cycle (season 2) averaged 62,368 metric tons during 1971-1979, or 50% of unfished SSB; 50,577 metric tons (34% of unfished SSB) during 1980−1989; 39,715 metric tons (28% of unfished SSB) during 1990−1999; 25,272 metric tons (19% of unfished SSB) during 2000−2009, and 23,717 metric tons (21% of unfished SSB) in 2010−2014. Compared to the 2013 stock assessment, the SSB estimates were higher in 1971−1991. Precision of SSB estimates gradually improved over time.

Recruitment (age-0 fish) estimates indicated a long-term fluctuation around a mean of approximately 897,000 (Figure 13c). Recruitment was low in the early part of time series (1971-1976) with an average of 741,000 recruits. The model estimated that several strong year classes (> 1000 thousand recruits) recruited to the fisheries in 1977-1979, 1982-1983, 1986-1987, 1992, 1997, 2009, and 2011 followed by several weak year classes. Compared to the 2013 stock assessment, the recruitment estimates were higher in 1977-1978, 1992, 1997 and 2011, but lower in 2009. Uncertainty in recruitment estimates in the update assessment was smaller than the 2013 assessment during 1985-1997, and comparable in other years.

Over the course of the modelled time series, estimated fishing mortality (average on ages 2 and older) gradually increased from the early 1970s to the 1990s, peaked at 0.38 year^{-1} in 2005 in response to higher catches, and afterward declined to 0.28 year^{-1} in the most recent years (2012-2014) (Figure 13d). Compared to the 2013 stock assessment, fishing mortality estimates were slightly higher in 2005 and 2010-2014, but overall were very similar.

Compared to MSY-based reference points, the current spawning biomass (average for 2012-2014) was 23% above SSB_{MSY} and the current fishing mortality (average for ages 2 and older in 2012-2014) was 14% below F_{MSY}. The Kobe plots indicate that the Pacific blue marlin spawning stock biomass decreased to the MSY level in the mid-2000's, and since then has increased slightly (Figure 14). The base case assessment model indicates that the Pacific blue marlin stock is currently not overfished and is not subject to overfishing relative to MSY-based reference points.

Future projection

The projected trajectories of SSB and yield from 2015 to 2024 were shown in Tables 13 and 14 and Figure 15. When the current fishing level was maintained (Scenario 3: $F_{2012-2014}$, equivalent to $F_{21\%}$), the SSB was projected to be stable at roughly 24,800 metric tons by 2024, which was above SSB at MSY level (19,852 metric tons). If fishing increased to the MSY level (Scenario 2: equivalent to $F_{18\%}$), the projected SSB was estimated to gradually decrease, and by 2024 it approached but remained above the SSB at MSY level. If fishing further increased to the 2003-2005 level (Scenario 1: $F_{16\%}$), the SSB was projected to be below SSB at MSY level by 2019. Conversely, if fishing mortality was reduced to be equivalent to $F_{30\%}$ (Scenario 4), the projected SSB would gradually increase to about 35,400 metric tons by 2024.

Fishing at the current level ($F_{21\%}$) and F_{MSY} ($F_{18\%}$) provided an expected safe/optimal level of harvest, where the average projected catches between 2015 and 2024 were near MSY at approximately 20,200 and 19,800 metric tons. Fishing at the 2003-2005 level ($F_{16\%}$) and $F_{30\%}$ provided average projected catches between 2015 and 2024 of about 21,900 and 17,000 metric tons, respectively.

Discussion

Data inconsistencies with previous assessment

We found there were some inconsistencies with the historical catch and length composition data between the 2013 assessment and the newly available data set (Figures A1 and A3). Because 2011 catch data were incomplete for the last assessment, updated catch data from 2011-2014 for all fisheries except JPNEarlyLL were used for the assessment update. In addition, updated time series of catch prior to 2011 were used for OthLL, WCPFCPS, JPNDrift, JPNOth, ASLL, JPNCLL, PYFLL, and the EPOPS fisheries. Overall, use of the updated catch data led to a small increase of 1.6% in reported blue marlin catch prior to 2011 compared to the 2013 assessment. Although the impact of updated time series of catch prior to 2011 on the historical population dynamics was not evaluated in this assessment, it was expected to be negligible.

The input log(SE) of S4_TWNLL and S5_TWNLL indices were much smaller than the 2013 assessment. Although the input log(SE) values were comparable with the RMSEs of residuals for the base case, it was unclear how sensitive the assessment result was to the lower input log(SE) value. The influence of the log(SE) value could be ascertained through a sensitivity run.

We used updated composition data for the whole time-series, with the exception of TWNLL. The differences in mean length or size between the updated and the previous dataset were generally less than 5% except smaller mean sizes (in kg) were observed in JPNDRIFT for all years and OthLL (in cm) in 2011 in the newer dataset. The change of JPNDRIFT size composition was caused by the misallocation of the size data into bins for the 2013 assessment, which affected the selectivity estimation of the fleet.

Alternative CPUE indices

The S3_HWLL was excluded from the total likelihood in the base case model. HWLL was positively correlated with JPNLateLL, but negatively correlated with S5_TWNLL and S6_TWNLL. Moreover, JPNLateLL was positively correlated with S5_TWNLL and S6_TWNLL, respectively. Contradictory correlation was found among the abundance indices. Different alternative model scenarios with subsets of CPUE indices may need to be explored.

Data weighting

The results of fishery stock assessments based on integrated model can be sensitive to the values used to weight each of the data types included in the objective function. Punt (2016) provided a comprehensive review and a comparison of various iterative re-weighting methods for length composition data. He found the McAllister-Ianelli-2 (1.B in his Table 2) generally performed better than other methods, including the McAllister-Ianelli-1 (1.A) that is often used in SS assessments, and the Francis method (1.C). The iterative re-weighting approach attempts to reduce the potential for particular data sources to have a disproportionate effect on total model fit, while creating estimates of uncertainty that are commensurate with the uncertainty inherent in the input data.

Our analysis revealed that the harmonic mean of the ratio between estimated effective N and input N was always greater than one for all fleets given the nature of blue marlin composition data (i.e., small variance of the residuals of the fit of the model to the data). For the iterative re-weighting approach, this implied up-weighting the length data, something not explored in the analyses by Punt (2016).

Noting that intra-set correlation commonly occurs in length composition data, necessitating down!weighting, the iterative re-weighting approach was not used. We recommend further examination of the precision of fleet-specific estimates of the length-frequency distributions of blue marlin in the Pacific Ocean. Future improvement for the sampling design of the fishery composition data as well as more appropriate fleet definition is also recommended.

We down weighted the composition data to give priority to the abundance indices because they provide direct information about the stock assessment quantities that are of most interest. However, the blue marlin abundance indices may not provide good contrast for estimating the stock assessment quantities. Therefore, the results from the assessment may be uncertain or difficult to interpret, as would be expected given that the generally increasing catch and slightly declining or flat trend of CPUEs reflect a one-way-trip (Hilborn and Walters, 1992).

Other considerations

Average fishing mortality was calculated over ages 2 through age 26 for this assessment, but the Kobe plots from the previous assessment were based on fishing mortality for ages 1-10 (Figures 6.1 and 6.2 in ISC (2013) even though the caption for Figure 6.1 says ages 2+). This difference affected the scale of fishing mortality but did not change the qualitative determination of stock status.

Sensitivity analyses

The current working paper only provided the basic updated base case model and its result to the BILLWG and not results from sensitivity analyses. Sensitivity analyses to examine the effects of changing the assumed values of the input parameters or alternative model configurations for the Pacific blue marlin will be conducted at the 2016 BILLWG assessment meeting.

Acknowledgments

We sincerely thank Darryl Tagami, William Walsh, Felipe Carvalho, Russell Ito, Hiroaki Okamoto, Michael Hinton, Nan-Jay Su, Chi-Lu Sun, and Jon Brodziak of ISC Billfish Working Group for their help in preparing and providing information for this assessment update.

References

Brodziak, J. (2013) Combining information on length-weight relationships for Pacific blue marlin. Working paper submitted to the ISC Billfish Working Group Workshop. 16-23 January, Honolulu, Hawaii, 9pp. Available at http://isc.fra.go.jp/pdf/BILL/ISC13_BILL_1/ISC13BILLWG-1-01.pdf

Brodziak, J. and Mangel, M. (2011) Probable values of stock-recruitment steepness for north Pacific striped marlin. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/11.

Brodziak, J., Mangel, M., Sun, C.L. (2015) Stock-recruitment resilience of North Pacific striped marlin based on reproductive ecology. Fisheries Research, 166: 140-150.

Carvalho, F., Walsh, W., Chang, Y.-J. (2016) Standardized catch rates of blue marlin (*Makaira nigricans*) in the Hawaii-based longline fishery (1995-2014). Working paper presented to the ISC BILLWG data workshop, January 13-20, Honolulu, Hawaii. ISC/16/BILLWG-1/05.

Chang, Y.-J., Brodziak, J., Lee, H.-H., DiNardo, G., and Sun, C.-L. (2013) A Bayesian hierarchical meta-analysis of blue marlin (*Makaira nigricans*) growth in the Pacific Ocean. Working paper submitted to the ISC Billfish Working Group Workshop. 16-23 January, Honolulu, Hawaii, 23pp. Available at: http://isc.fra.go.jp/pdf/BILL/ISC13_BILL_1/ISC13BILLWG-1-02.pdf

Chang, Y.J., Yau, A., Brodziak, J. (2016) Summary of blue marlin catch and size data from the Western and Central Pacific Fisheries Commission and the Inter-

American Tropical Tuna Commission. Working paper presented to the ISC BILLWG data workshop, January 13-20, Honolulu, Hawaii. ISC/16/BILLWG-1/04.

Hinton, M. and Nakano, H. (1996) Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and biological data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. Bull. Int. Am. Trop. Tuna Comm., 21(4):171-200.

Hilborn, R. and Walters, C.J. (1992) Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. New York, Chapman & hall. 592 p.

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

Ijima, H. and Shiozaki, K. (2016) Japanese catch statistics of the Pacific blue marlin (*Makaira nigricans*): Update for a stock assessment. Working paper presented to the ISC BILLWG data workshop, January 13-20, Honolulu, Hawaii. ISC/16/BILLWG-1/02.

ISC (2013) Stock assessment of blue marlin in the Pacific Ocean in 2013, Report of the Billfish Working Group. July, Busan, Korea. Annex 10. Available at: http://isc.fra.go.jp/pdf/ISC13pdf/Annex%2010-%20Blue%20marlin%20stock%20assessment.pdf.

Ito, R.Y. (2016) U.S. commercial fisheries for marlins in the North Pacific Ocean. Working paper submitted to the ISC Billfish Working Group Meeting, 13-20 January, Honolulu, Hawaii. ISC/16/BILLWG-1/01.

Kai, M.,Okamoto, H., Shiozaki, K., Hinton, M. (2016). Update of Japanese longline CPUE for blue marlin (*Makaira nigricans*) in the Pacific Ocean standardized applying habitat model. Working paper presented to the ISC BILLWG data workshop, January 13-20, Honolulu, Hawaii. ISC/16/BILLWG-1/06.

Kanaiwa, M., Kimoto, A., Yokawa, K., and Hinton, M.G. 2013. Standardized abundance indices for blue marlin (*Makaira nigricans*) in Pacific Ocean from analyses of catch and fishing effort from offshore and distance water longline vessels of Japan. Working paper submitted to the ISC Billfish Working Group Workshop, 16-23 January, Honolulu, Hawaii, 87 pp. Available at: http://isc.fra.go.jp/pdf/BILL/ISC13_BILL_1/ISC13BILLWG-1-05.pdf

Langseth, B.J. and Fletcher, E. (2016) Size composition for blue marlin (*Makaira nigricans*) in the Hawaii-based pelagic longline fishery, 1994-2014. Working paper submitted to the ISC Billfish Working Group Meeting, 13-20 January, Honolulu, Hawaii. ISC/16/BILLWG-1/09.

Lee, H.-H. and Chang, Y.-J. (2013) Age-structure natural mortality for Pacific blue marlin based on meta-analysis and an ad hoc mortality model. Working paper submitted to the ISC Billfish Working Group Meeting, 16-23 January, Honolulu, Hawaii. ISC/13/BILLWG-1/07

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

McAllister, M.K. and Ianelli, J.N. (1997) Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences, 54:1094-1151

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.

Punt, A.E. (in press) Some insights into data weighting in integrated size-structured models. Fisheries Research.

Shimose, T., Fujita, M., Yokawa, K., Saito, H., and Tachihara, K. (2009) Reproductive biology of blue marlin *Makaira nigricans* around Yonaguni Island, southwestern Japan. Fish. Sci., 75:109-119.

Su, N.J., Sun, C.L., and Yeh, S.Z. (2016) CPUE standardization of blue marlin (*Makaira nigricans*) for the Taiwanese distant-water tuna longline fishery in the Pacific Ocean. Working paper submitted to the ISC Billfish Working Group Meeting, 13-20 January, Honolulu, Hawaii. ISC/16/BILLWG-1/10

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. Fish. Bull. 107:420-432.

Tagami, D.T., and Wang, H. (2016) Updated blue marlin catches (2012-2014) in the North and South Pacific from WCPFC data. Working paper submitted to the ISC Billfish Working Group Meeting, 13-20 January, Honolulu, Hawaii. ISC/16/BILLWG-1/03.

Table 1. Descriptions of fisheries included in the base case model for the stock assessment update including fishing countries, gear types, catch units (biomass (B) or numbers (#)), and reference sources for catch data.

Fishery number	Reference Code	Fishing Countries	Gear Types	Units	Source
F1	JPNEarlyLL	Japan	Offshore and distant-water longline (early period)	B	Ijima and Shiozaki (2016)
F2	JPNLateLL	Japan	Offshore and distant-water longline (late period)	B	Ijima and Shiozaki (2016)
F3	JPNCLL	Japan	Coastal longline	B	Ijima and Shiozaki (2016)
F4	JPNDRIFT	Japan	High-sea large-mesh driftnet and coastal driftnet	B	Ijima and Shiozaki (2016)
F5	JPNBait	Japan	Bait fishing	B	Ijima and Shiozaki (2016)
F6	JPNOth	Japan	Other gears	B	Ijima and Shiozaki (2016)
F7	HWLL	USA (Hawaii)	longline	B	Ito (2016)
F8	ASLL	USA (American Samoa)	longline	#	Russell Ito, pers. comm., Jan 13, 2016
F9	HWOth	USA (Hawaii)	Troll and handline	B	Ito (2016)
F10	TWNLL	Taiwan	Distant-water longline	B	Nan-Jay Su, pers. comm., Jan 13, 2016
F11	TWNOth	Taiwan	Offshore longline, coastal longline, gillnet, harpoon, and others	B	Nan-Jay Su, pers. comm., Jan 13, 2016
F12	OthLL	Various flags	Longline	B	Chang et al. (2016); Tagami and Wang (2016)
F13	PYFLL	French Polynesia	Longline	B	Chang et al. (2016)
F14	EPOPS	Various flags	Purse seine	#	Chang et al. (2016)
F15	WCPFCPS	Various flags	Purse seine	B	Chang et al. (2016)
F16	EPOOth	French Polynesia	Troll, handline, and harpoon	B	Chang et al. (2016)

Table 2. Descriptions of standardized relative abundance indices (catch-per-unit-effort, CPUE) of Pacific blue marlin used in the stock assessment update including whether the index was used in the base case, sample size (*n*), years of coverage, and reference source. For all indices, catch was in numbers and effort was in 1000 hooks.

Reference Code	**Used**	**Fishery Description**	***n***	**Time series**	**Source**
S1_JPNEarlyLL (F1)	Yes	Japanese offshore and distant-water longline (early period)	19	1975-1993	Kai et al. (2016)
S2_JPNLateLL (F2)	Yes	Japanese offshore and distant-water longline (late period)	21	1994-2014	Kai et al. (2016)
S3_HWLL (F7)	No	Hawaiian longline	20	1995-2014	Carvalho et al. (2016)
S4_TWNLL (F10)	Yes	Taiwanese distant-water longline (early period)	8	1971-1978	Su et al. (2016)
S5_TWNLL (F10)	Yes	Taiwanese distant-water longline (middle period)	21	1979-1999	Su et al. (2016)
S6_TWNLL (F10)	Yes	Taiwanese distant-water longline (late period)	15	2000-2014	Su et al. (2016)

Table 3. Description of length composition data (eye-fork lengths, EFL, cm) and size composition data (kg) for Pacific blue marlin used in the stock assessment update, including bin size definitions, number of observations (n), years of coverage, and reference sources.

Reference Code	**Fleet**	**Fishery Description**	**Unit**	**Bin**	***n***	**Time series**	**Source**
JPNEarlyLL	F1	Japanese offshore and distant-water longline (early period)	cm	5	92	1971-1993	Ijima and Shiozaki (2016)
JPNLateLL	F2	Japanese offshore and distant-water longline (late period)	cm	5	84	1994-2014	Ijima and Shiozaki (2016)
JPNDRIFT	F4	High-sea large-mesh driftnet and coastal driftnet	kg	Proportional to length	19	1977-1989; 1993; 1998	Ijima and Shiozaki (2016)
HWLL	F7	Hawaiian longline	cm	5	70	1994-2014	Langseth (2016)
TWNLL	F10	Taiwanese distant-water longline	cm	5	23	2005-2010	ISC (2013)
OthLL	F12	Various flags longline	cm	10	83	1992-2014	Chang et al. (2016)
PYFLL	F13	French Polynesia longline	cm	10	52	1996-2014	Chang et al. (2016)
EPOPS	F14	Various flags purse seine	cm	5	95	1990-2014	Chang et al. (2016)

Table 4. Key life history parameters and model structures for Pacific blue marlin used in the stock assessment update including values, pertinent comments, and references.

Parameter	Value		Comments	Source
Gender	2		Two genders model	ISC(2013)
Natural mortality	Female: 0.42 (age 0) 0.37 (age 1) 0.32 (age 2) 0.27 (age 3) 0.22 (age 4-25)	Male: 0.42 (age 0) 0.37 (age 1-25)	Age-specific natural mortality	Lee and Chang (2013)
Reference age (a1)	1		Fixed parameter	Refit from Chang et al. (2013); ISC(2013)
Maximum age (a2)	26		Fixed parameter	
Length at a1 (L1)	144 (Female); 144 (Male)		Fixed parameter	Refit from Chang et al. (2013); ISC(2013)
Length at a2 (L2)	304.18 (Female) 226 (Male)		Fixed parameter	Refit from Chang et al. (2013); ISC(2013)
Growth rate (K)	0.107 (Female) 0.211 (Male)		Fixed parameter	Refit from Chang et al. (2013); ISC(2013)
CV of L1 (CV=f(LAA))	0.14 (Female); 0.14 (Male);		Fixed parameter	Chang et al. (2013); ISC(2013)
CV of L2	0.15 (Female); 0.1 (Male);		Fixed parameter	Chang et al. (2013); ISC(2013)
Weight-at-length	$W=1.844 \times 10^{-5}L^{2.956}$ (Female); $W=1.37 \times 10^{-5}L^{2.975}$ (male)		Fixed parameter	Brodziak 2013
Length-at-50% Maturity	179.76		Fixed parameter	Sun et al. (2009); Shimose et al. (2009)

Slope of maturity ogive	-0.2039	Fixed parameter	Sun et al. (2009); Shimose et al. (2009)
Fecundity	Proportional to spawning biomass	Fixed parameter	Sun et al. (2009)
Spawning season	2	Model structure	Sun et al. (2009)
Spawner-recruit relationship	Beverton-Holt	Model structure	Brodziak and Mangel (2011); Brodziak et al. (2015)
Spawner-recruit steepness (*h*)	0.87	Fixed parameter	Brodziak and Mangel (2011); Brodziak et al. (2015)
Recruitment variability (σ_R)	0.28	Fixed parameter	
Initial age structure	5 yrs (1966-1970)	Estimated	
Main recruitment deviations	1971-2013	Estimated	
Bias adjustment	1971-2013	Fixed	ISC (2013)

Table 5. 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.

Fishery number	Reference Code	Selectivity assumption	Mirror gear
F1	JPNEarlyLL	Cubic Spline (nodes=4)	
F2	JPNLateLL	Double-normal	
F3	JPNCLL	Double-normal	F2
F4	JPNDRIFT	Double-normal	
F5	JPNBait	Double-normal	F4
F6	JPNOth	Double-normal	F2
F7	HWLL	Cubic Spline (nodes=3)	
F8	ASLL	Double-normal	F7
F9	HWOth	Double-normal	F7
F10	TWNLL	Double-normal	
F11	TWNOth	Double-normal	F10
F12	OthLL	Double-normal	
F13	PYFLL	Double-normal for 1971-2002; 2003-2014	
F14	EPOPS	Double-normal	
F15	WCPFCPS	Double-normal	F14
F16	EPOOth	Double-normal	F14

Table 6. Standardized catch-per-unit-effort (CPUE; in number per 1000 hooks) indices for blue marlin from the Pacific Ocean used in the stock assessment update. Season refers to the calendar quarter(s) in which most of the catch was taken by each fishery, where 1 = Jan-Mar, 2 = Apr-June, 3 = July-Sept, and 4 = Oct-Dec.

Index	**JPNEarlyLL**	**JPNLateLL**	**HWLL**	**TWNLL**		
	S1	**S2**	**S3**	**S4**	**S5**	**S6**
Season	1	1	3	1	1	1
1971				0.076		
1972				0.080		
1973				0.082		
1974				0.079		
1975	0.333			0.073		
1976	0.329			0.081		
1977	0.247			0.070		
1978	0.399			0.074		
1979	0.456				0.153	
1980	0.468				0.129	
1981	0.548				0.136	
1982	0.546				0.124	
1983	0.439				0.118	
1984	0.697				0.127	
1985	0.476				0.138	
1986	0.492				0.115	
1987	0.482				0.103	
1988	0.459				0.118	
1989	0.476				0.113	
1990	0.463				0.102	
1991	0.443				0.123	
1992	0.454				0.084	
1993	0.567				0.103	
1994		12.455			0.127	
1995		15.023	0.510		0.106	
1996		8.237	0.570		0.103	
1997		11.338	0.480		0.081	
1998		10.845	0.470		0.088	
1999		8.800	0.140		0.102	
2000		9.100	0.450			0.092

2001	7.611	0.300	0.099
2002	8.282	0.140	0.089
2003	10.174	0.230	0.108
2004	12.472	0.170	0.094
2005	10.816	0.120	0.127
2006	10.682	0.230	0.114
2007	8.864	0.050	0.111
2008	7.998	0.120	0.095
2009	11.265	0.110	0.095
2010	10.350	0.070	0.101
2011	7.487	0.100	0.094
2012	11.400	0.160	0.094
2013	9.457	0.070	0.111
2014	10.828	0.110	0.105

Table 7. Input standard error (SE) in log-space (i.e., log(SE)) of lognormal error for the catch-per-unit-effort (CPUE) series for Pacific blue marlin used in the stock assessment update. Lognormal errors were assumed. Season refers to the calendar quarter(s) in which most of the catch was taken by each fishery, where 1 = Jan-Mar, 2 = Apr-June, 3 = July-Sept, and 4 = Oct-Dec.

Index	**JPNEarlyLL**	**JPNLateLL**	**HWLL**	**TWNLL**		
	S1	**S2**	**S3**	**S4**	**S5**	**S6**
Season	1	1	3	1	1	1
1971				0.063		
1972				0.064		
1973				0.063		
1974				0.059		
1975	0.015			0.069		
1976	0.019			0.068		
1977	0.015			0.065		
1978	0.023			0.070		
1979	0.027				0.065	
1980	0.027				0.066	
1981	0.032				0.064	
1982	0.032				0.067	
1983	0.026				0.073	
1984	0.041				0.071	
1985	0.028				0.077	
1986	0.029				0.079	
1987	0.028				0.071	
1988	0.027				0.077	
1989	0.028				0.077	
1990	0.027				0.091	
1991	0.026				0.082	
1992	0.027				0.079	
1993	0.033				0.069	
1994		0.011			0.072	
1995		0.013	0.464		0.085	
1996		0.014	0.394		0.072	
1997		0.014	0.349		0.075	
1998		0.013	0.275		0.078	
1999		0.013	0.159		0.068	
2000		0.012	0.256			0.051

2001	0.011	0.179	0.042
2002	0.012	0.129	0.041
2003	0.014	0.149	0.040
2004	0.012	0.129	0.040
2005	0.015	0.129	0.040
2006	0.017	0.129	0.041
2007	0.013	0.070	0.044
2008	0.017	0.100	0.045
2009	0.022	0.100	0.045
2010	0.013	0.080	0.044
2011	0.016	0.090	0.044
2012	0.013	0.110	0.046
2013	0.016	0.100	0.050
2014	0.018	0.110	0.078

Table 8. Fishery-specific initial multinomial effective sample sizes (N) and re-scaled effective sample sizes for length composition data of Pacific blue marlin as used in the stock assessment update. Estimated mean N was the effective sample size from the initial run of SS.

Reference Code	Fleet	Initial mean N	Estimated mean N	Re-scale mean N
JPNEarlyLL	F1	49.65	269.25	27.11
JPNLateLL	F2	44.97	114.21	26.98
JPNDRIFT	F4	45.11	107.03	24.60
HWLL	F7	13.19	57.61	No rescale
TWNLL	F10	48.89	423.39	29.27
OthLL	F12	27.25	85.90	No rescale
PYFLL	F13	6.91	22.74	No rescale
EPOPS	F14	49.32	213.36	27.58

Table 9. Relative negative log-likelihoods of abundance index data components in the base case model over a range of fixed levels of virgin recruitment in log-scale (log(R0)). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of log(R0) was 6.88. See Table 2 for a description of the abundance indices. S3_HWLL was not included in the total likelihood.

log(R0)	S1_JPNEarlyLL	S2_JPNLateLL	S3_HWLL	S4_TWNLL	S5_TWNLL	S6_TWNLL
6	20.72	19.37	0.00	1.80	13.67	7.71
6.1	8.01	14.91	20.47	0.97	6.70	5.90
6.2	1.56	14.27	19.57	0.96	4.58	5.46
6.3	7.51	13.76	17.97	0.78	3.72	5.27
6.4	0.00	11.68	16.79	1.01	2.92	4.46
6.5	2.88	9.77	13.24	1.19	2.61	3.80
6.6	1.04	7.67	8.85	1.40	1.40	3.06
6.7	1.69	4.72	3.67	0.68	1.34	1.74
6.8	3.75	2.55	4.14	0.04	0.91	0.78
6.9	5.50	1.62	9.43	0.00	0.25	0.50
7	6.60	1.20	15.62	0.11	0.00	0.50
7.1	6.28	0.73	19.60	0.24	0.22	0.33
7.2	5.97	0.41	22.10	0.32	0.35	0.14
7.3	6.14	0.29	27.26	0.39	0.39	0.27
7.4	5.38	0.07	27.67	0.24	0.59	0.04
7.5	5.05	0.00	29.78	0.16	0.71	0.00

Table 10. Relative negative log-likelihoods of length composition data components in the base case model over a range of fixed levels of virgin recruitment in log-scale (log(R0)). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of log(R0) was 6.88. See Table 3 for a description of the composition data.

log(R0)	JPNEarlyLL	JPNLateLL	HWLL	TWNLL	EPOPS	JPNDRIFT	OthLL	PYFLL
6	75.21	41.70	2.76	7.96	11.95	1.45	8.64	3.75
6.1	40.09	29.17	11.64	6.29	6.91	1.87	6.10	2.86
6.2	42.17	23.28	6.62	4.12	5.45	2.43	2.40	2.33
6.3	37.92	20.38	4.15	2.82	5.23	2.49	0.46	2.12
6.4	25.44	14.93	3.15	1.32	3.39	1.69	0.00	1.39
6.5	11.57	11.25	3.17	0.77	2.41	0.81	0.41	0.91
6.6	4.28	7.75	3.50	0.44	1.55	0.65	1.19	0.53
6.7	0.59	3.23	3.73	0.00	0.54	0.28	2.22	0.17
6.8	0.00	0.25	3.33	0.15	0.00	0.17	3.46	0.00
6.9	1.14	0.00	1.54	0.52	0.42	0.12	4.86	0.00
7	1.73	3.60	0.43	0.85	1.10	0.07	6.31	0.77
7.1	1.58	4.35	0.00	1.20	2.11	0.05	8.07	2.40
7.2	1.53	6.97	0.13	1.60	2.70	0.02	10.15	1.08
7.3	1.70	9.06	0.58	1.84	3.32	0.00	12.00	5.83
7.4	2.19	11.49	0.96	2.16	3.72	0.01	13.93	1.69
7.5	2.64	13.46	1.39	2.38	4.11	0.01	15.58	1.95

Table 11. Mean input standard error (SE) in log-space (i.e., log(SE)) of lognormal error, root-mean-square-errors (RMSE), and standard deviations of the normalized residuals (SDNR) for the relative abundance indices for Pacific blue marlin used in the 2013 stock assessment and in this stock assessment update. S3_HWLL was not included in the total likelihood. An SDNR value greater than the chi-squared statistic (χ^2) indicates a statistically poor fit.

Reference code	2013 assessment					2016 update				
	N	**Input log(SE)**	**RMSE**	**SDNR**	χ^2	***n***	**Input log(SE)**	**RMSE**	**SDNR**	χ^2
S1_JPNEarlyLL (F1)	19	0.14	0.14	1.05	1.27	19	0.14	0.14	1.07	1.27
S2_JPNLateLL (F2)	18	0.14	0.16	1.16	1.27	21	0.14	0.17	1.28	1.25
S3_HWLL (F7)	17	0.14	0.48	3.39	1.28	20	0.18	0.83	4.36	1.26
S4_TWNLL (F10)	8	0.64	0.09	0.18	1.42	8	0.14	0.06	0.45	1.42
S5_TWNLL (F10)	21	0.45	0.21	0.39	1.25	21	0.14	0.12	0.89	1.25
S6_TWNLL (F10)	12	0.14	0.17	1.29	1.34	15	0.14	0.11	0.86	1.30

Table 12. Mean input multinomial effective sample sizes (N) and model estimated effective sample sizes (effN) in the 2013 stock assessment and this stock assessment update.

Reference code	**Fleet**	**2013 assessment**		**2016 update**	
		Input mean N	**Mean effN**	**Input mean N**	**Mean effN**
JPNEarlyLL	1	30.00	249.59	27.11	261.22
JPNLateLL	2	30.00	122.38	26.98	112.96
JPNDRIFT	4	30.00	121.68	24.60	116.58
HWLL	7	14.50	61.35	13.19	58.36
TWNLL	10	30.00	408.63	29.27	407.60
OthLL	12	26.49	85.14	27.25	86.09
PYFLL	13	6.95	19.38	6.91	22.44
EPOPS	14	30.00	209.53	27.58	210.63

Table 13. Projected trajectory of spawning stock biomass (SSB in metric tons) for alternative harvest scenarios. Fishing intensity ($F_{x\%}$) alternatives are based on $F_{16\%}$ (average 2003-2005), F_{MSY} ($F_{18\%}$), $F_{2012-2014}$ ($F_{21\%}$) (average 2012-2014 defined as current), and $F_{30\%}$. Green blocks indicate the projected SSB is greater than MSY level (SSB_{MSY} =19,853 metric tons).

Run	Harvest scenario	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Average
1	$F_{2003-2005}$ ($F_{16\%}$)	24,545	22,683	21,163	20,014	19,167	18,546	18,086	17,741	17,481	17,283	19,671
2	F_{MSY} ($F_{18\%}$)	24,810	23,850	22,972	22,260	21,710	21,295	20,982	20,745	20,564	20,426	21,961
3	$F_{2012-2014}$ ($F_{21\%}$)	25,114	25,242	25,217	25,144	25,063	24,995	24,942	24,901	24,869	24,845	25,033
4	$F_{30\%}$	25,638	27,797	29,585	31,042	32,212	33,151	33,903	34,506	34,985	35,367	31,819

Table 14. Projected trajectory of yield (metric tons) for alternative harvest scenarios. Fishing intensity ($F_{x\%}$) alternatives are based on $F_{16\%}$ (average 2003-2005), F_{MSY} ($F_{18\%}$), $F_{2012-2014}$ ($F_{21\%}$) (average 2012-2014 defined as current), and $F_{30\%}$. MSY = 19,901 metric tons.

Harvest scenario	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**Average**
1: $F_{2003-2005}$ ($F_{16\%}$)	25,688	24,044	22,890	22,089	21,522	21,111	20,806	20,576	20,402	20,268	21,940
2: F_{MSY} ($F_{18\%}$)	23,194	22,336	21,693	21,234	20,905	20,667	20,491	20,359	20,259	20,182	21,132
3: $F_{2012-2014}$ ($F_{21\%}$)	20,267	20,162	20,047	19,958	19,895	19,852	19,822	19,800	19,785	19,774	19,936
4: $F_{30\%}$	15,015	15,802	16,386	16,833	17,177	17,442	17,648	17,808	17,932	18,028	17,007

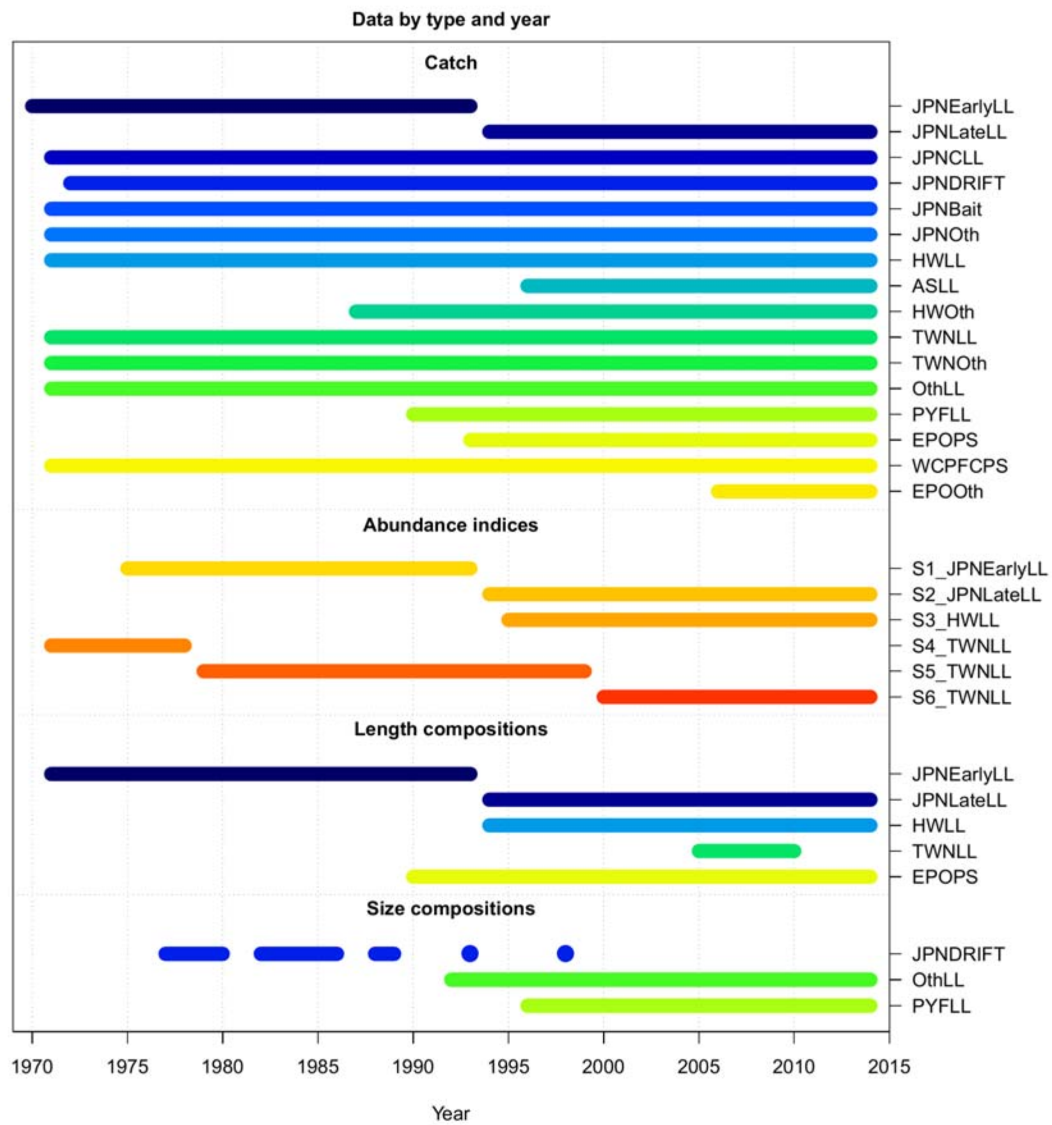

Figure 1. Available temporal coverage and sources of catch, CPUE (abundance indices), and length and size composition for the stock assessment update of the Pacific blue marlin.

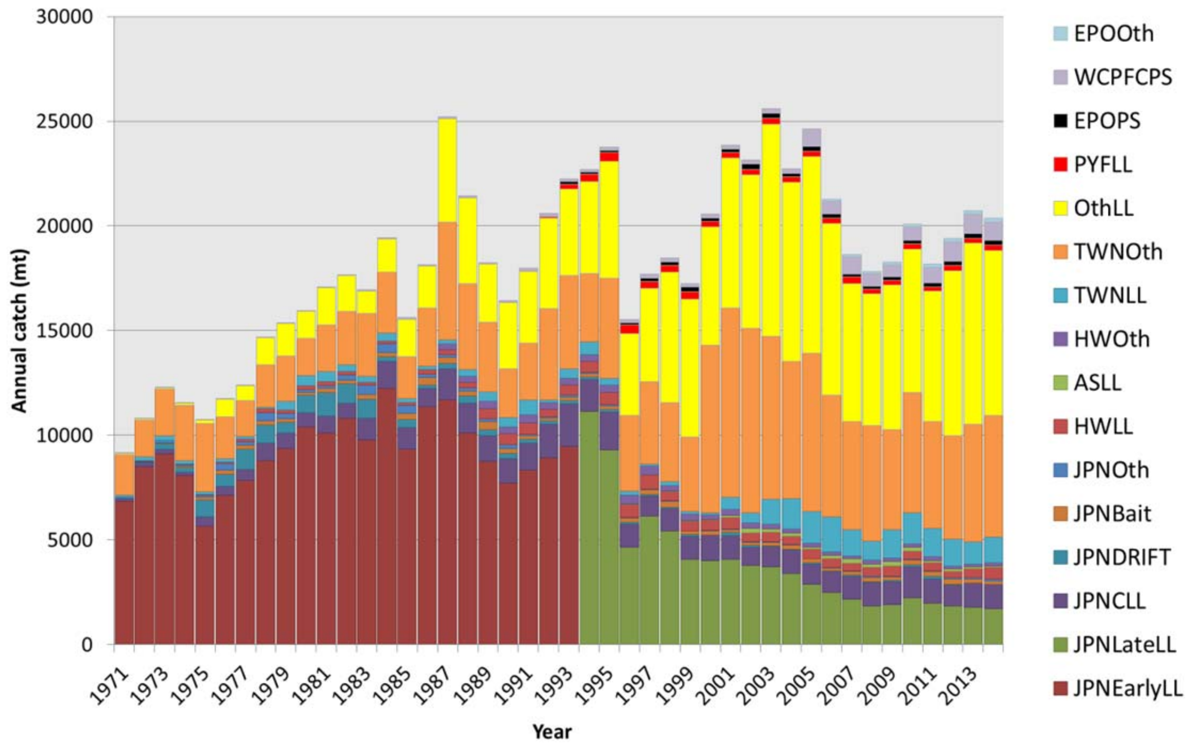

Figure 2. Total annual catch of the Pacific blue marlin by all fisheries harvesting the stock during 1971-2014. See Table 1 for the reference code for each fishery.

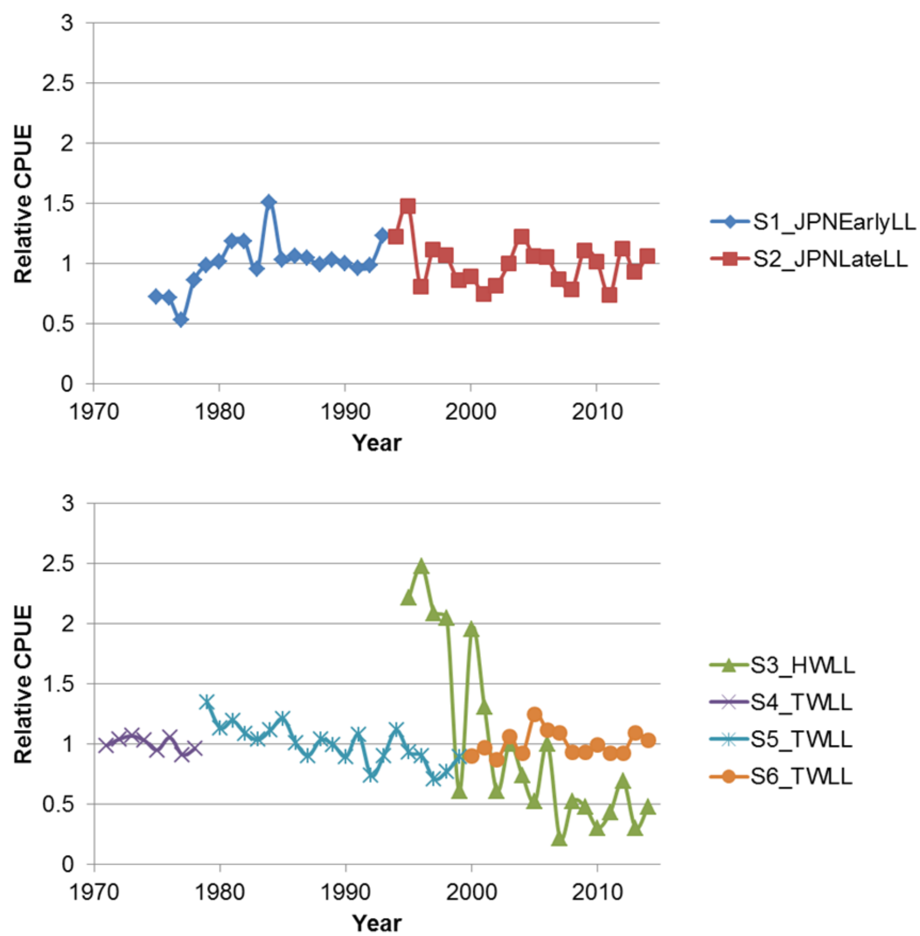

Figure 3. Time series of annual standardized indices of catch-per-unit-effort (CPUE) for the Japanese distant water longline fisheries (top panel); Hawaii-based longline and Taiwan distant water longline fisheries (bottom panel) for the Pacific blue marlin as described in Table 2. Index values were rescaled by the mean of each index for comparison purposes.

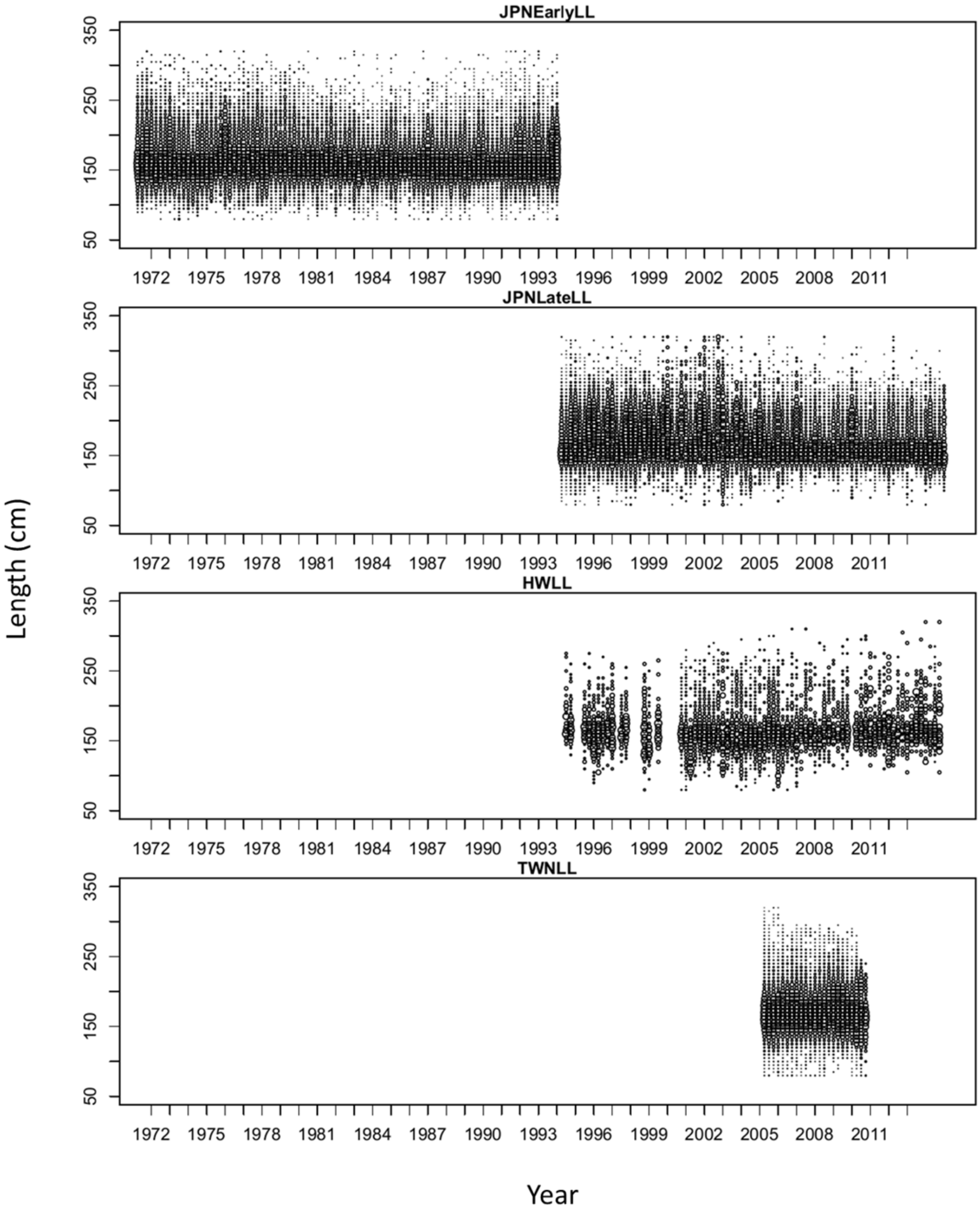

Figure 4. Quarterly length and size composition data by fishery used in the stock assessment update (see Table 3). The sizes of the circles are proportional to the number of observations. All measurements were eye- fork lengths (EFL, cm) except JPNDRIFT (kg).

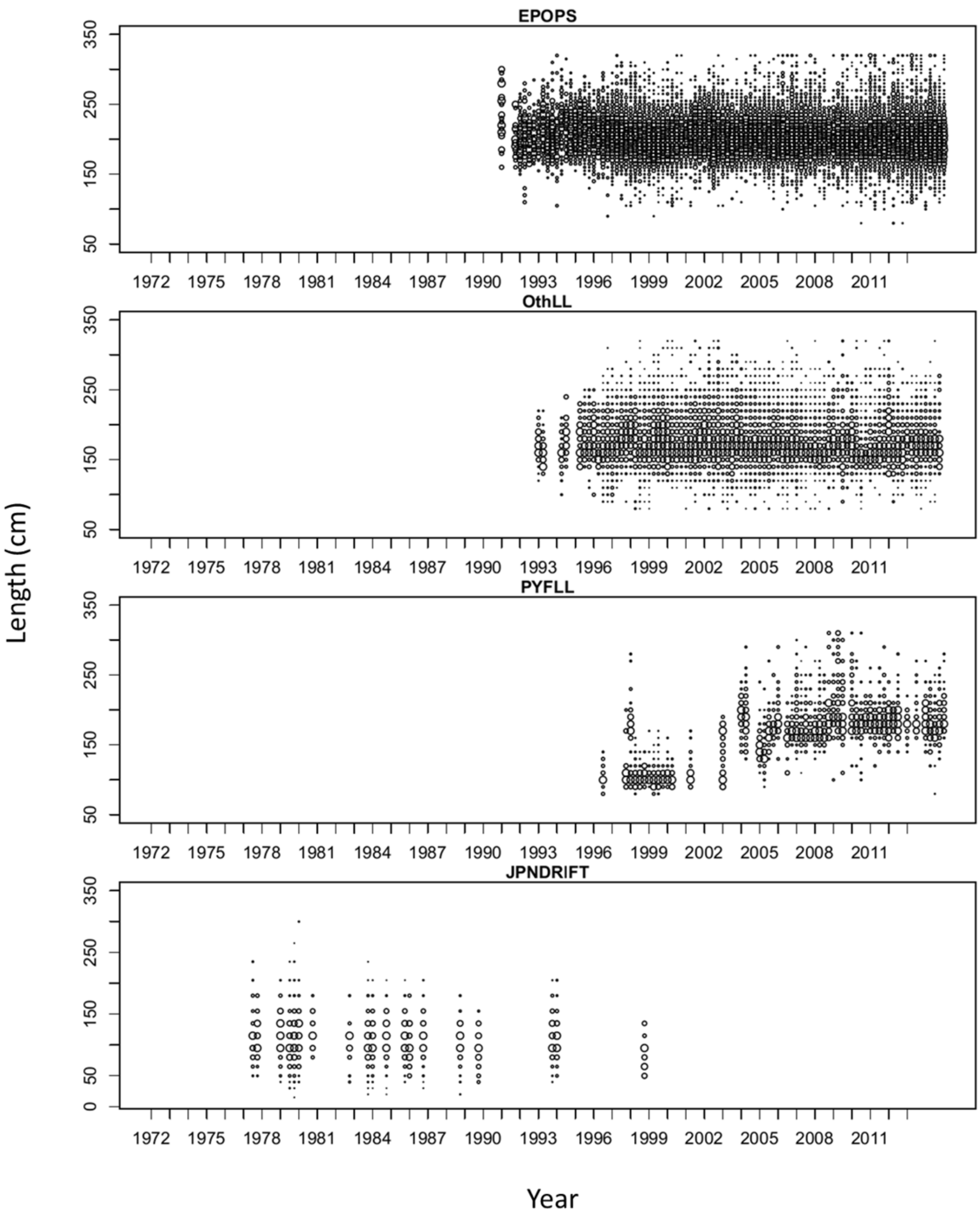

Figure 4. Continued.

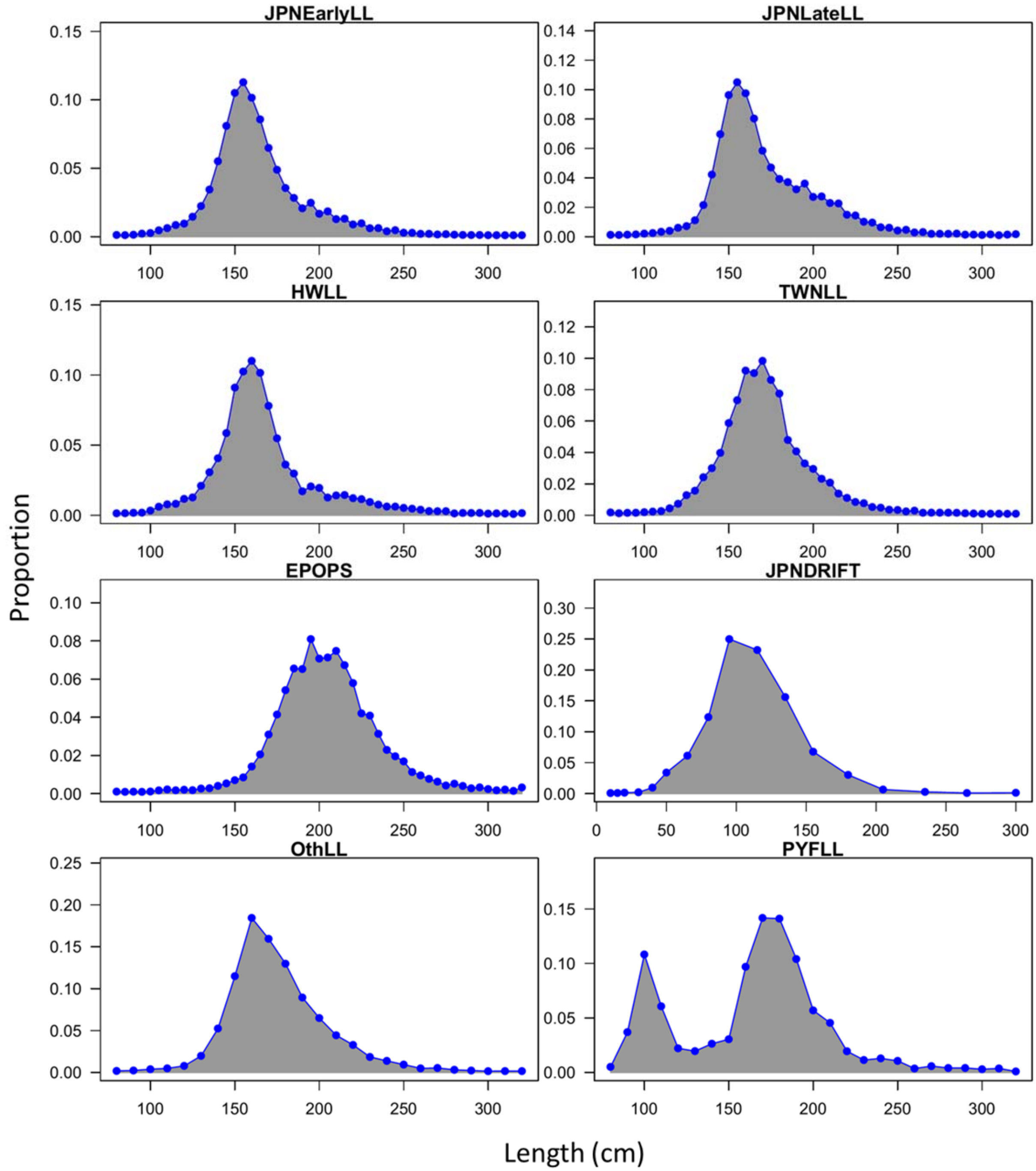

Figure 5. Aggregated length and size compositions used in the stock assessment update (see Table 3 for descriptions of the composition data). All measurements were eye- fork lengths (EFL, cm) except JPNDRIFT (kg).

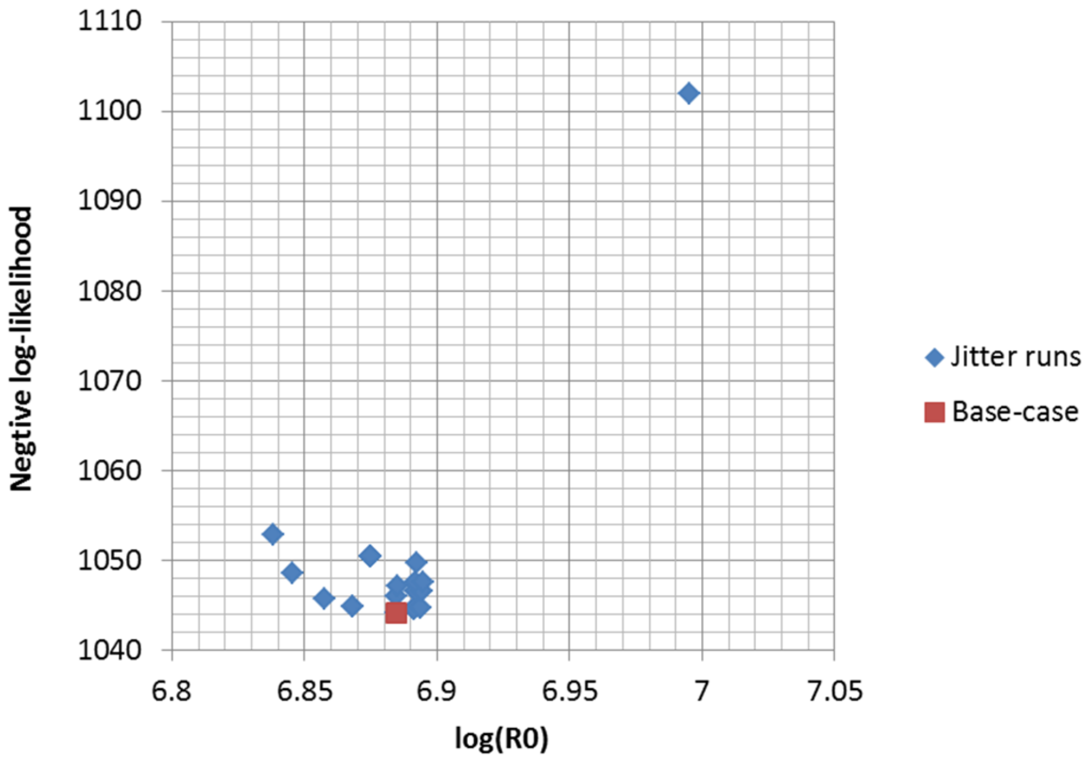

Figure 6. Total negative log-likelihood and estimated virgin recruitment in log-scale (log(R0)) from 30 model runs with different initial values estimated parameters in the base case model. Red triangle indicates results from model run using agreed upon parameters for the updated base case model, which had the lowest total negative log-likelihood (1044.2) of all 30 model runs.

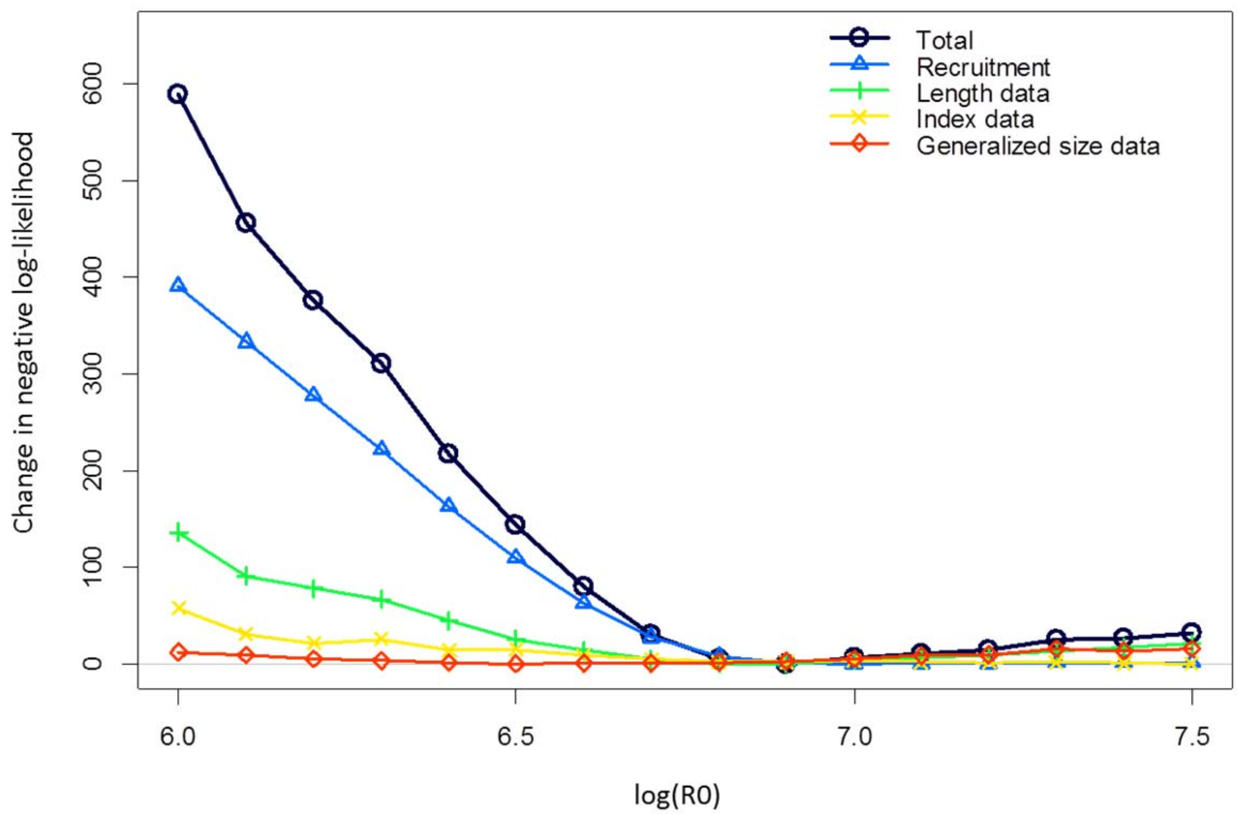

Figure 7.1. Profiles of the relative-negative log likelihoods by different likelihood components for the virgin recruitment in log-scale (log(R0)) ranged from 6.0 to 7.5 of the base case scenario.

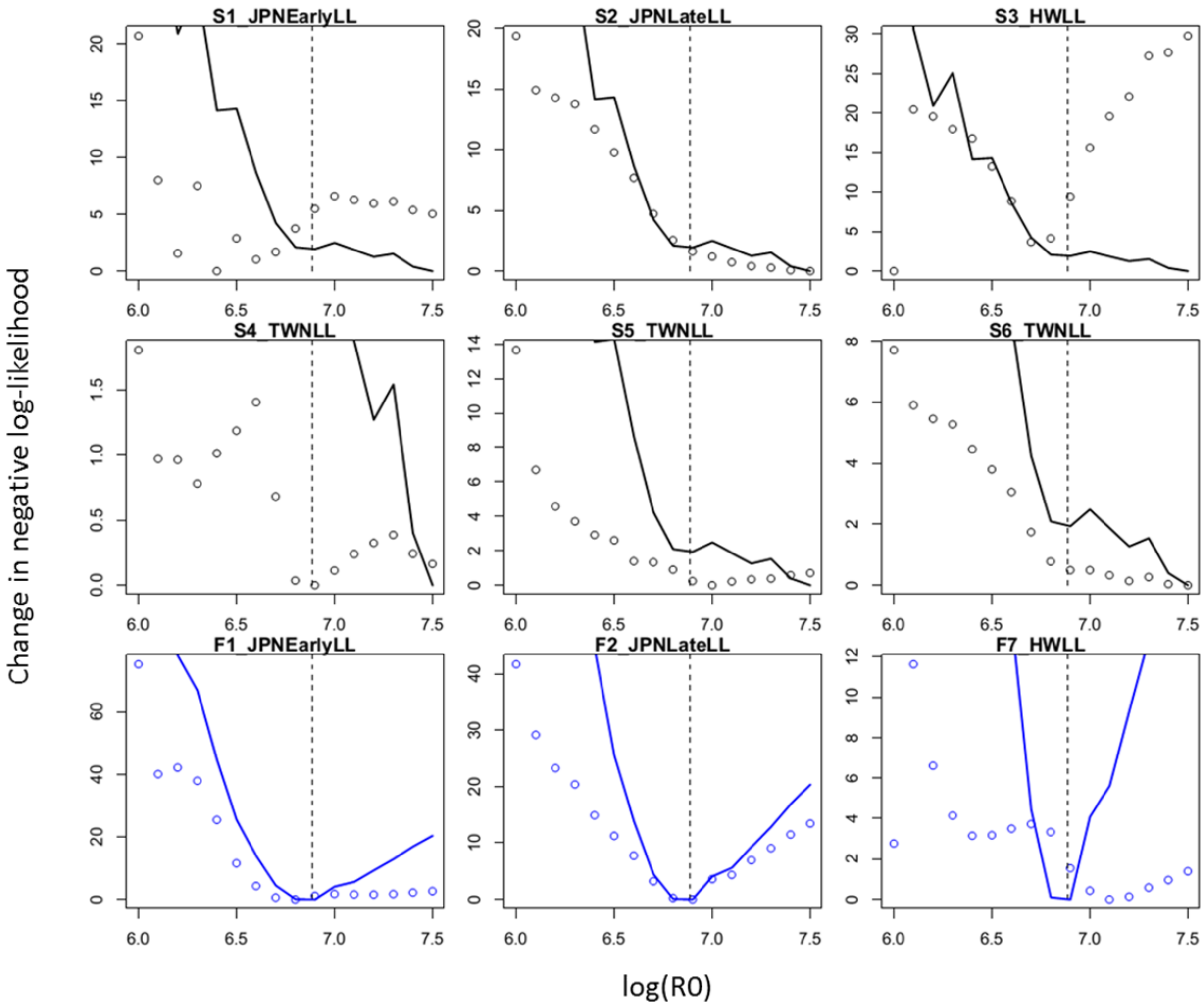

Figure 7.2. Profiles of the relative-negative log likelihoods by index (black circles), length composition (blue circles), generalized-size composition (red circles) likelihood components for the virgin recruitment in log-scale (log(R0)) ranged from 6.0 to 7.5 of the base case scenario. Black, blue, and red lines denote the fleet combined changed in log-likelihood for index, length composition, and generalized-size composition, respectively. See Tables 2 and 3 for descriptions of the index and composition data. S3_HWLL was not included in the total likelihood.

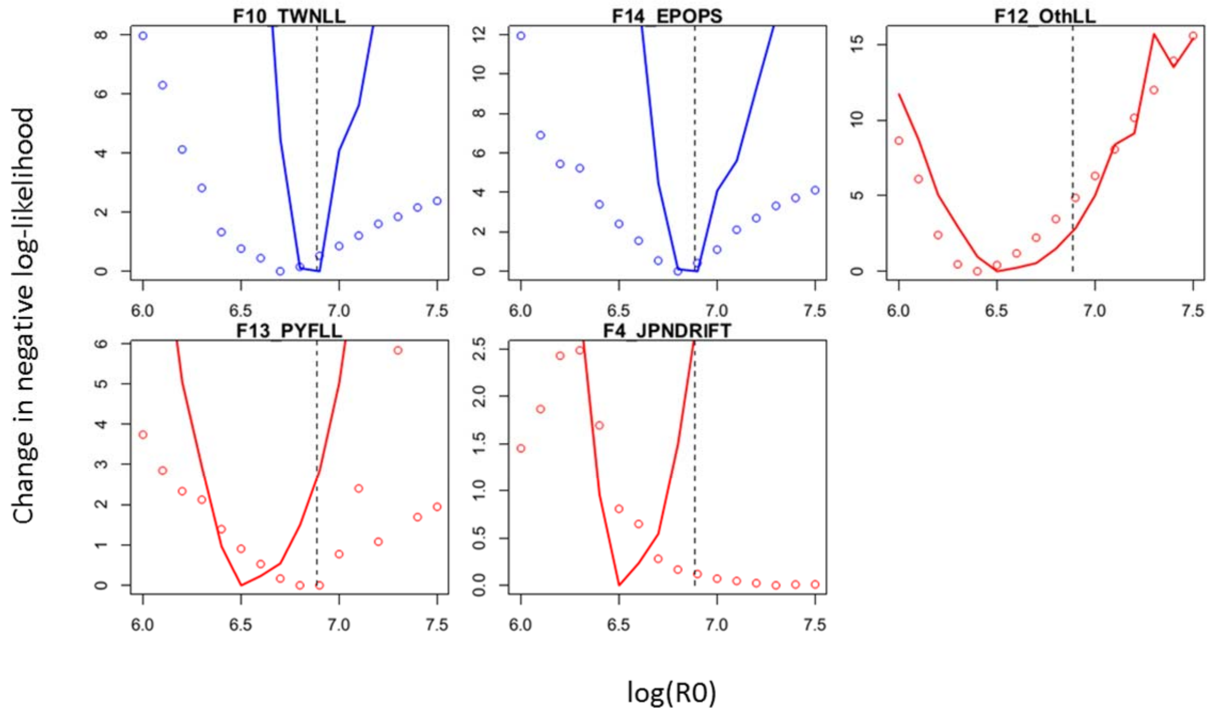

Figure 7.2 Continued.

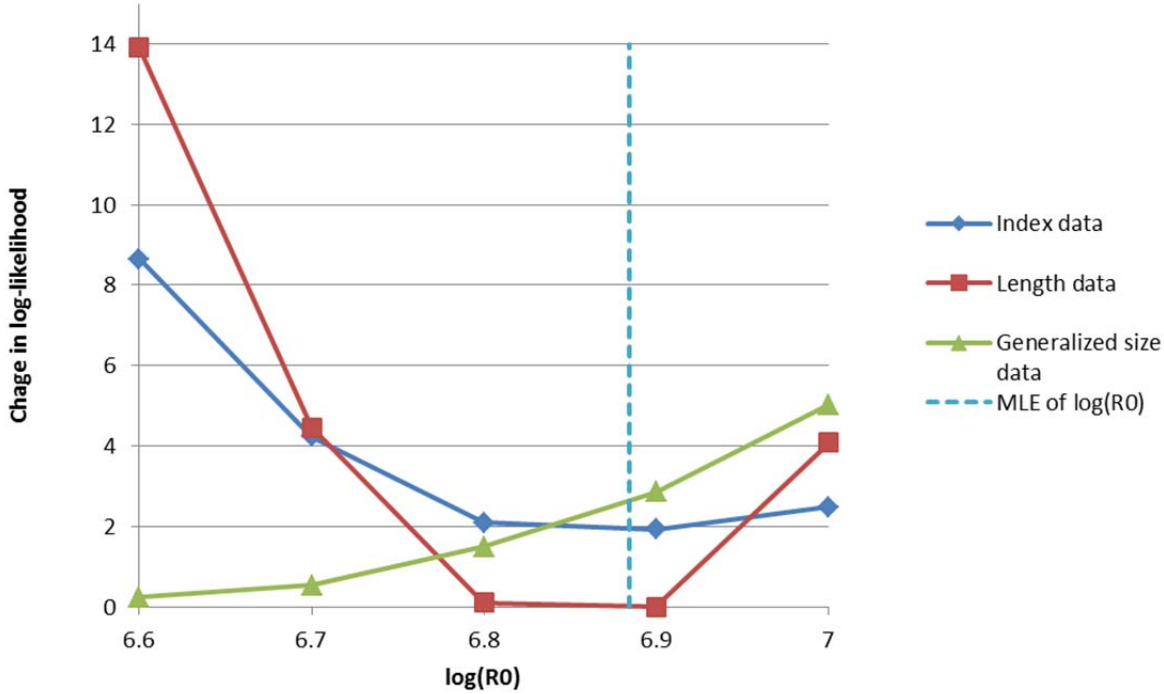

Figure 7.3. Zoomed in profiles of the relative-negative log likelihoods by index, length composition, and generalized-size composition likelihood components for the virgin recruitment in log-scale (log(R0)) ranged from 6.6 to 7.0 of the base case scenario.

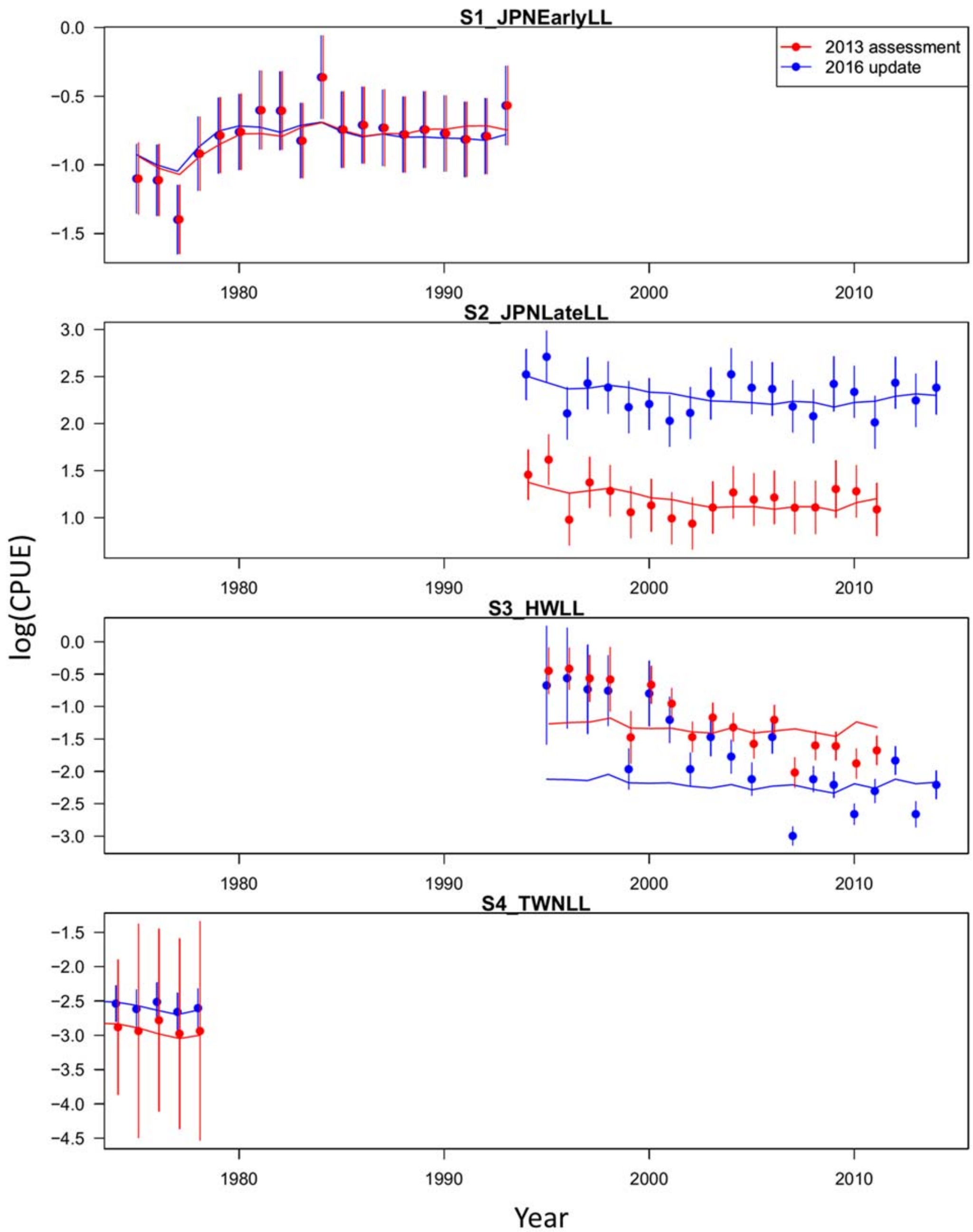

Figure 8. Model fits to the standardized catch-per-unit-effort (CPUE) data sets from different fisheries for the base case scenario. The line is the model predicted value and the points are observed (data) values. The vertical lines represent the estimated confidence intervals (± 1.96 standard deviations) around the CPUE values. Red color = 2013 assessment, blue color = 2016 update. S3_HWLL was not included in the total likelihood.

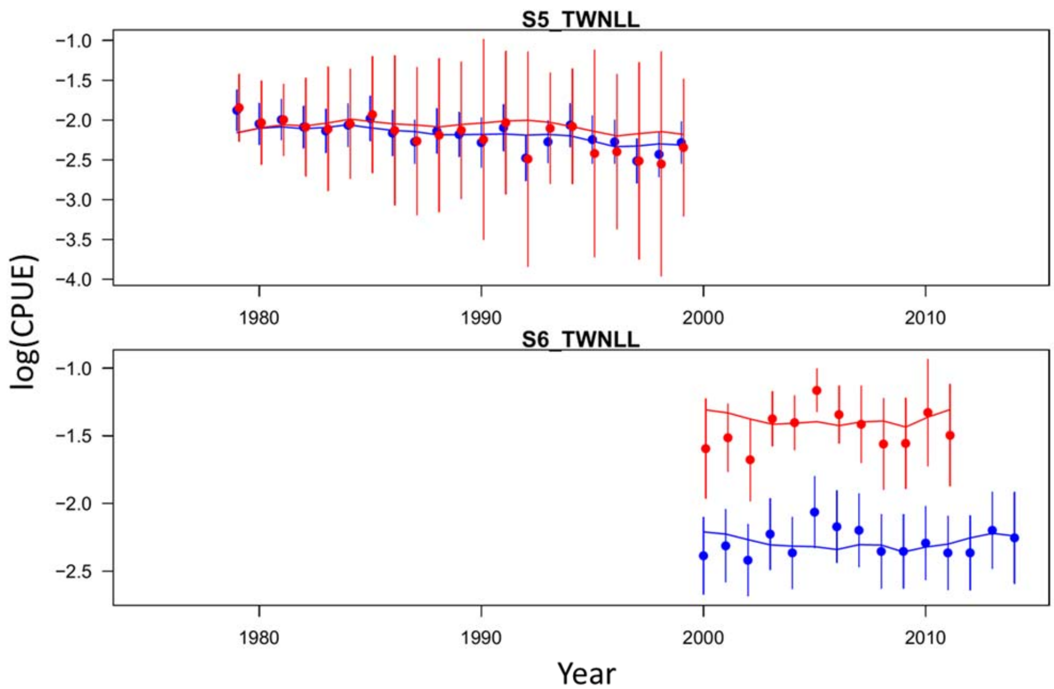

Figure 8. Continued.

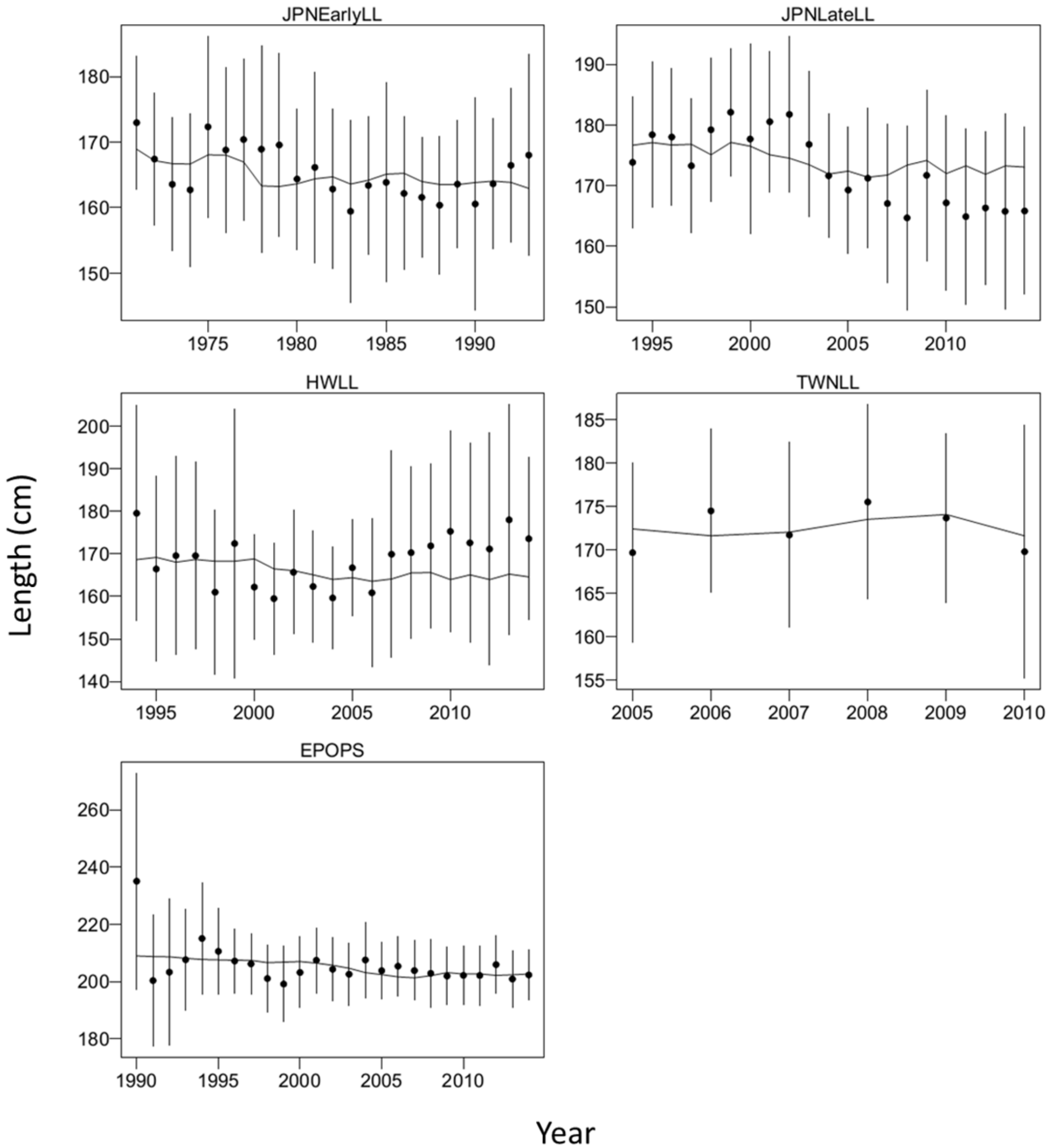

Figure 9. Model fit (lines) to mean length of the composition data (points, showing the observed mean age and 95% credible limits around mean age (vertical lines)). See Table 3 for descriptions of the data. All measurements were eye-to-fork lengths (EFL, cm) except JPNDRIFT (kg).

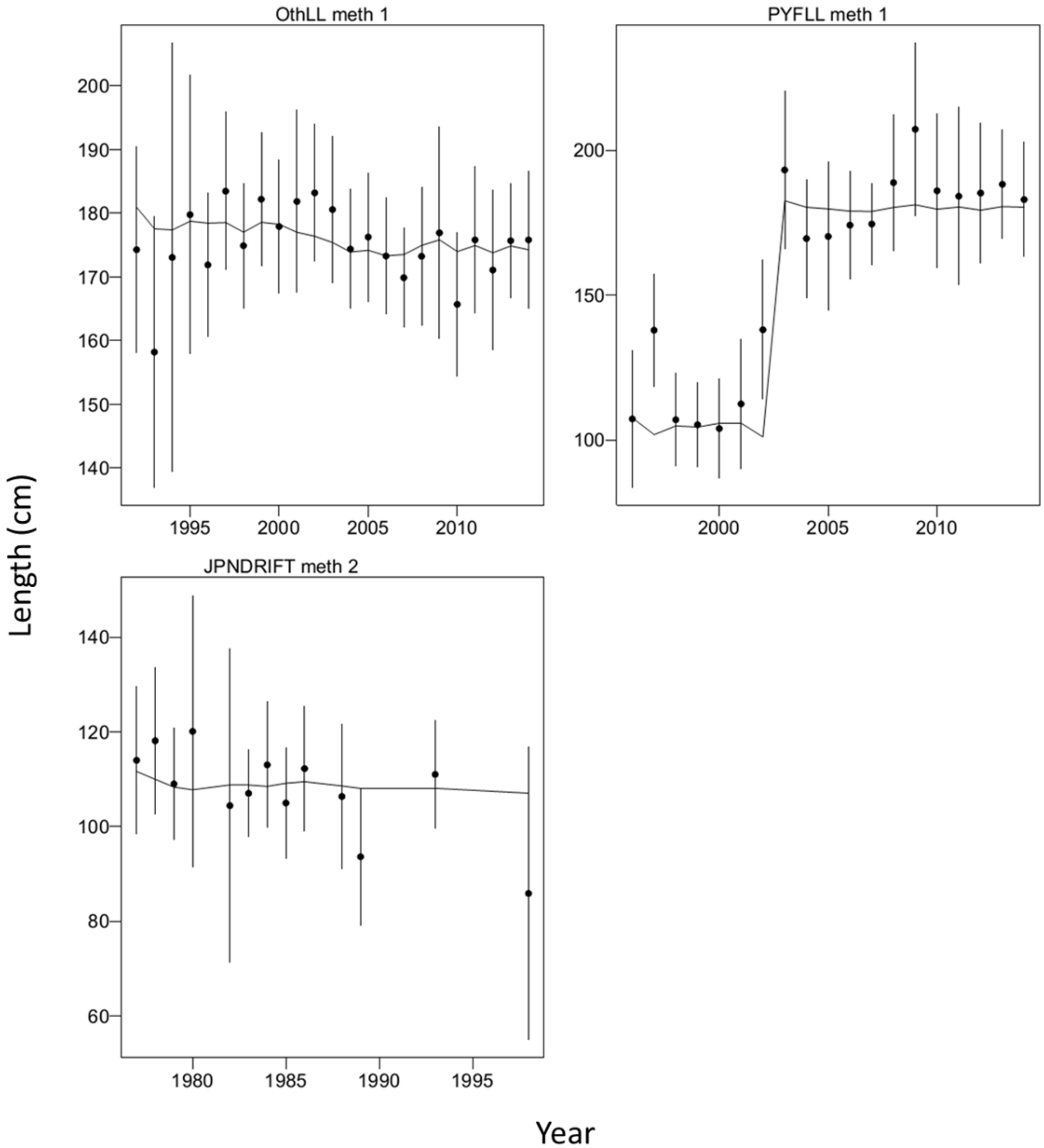

Figure 9. Continued.

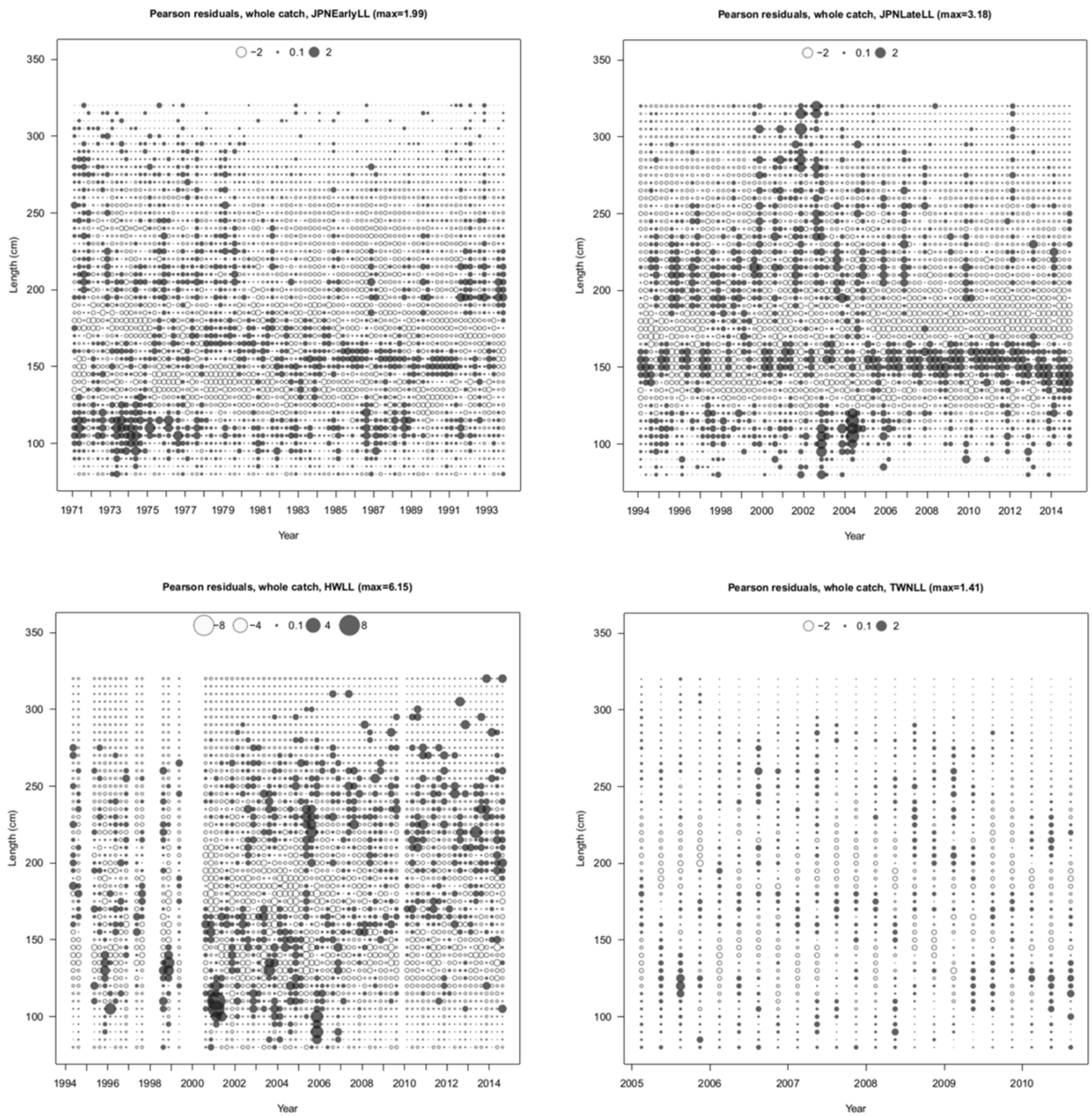

Figure 10. Pearson residual plots of model fits to the various length-composition data for the Pacific blue marlin fisheries used in the assessment model.

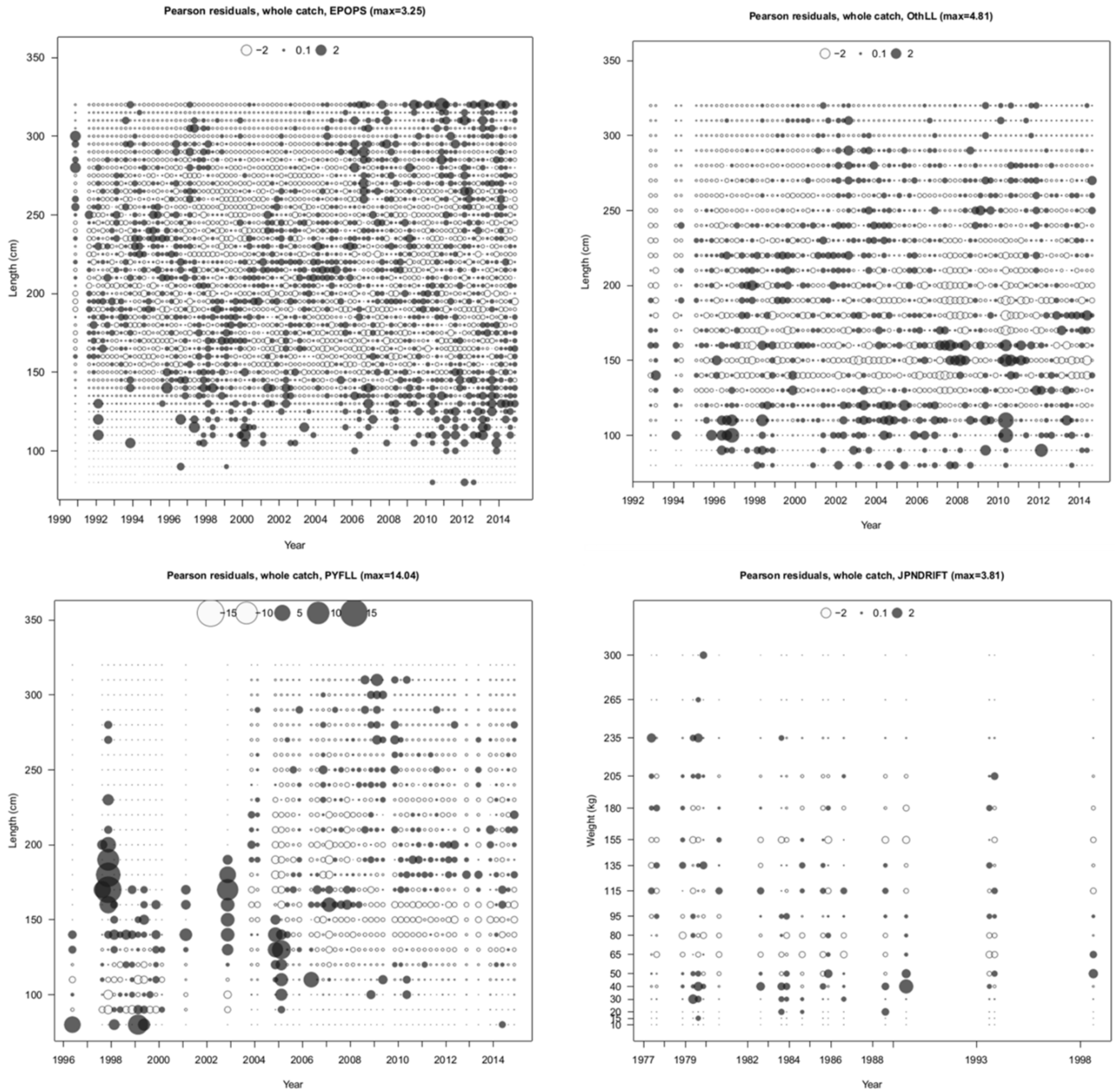

Figure 10. Continued.

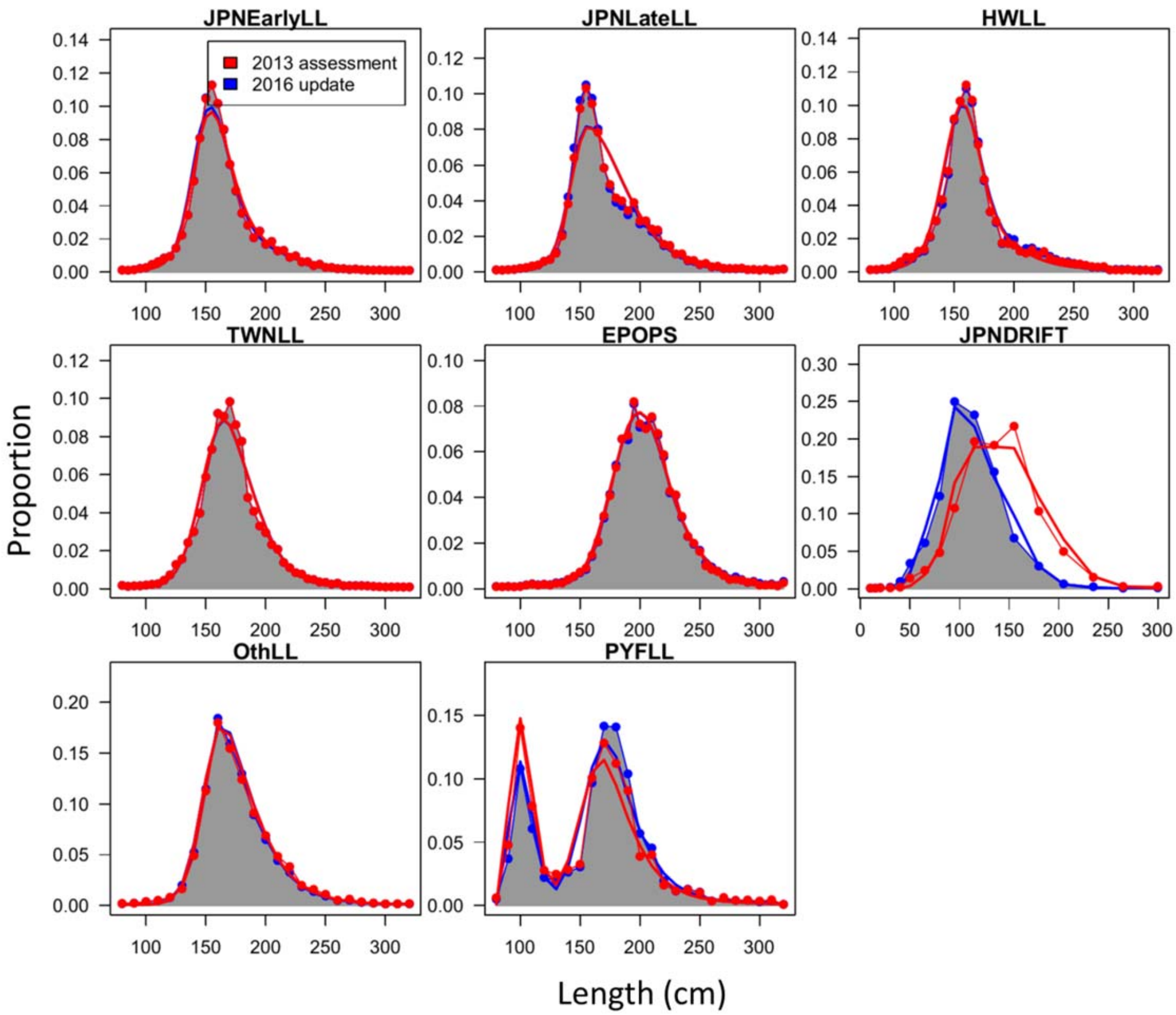

Figure 11. Comparison of observed (gray shaded area and blue dots) and model predicted (blue solid line) length compositions for fisheries used in the updated stock assessment for the Pacific blue marlin. Red colors indicate observed (dots) and predicted (line) length compositions from the 2013 assessment. All measurements were eye-to-fork lengths (EFL, cm) except JPNDRIFT (kg).

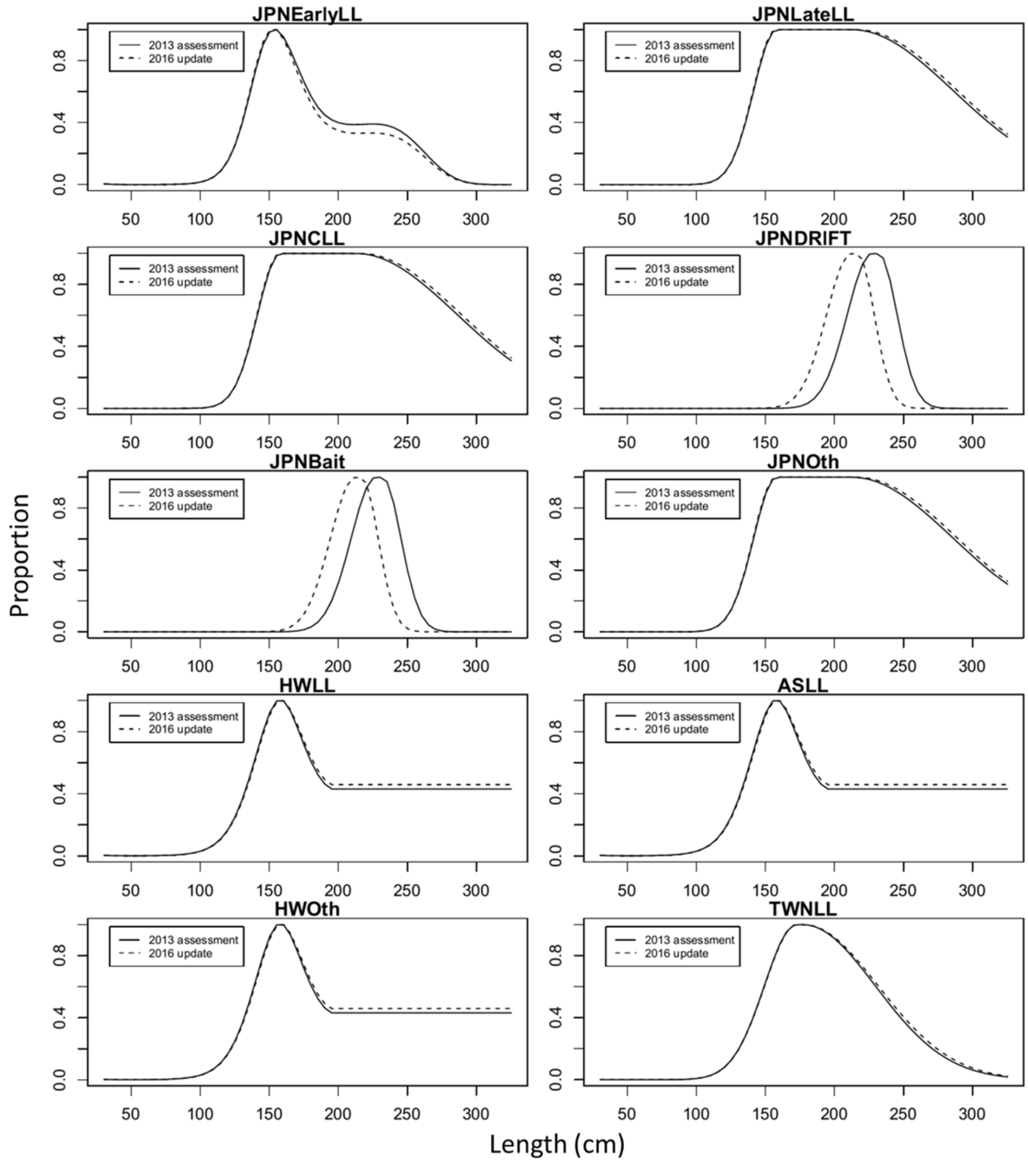

Figure 12. Comparison of length-based selectivity of fisheries for Pacific blue marlin between the 2013 stock assessment (solid lines) and the 2016 update (dash lines). Different colors denote the selectivity curves by time blocks.

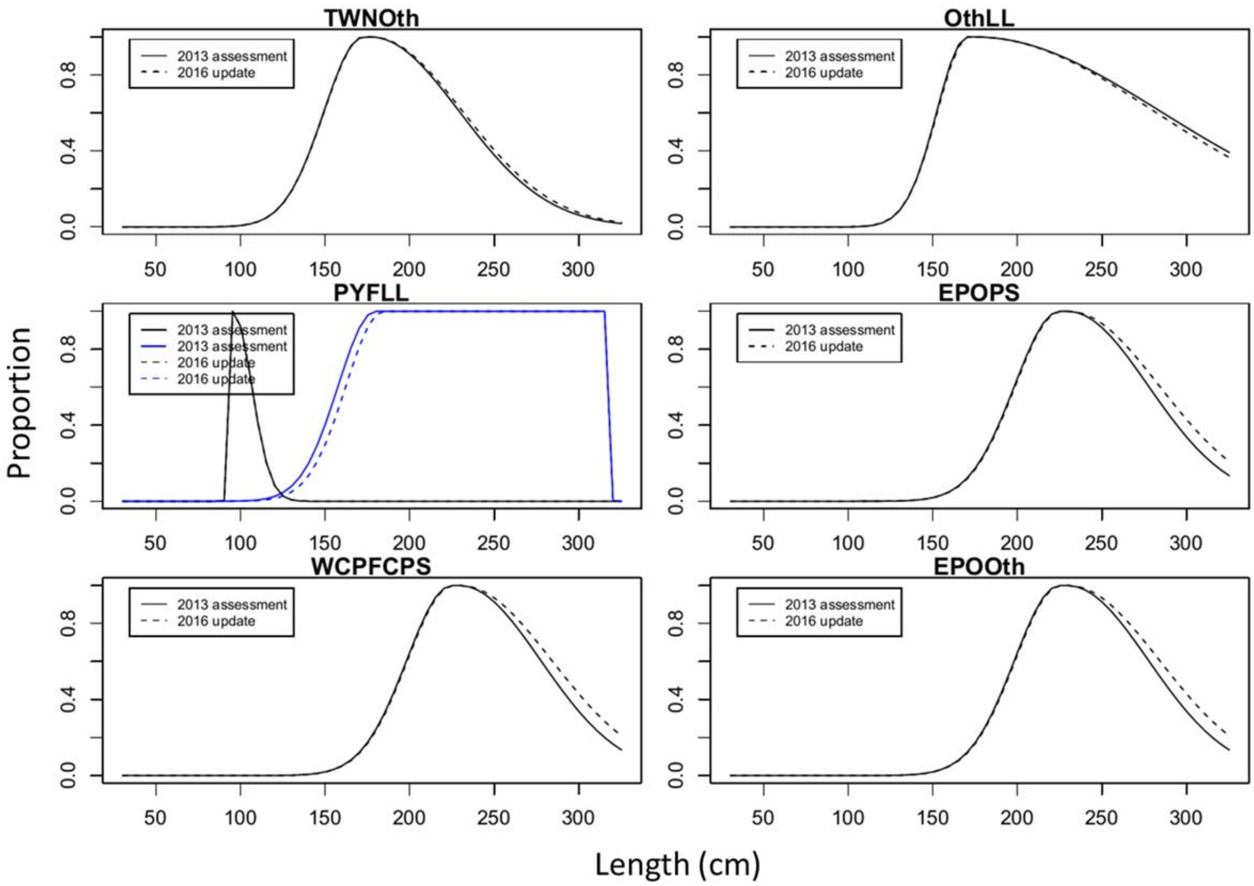

Figure 12. Continued.

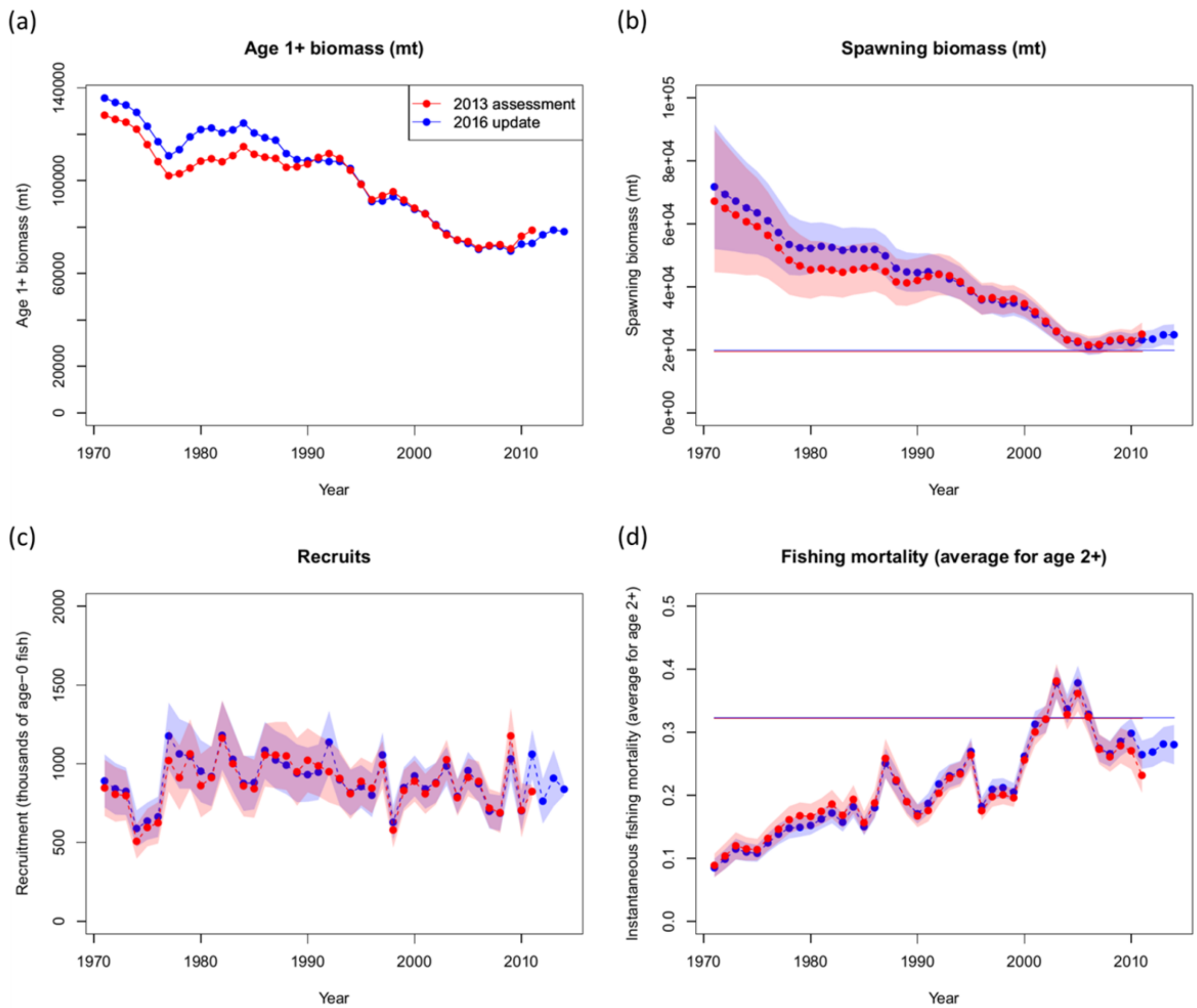

Figure 13. Comparison of time series of (a) total biomass (age 2 and older), (b) spawning biomass, (c) age-0 recruitment, and (d) instantaneous fishing mortality (year^{-1}) for Pacific blue marlin between the 2013 stock assessment (red) and the 2016 update (blue). The solid line with circles represents the maximum likelihood estimates for each quantity and the shadowed area represents the uncertainty of the estimates (± 1 standard deviations). The solid horizontal lines indicated the MSY-based reference points.

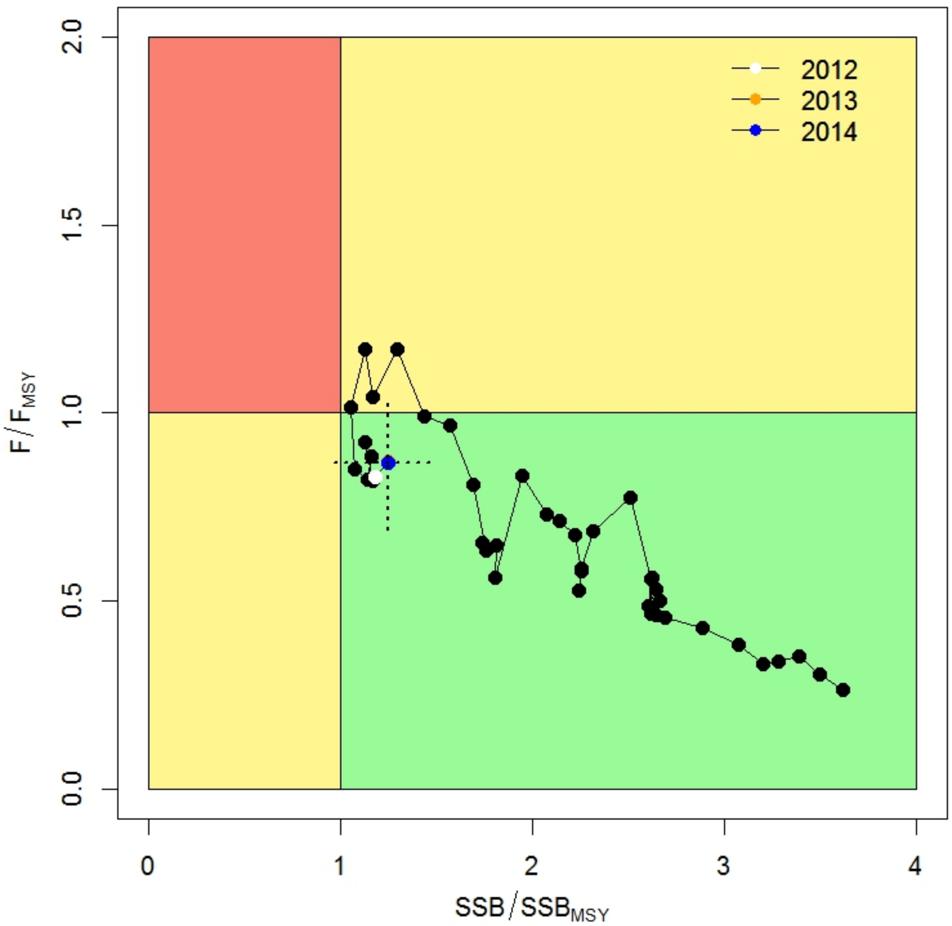

Figure 14. Kobe plot of the trends in estimates of relative fishing mortality (average of age 2+) and spawning stock biomass of Pacific blue marlin (*Makaira nigricans*) during 1971-2014. The dash lines denote the 95% confidence intervals of year 2014.

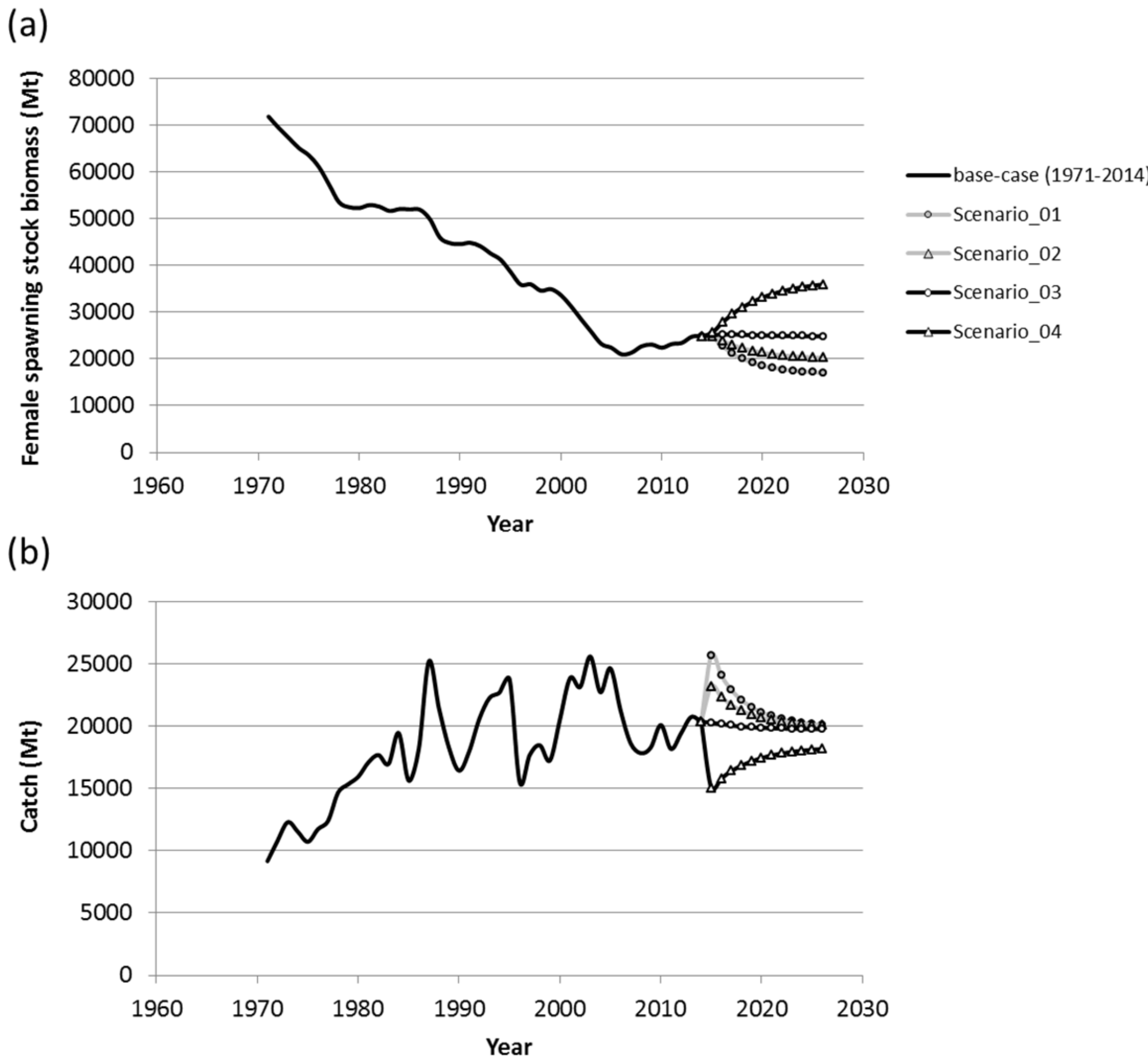

Figure 15. Historical and projected trajectories of (a) spawning biomass and (b) total catch from the Pacific blue marlin base case model. Scenario 1 = average fishing intensity during 2003-2005 ($F_{2003-2005}= F_{16\%}$); scenario 2 = F_{MSY} ($F_{18\%}$); scenario 3 = average fishing intensity during 2012-2014 ($F_{2012-2014}= F_{21\%}$); scenario 4 = $F_{30\%}$.

Appendix I

Table A1. Correlation matrix of abundance indices. Lower diagonal values are correlation coefficient and upper diagonal values indicate number of overlapped years. Colors indicate levels of correlation (blue: high positive correlation, red: high negative correlation). See Table 2 for descriptions of each abundance index.

	S1	S2	S3	S4	S5	S6
S1 (1975-1993)	19	0	0	4	15	0
S2 (1994-2014)	NA	21	20	0	6	15
S3 (1995-2014)	NA	0.24	20	0	5	15
S4 (1971-1978)	0.38	NA	NA	8	0	0
S5 (1979-1999)	0.11	0.23	-0.14	NA	21	0
S6 (2000-2014)	NA	0.22	-0.24	NA	NA	15

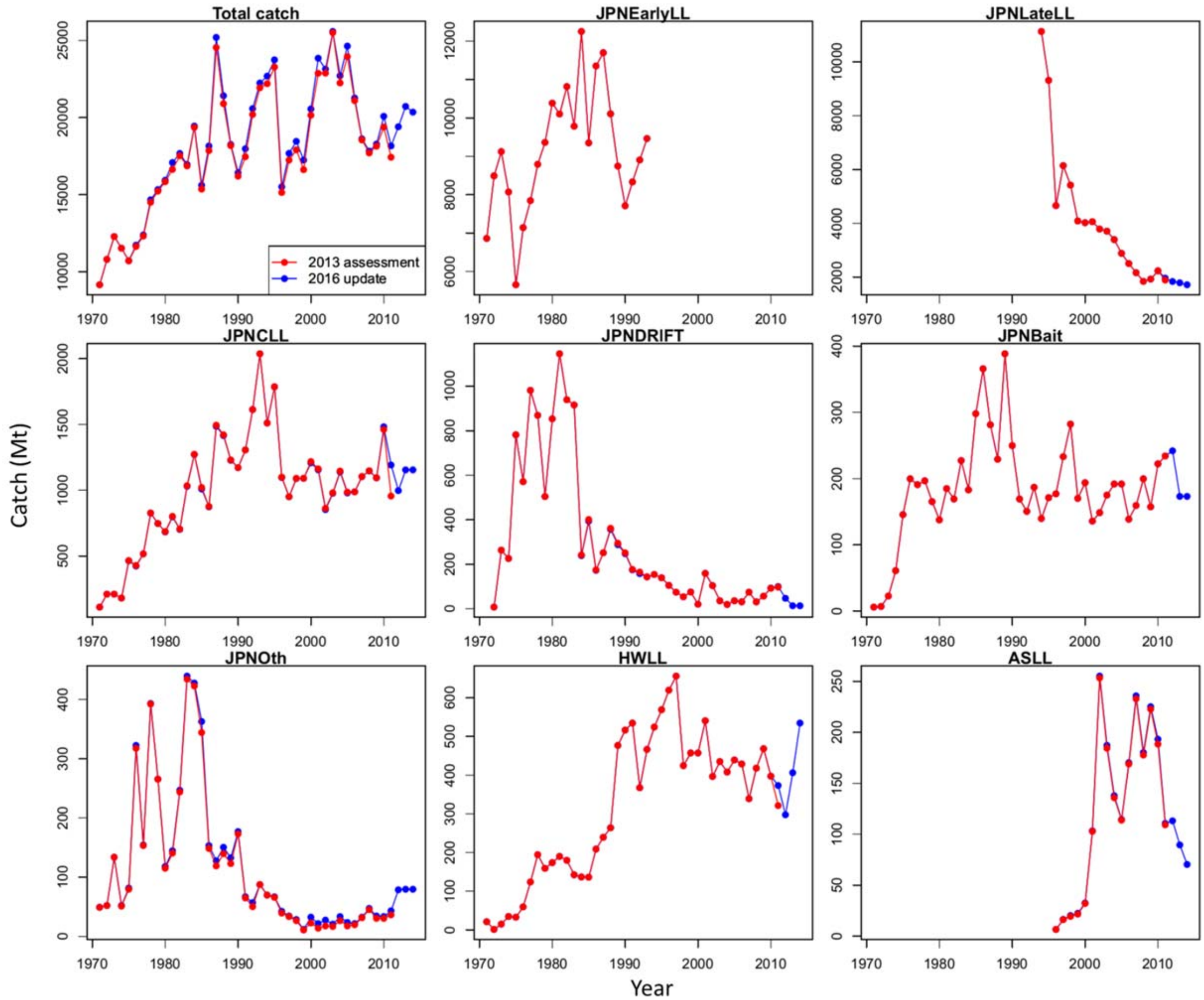

Figure A1. Comparison of time-series of catch of Pacific blue marlin used in the 2013 stock assessment (red line) and the 2016 update (blue line).

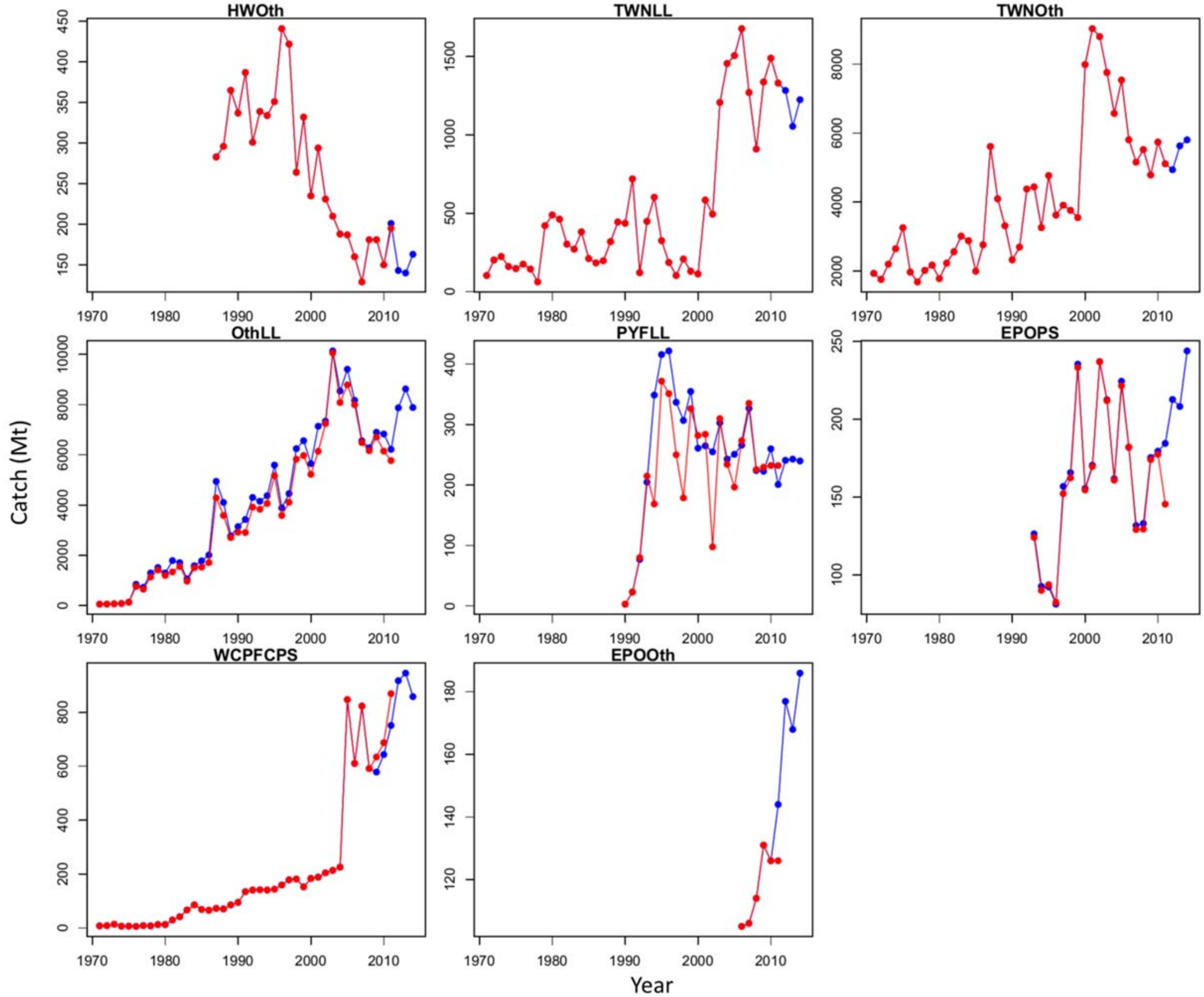

Figure A1. Continued.

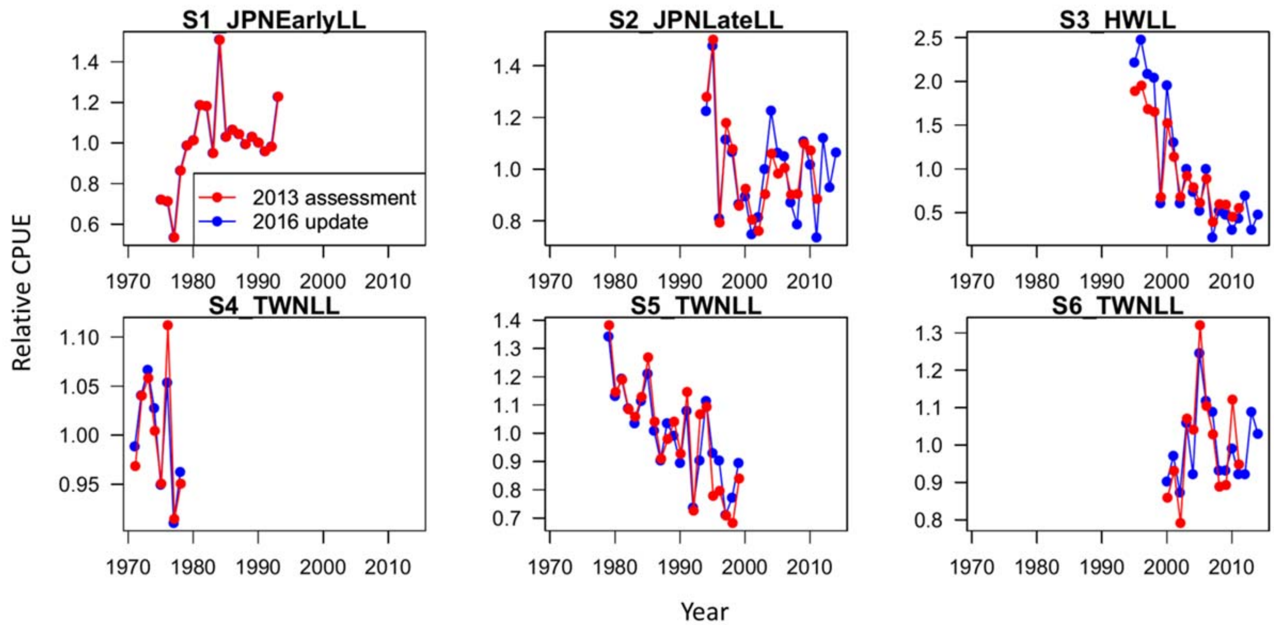

Figure A2. Comparison of relative abundance indices (in relative scale) of catch-per-unit-effort (CPUE) for Pacific blue marlin in the 2013 stock assessment (red line) and the 2016 update (blue line).

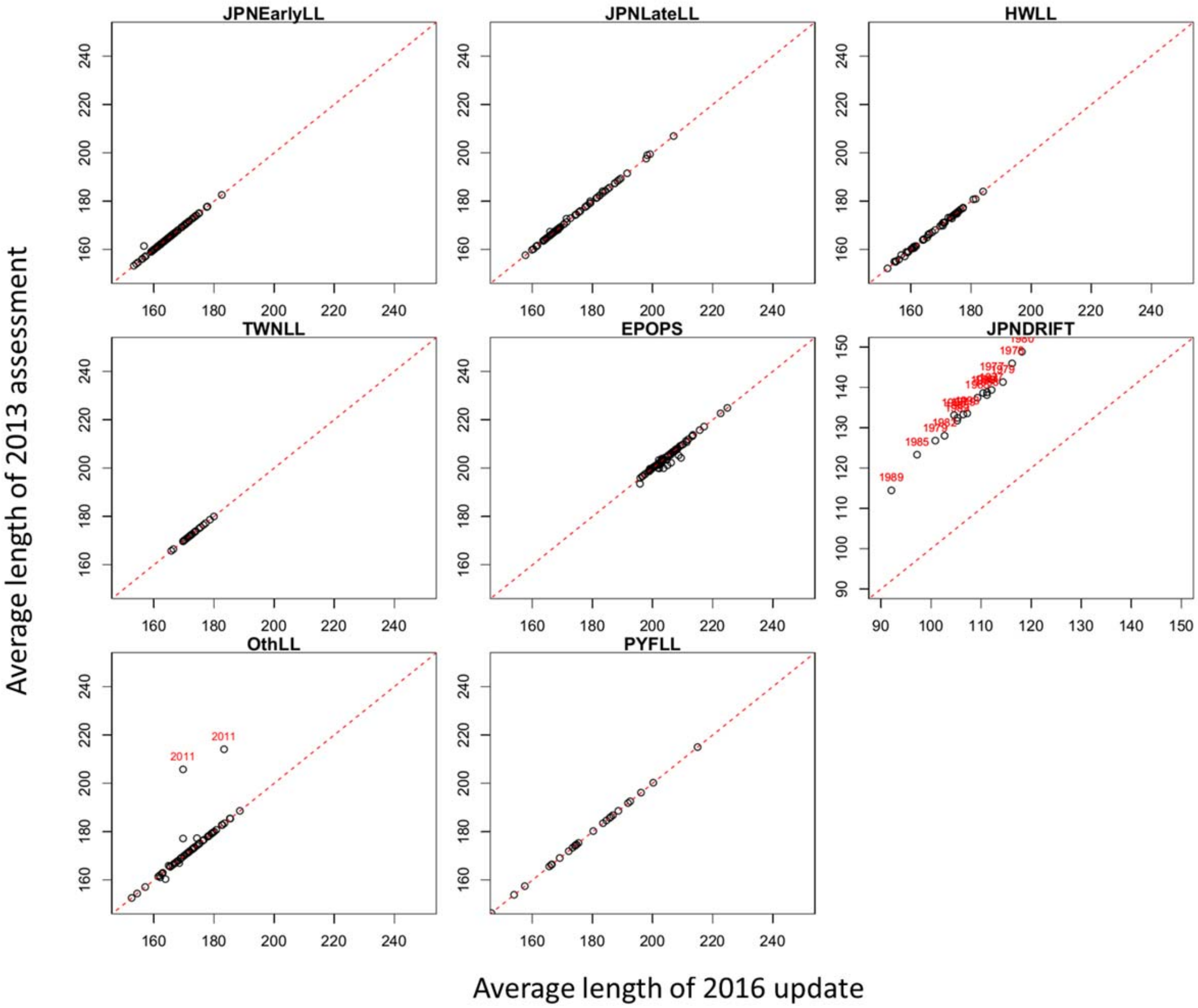

Figure A3. Comparison of average length of the input composition data for the Pacific blue marlin in the 2013 stock assessment and the 2016 update. All measurements were eye-to-fork lengths (EFL, cm) except JPNDRIFT (kg). The red labels denote the years with differences larger than 5%.

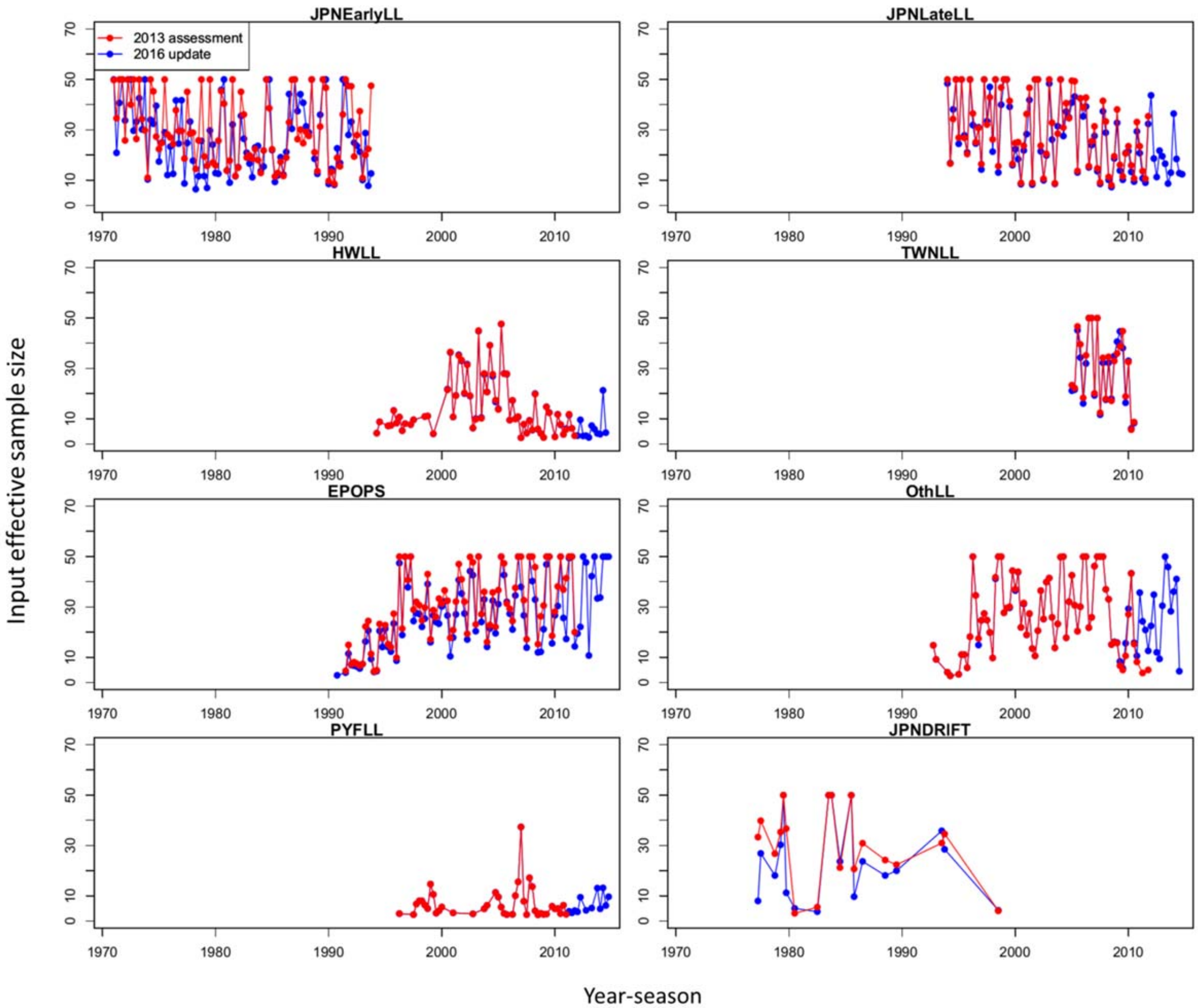

Figure A4. Comparison of the input effective sample size of the multinomial length composition for the Pacific blue marlin in the 2013 stock assessment (red line) and the 2016 update (blue line).