

CURRENT STATUS OF THE BLUE MARLIN (*MAKAIRA NIGRICANS*) STOCK IN THE ATLANTIC OCEAN 2024: PRE-DECISIONAL STOCK ASSESSEMENT MODEL

Michael Schirripa¹

SUMMARY

*This document describes the pre-decisional base case model configured to estimate the status of the blue marlin (*Makaira nigricans*) stock for the June 2024 stock assessment meeting. The model configuration is based on the 2018 model used to provide management advice. Uncertainties specifically accounted for were growth, stock-recruitment steepness, natural mortality and conflicting CPUE trends. Uncertainties not accounted for were, inter alia, seasonal and/or aerea differences in life history traits and illegal, unreported, and unregulated (IUU) landings. Several assumptions were investigated via different model configurations, namely four steepness values (0.40, 0.50, 0.70 and model estimated) and two natural mortality values (0.148 and model estimated for females).*

RÉSUMÉ

*Le présent document décrit le cas de base du modèle pré-décisionnel configuré pour estimer l'état du stock de makaira bleu (*Makaira nigricans*) pendant la session d'évaluation du stock réalisée en juin 2024. La configuration du modèle est basée sur le modèle de 2018 utilisé pour formuler un avis de gestion. Les incertitudes spécifiquement prises en compte étaient la croissance, la pente à l'origine de la relation stock-recrutement (steepness), la mortalité naturelle et les tendances contradictoires de la CPUE. Les incertitudes non prises en compte étaient, entre autres, les différences saisonnières et/ou géographiques dans les caractéristiques du cycle vital et les débarquements illégaux, non déclarés et non réglementés (IUU). Plusieurs postulats ont été étudiés par le biais de différentes configurations du modèle, à savoir quatre valeurs de steepness (0,40, 0,50, 0,70 et une valeur estimée par le modèle) et deux valeurs de mortalité naturelle (0,148 et une valeur estimée par le modèle pour les femelles).*

RESUMEN

*Este documento describe el caso base del modelo pre-decisionario configurado para estimar el estado del stock de aguja azul (*Makaira nigricans*) para la reunión de evaluación de stock de junio de 2024. La configuración del modelo se basa en el modelo de 2018 utilizado para proporcionar asesoramiento en materia de ordenación. Las incertidumbres que se tuvieron en cuenta específicamente fueron el crecimiento, los valores de inclinación de la relación stock-recrutamiento, la mortalidad natural y las tendencias contradictorias de CPUE. Las incertidumbres que no se tuvieron en cuenta fueron, entre otras, las diferencias estacionales y/o geográficas en los rasgos del ciclo vital y los desembarques ilegales, no declarados y no reglamentados (IUU). Se investigaron varios supuestos con distintas configuraciones de modelo, a saber, cuatro valores de inclinación (0,40, 0,50, 0,70 y modelo estimado) y dos valores de mortalidad natural (0,148 y modelo estimado para las hembras).*

KEYWORDS

Stock Synthesis, Blue Marlin, stock assessment, diagnostics, population dynamics

¹ NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA. Michael.Schirripa@noaa.gov

1. Introduction

Blue marlins (*Makaira nigricans*) are a large apex pelagic predator species within the family of Istiophorida (the billfish). Their growth rate is among the fastest documented for a fish species and can reach weights in excess of one thousand pounds with females growing significantly larger than males (Wilson *et al.* 1991). They display extensive movements within and across the Atlantic, as demonstrated from various tagging studies (Orbesen *et al.* 2011). Spawning takes place over a wide seasonal time range with a peak occurring in the summer months. Little information is available on the larval and juvenile stages as they rarely occur in biological samples.

The International Commission for the Conservation of Atlantic Tuna (ICCAT) is responsible for the management of the Atlantic blue marlin. Catches of blue marlin are predominately incidental from surface longline fisheries targeting tuna, however, there are also documented catches from near-shore gillnet fisheries. There is also a substantial targeted recreational fishery in North and South America as well as the Caribbean, with most of the fish being released. The ICCAT manages the Atlantic as a single unit stock.

The Atlantic Blue Marlin (*Makaira nigricans*) stock assessment meeting was held June 17-21, 2024. One of the modeling platforms employed was the fully integrated Stock Syntheses (SS3) (Methot and Wetzel 2013). This document describes the post-decisional stock assessment model configuration and diagnostics. Diagnostic tests used are those described in Carvalho *et al.* 2021.

2. Methods

This assessment was carried out using Stock Synthesis (SS3) [V3.30.22.1;_safe;_compile_date:_Jan 30 2024]. Assumptions and configurations were guided by the decisions provided during the 2024 blue marlin data preparatory meeting (Anonymous 2024). Some analyst discretion was provided via the Group to account for unforeseen circumstances that arose during the assessment process.

The model configured with one area, two sexes, five fishing fleets and twelve surveys. Fishing fleets included artisanal gillnet, longline, purse seine, rod & reel and moored fish aggregation devices (mFAD). All observational data used to fit the model are shown in **Figure 1**.

Fleet Structure. The assessment model was configured with five fleets: (1) artisanal-gillnet, (2) longline, (3) purse seine, (4) rod & reel, and (5) Fish Aggregation Device (FAD) fleet. The artisanal fleet uses mostly gillnet gear. The FAD fishery uses mostly purse seine gear. The FAD fishery fleet is an additional fleet not considered in the 2011 assessment. Even though no CPUE and only a small number of lengths are available from this fishery, given the increased use of FADs in general, it was believe to be important to maintain the individual identity of this fishery for future examination.

The decisions regarding the reclassification of “gillnet” to “artisanal” as well as the addition of a FAD fishery were made at the Blue Marlin DP meeting. The classification from “gillnet” to “artisanal” was made to better reflect the variety of gears used for this fishery: drift-gillnets, trolling-hooks, longline). The FAD fleet is primarily industrial purse seines fishing for tropical tuna. The addition of the FAD fleet was made to provide a finer resolution and to more accurately classify the landings of this type of fishing as it has become apparent that fishing on FADs has been gradually increasing.

Landings. Landing were provided by the ICCAT Secretariat Task 1 data (**Figure 2**). Four catch scenarios were considered: (1, baseline), landings plus reported dead discards; (2, sensitivity), landings plus dead discards plus live discards multiplied by the minimum longline post-release mortality obtained from the scientific literature with the calculation being performed outside the assessment model; (3, sensitivity), landings plus dead discards plus live discards multiplied by the maximum longline post-release mortality obtained from the scientific literature with the calculation being performed outside the assessment model; and (4, sensitivity), landings plus discards and assuming a five percent post-release mortality on the rod & reel fishery estimated internal to the assessment model, as was done in the 2018 stock assessment.

Biology. Natural mortality (M) was allowed to vary between sexes. For males M was fixed at 0.148. For females M was either fixed at 0.148 or freely estimated. The observational data that informed the estimation of female M was contained primarily within the sex specific age-at-size data. The first mature age for females was assumed to be age_1 with fecundity being defined as a logistic function of length. Length at 50% maturity was set to 206 cm LJFL (Shimose *et al.* 2015).

Growth in Blue Marlin remains poorly determined. Data was available from several studies; however, there were very meaningful differences in the estimated size-at-age between the investigations. Four sources were identified that could help estimate growth. The first, and oldest, source was available from Wilson (1991) via Goodyear (2000). Only the estimated von Bertalanffy growth parameters were available from this source. The second source was Shimose *et al.* (2015) which used otoliths and fin spines to describe growth in the western North Pacific. The third source was provided by Hoolihan *et al.* 2019 who used transverse section of the second spine from the first anal fin, coupled with estimated compensation for vascularization. The fourth source was presented by Krusic-Golub *et al.* at the 2024 blue marlin data preparatory meeting (SCRS/P/2024/007), who used otoliths to estimate age. Although overall size-at-age does not seem to be definitively determined, all studies agree that growth is extremely rapid in the first one to two years with a significant slowing thereafter. This pattern of growth is not well suited for use of the traditional three parameter von Bertalanffy growth model. I used a Richards growth function to fit the observations of age-at-size available within the SS3 framework (Methot and Wetzel 2013)

Indices of abundance. The indices of abundance (i.e. survey) used were made available during the 2024 data preparatory meeting (Anonymous 2024). Indices used included: (1) US longline/observer, 1993-2022; (2) Venezuela longline, 1991-2018; (3) Venezuela artisanal (gillnet), 1991-2022; (4) Venezuela recreational rod & reel, 1961-2001; (5) Chinese Taipei longline-early, 1968-1989; (6) Chinese Taipei longline-mid, 1990-1997; (7) Chinese Taipei longline-late, 1998-2022; (8) Brazil longline, 1978-2005; (9) Japan longline, 1994-2022; (10) Ghana gillnet, 2000-2009; and (11) Japan longline-historic, 1959-1993.

The coefficient of variation of each index of abundance was derived by maintaining the same trend as reported by the authors but scaling the trend to a minimum of 0.30.

Time varying catchability. The Atlantic yellowfin tuna market began a shift towards bigeye tuna in the late 20th century. This led to a shift in the Japanese and Chinese-Taipei fleets targeting effort away from yellowfin tuna and on to bigeye tuna. This targeting bigeye tuna was accomplished through setting longlines relatively deeper than those targeting yellowfin tuna. Blue marlin and yellowfin tuna have much greater habitat overlap than do blue marlin and bigeye tuna. Thus, effort directed away from yellowfin tuna was also directed away from blue marlin. The flag-specific vector of the ratio of yellowfin tuna in the yellowfin tuna plus bigeye tuna in the Japanese and Chinese-Taipei fisheries up until 1980 was used as a proxy for the targeting of yellowfin tuna and thus a catchability for blue marlin. The flag-specific ratio vectors were used “as data” to modulate the catchability of the two fleets by estimating a parameter to relate the relationship between the catchability and the ratio.

Lengths. Length compositions by gear type of Blue Marlin were provided by the ICCAT Secretariat.

Data weighting. Length and age composition data was weighted according to Francis, R.I.C.C. (2011). The method uses the quantitative fit of the model parameters to the length and age data, in the form of residuals, relative to the fit to the survey data to balance the fit between the three sets of data.

Stock-recruitment relation. A Beverton-Holt stock-recruitment relation was used with virgin recruitment (R0) freely estimated and steepness either fixed or freely estimated, depending on the prescribed configuration.

Selectivity. Selectivity was length-based with age 1 fish and older being fully selected. The double-normal function was used for those fisheries whose selectivity was estimated.

Retention. A retention function was applied to recreational fisheries starting in 1993 to accommodate the minimum legal size regulation.

Discard mortality. For the catch scenario that required use of recreational release mortality (i.e. catch scenario #4) a value of 5 percent was.

3. Results

The resulting fitted parameter values for the baseline model are given in **Table 1**.

Growth and maturity. The resulting fit to the female growth information is shown in **Figure 3**, top. The sex-specific growth curves are shown in **Figure 3**, bottom. While upon initial examination, the fitted growth curve seems to under estimate the female size at age it needs to be kept in mind that the selectivity of the gear is considered in the fitting process. In this case, the selectivity of the gear suggests that the observational data reflects mostly the faster growing individuals.

Fit to indices of abundance. The model fit to each of the indices of abundance, along with the time varying catchability of select fleets, are shown in **Figures 4** and **5**. Note that because of the variance adjustment that was used some indices are seen as more/less precise as others and thus varying degrees of fit are expected.

The parameter associated with the relation between the Chinese-Taipei (early) CPUE and the yellowfin tuna:bigeye tuna species ratio (-0.286) was meaningful and improved the fit to the index. A similar result was found with the Japanese time varying catchability (-0.089) but to a lesser extent. The trends in the time varying catchabilities are shown in **Figures 5** and **6**.

For experimental purposes, a single spline was fit to all CPUEs used in the model. Limited experimentation with this index was not fruitful and although it remained in the model, it was given zero weight. Nonetheless, within the lack of fit to spline fit index can be seen some of the effect of the time varying catchability. The spline fit did not consider time varying catchability and shows a marked decrease from 1965 to 1980, a decrease not reflected in the estimated stock trend, which did use the time varying feature.

Several CPUEs (Chinese-Taipei_LL (late) and Japan_LL) showed increasing trends in the last few years of data. The spline fit to all CPUEs also showed a slight increasing trend. Other CPUEs were mostly flat or without trend.

Fit to length compositions. The SS3 fit to the gear specific length compositions for all years is shown in **Figure 6**. Fit the length compositions aggregated across years fit very well. However, there were signs of year-specific residual patterns in some fleets. Efforts to alleviate these patterns lead to over-parameterization that eventually would have rendered the model less stable had they been maintained. It was determined that the better option was to keep the number the number of parameters at a minimum to maintain the stability of the model and to attribute the lack of fit to less than random sampling of the catch.

Selectivity. The three SS3 estimated gear specific selectivities are shown in **Figure 7**. The artisanal gillnet fleet was allowed to be dome-shaped yet was estimated to be asymptotic. Both the longline and recreational rod and reel fleets were assumed asymptotic. There was no *a priori* reason to believe that selectivity had changed over time.

Stock-recruitment. The estimated stock-recruitment function and resulting recruitment deviations are shown in **Figure 8**. Annual recruitment deviations did not show any sort of linear trend ($p = 0.386$) which suggests a random dispersal of points around the stock-recruitment function.

Model Diagnostics

Several standard diagnostic techniques were investigated and/or applied to the baseline model (Model_1.2) to evaluate the stability of the parameter estimates and general model performance.

1. Joint residual and Run Test on CPUEs
2. Profiling $R0$ (virgin recruitment) and h (steepness of S/R function)
3. Retrospective analysis
4. Hindcast cross-validation
5. Jitter analysis

A summary of the results are given in **Table 2**.

Joint residual and Run Test on CPUEs. To evaluate the overall model fit of the relative abundance indices and composition data, the joint-index residual plot was applied to the residuals from the fits to indices and mean length for multiple time series simultaneously. The code for this diagnostic plot was adapted from the JABBA R package (github.com/jabbamodel/JABBA) to Stock Synthesis output files and implemented as the plotting function `ss3diags::SSplotJABBAres()`, which provides the option to specify the type of data input. A summary of these results are provided in **Tables 2** and **3**.

Overall, the joint-index residual plot indicated only a “fair” fit with the RMSE around 52% for the CPUE data (**Figure 9**, top). A better fit was obtained to the mean lengths, which had an RMSE of 5.1% (**Figure 9**, bottom).

The runs test (Carvalho *et al.* 2017) is used to test for randomness of residuals of the model fit the CPUE and mean length time series (**Figure 10**). Failure to pass run tests is seen as an indication that process error is not being adequately characterized and that further model development is warranted. There was sufficient evidence ($p \geq 0.05$) to reject the hypothesis of randomly distributed residuals for 8 of the 12 indices of abundance.

Profiles on R0 and steepness. Profile analysis for R0 by data type is shown in **Figure 11**. Examination of profiling by individual data sources suggests agreement between the length and age data with a well-defined minimum. (**Figure 11**, top). However, there was conflict between the size-at-age and the CPUE data. The priors had very little influence on the estimation of R0. The profile of R0 was also examined by individual CPUE (**Figure 11**, middle). Most of the CPUEs fit better at lower values of R0 while the US_LL and Ven_Art fit best at high values (**Figure 11**, bottom). Profiles were also examined by individual length composition data. Conflict between the information in the length data was also apparent. However, the vast majority of the data (coming from longline) showed a clear minimum. There was a direct conflict between the Art_gillnet and the rod and reel length data.

Profile on steepness is shown in **Figure 12**. As with the estimate of R0, there was contrasting signals coming from the various data sources (**Figure 12**, top). However, the individual CPUEs where is relative agreement with and most showed a clear minimum $-\log(\text{likelihood})$ (**Figure 12**, middle). The individual length compositional data did not show as clear of a minimum (**Figure 12**, bottom).

Retrospective analysis. This analysis was carried out with 5-year peel (**Figure 13**). A 'rule of thumb', proposed by Hurtado-Ferro *et al.* (2015), suggests values of ρ_M that fall outside (-0.15 to 0.20) for SSB for longer-lived species, or outside (-0.22 to 0.30) for shorter-lived species indicates an undesirable retrospective pattern. In addition, the direction of the retrospective bias has implications for characterizing risk associated with management advice. A positive ρ_M for SSB is of particular concern because it implies a systemic overestimation of biomass, which would lead to over-optimistic quota advice if not taken into consideration (Hurtado-Ferro *et al.*, 2015).

Results from this diagnostic show that an undesirable retrospective pattern was evident. The values of ρ_M were as follows: SSB = -0.59; Recruitment = -0.27; Bratio = -0.62; and F = 1.09. Each of these values are outside the range proposed by Hurtado-Ferro *et al.* (2015).

No further diagnostics were investigated until the retrospective pattern was sufficiently addressed.

Hindcast cross validation. Implementing the HCxval diagnostic in Stock Synthesis required using the outputs produced for the retrospective routine generated by the R package r4ss. The forecasts are based on the settings in 'forecast.ss', which are also evoked when conducting future projections with the same model, only that the observed catches are used for the retrospective forecasts. A desirable feature of Stock Synthesis is that the software also computes the expected values of the observational data (e.g., abundance indices, length- or age-composition data) based on the forward-projections of the numbers- and catch-at-age matrices. Therefore, there are no additional computationally intensive tasks needed if HCxval is conducted in conjunction with retrospective analysis.

Of the ten indices of abundance used, six had MASE scores > 1 (failed) and four < 1 (passed), indicating that six had a superior prediction skill than the naïve baseline forecast of a random draw (**Figure 14**). The most accurate predictions were observed for the Age_3 index with a prediction residual MASE of 0.67 and the least accurate for the Portuguese longline CPUE with a prediction residual MASE of 3.08. The hindcast cross validation diagnostic again highlighted the lack of consistent signal emanating from all the CPUE indices considered.

Jitter analysis. A commonly used jitter value of 0.1 was used to test model stability. Results of the jitter analysis suggest relatively good model stability (**Figure 15**). Stability of the model configuration was challenged by the attempted estimation of the CPUE added variance parameters. Presumably due to the conflicting signal contained in the CPUE series, attempting to assign added variance to each on simultaneously lead to equally likely model fits. This inability is itself a useful model diagnostic highlighting the issues of conflicting CPUE times series.

Jack-knife analysis. To examine model sensitivity to each of the CPUE's each CPUE index was removed one at a time and the trends in SSB were compared (**Figure 16**, top). The removal of any one individual CPUE time series had virtually no effect on either the trend or the absolute value of SSB. This is likely due to the lack of a consistent signal emanating from the CPUEs as a whole. With thirteen indices, many of which are inconsistent with the others, the elimination of any one index will not lead to a significant change in the estimated trend in SSB.

However, when all CPUE's were removed and only the information in the length compositions considered, a different absolute value of SSB was estimated as well a different trend. While the CPUEs in isolation suggest an increasing stock size over the past two decade, the length compositions, in isolation, suggest a decreasing trend over the most recent years.

To examine model sensitivity to each of the length compositions each fleet specific composition was removed one at a time and the trends in SSB were compared (**Figure 16**, bottom). The most influential length composition was the US fleet. This is likely due to this is the only fleet in which the lengths of discards were separated from the lengths of retained fish; the lengths from all other fleets were categorized as "combined". The model would very likely be improved by a greater distinction between the lengths of fish that were discarded and those that were retained.

Sensitivity to natural mortality and steepness assumptions

Eight models were configured according to the decisions made at the data preparatory meeting. These models consisted of four assumptions pertaining to steepness (0.5*, 0.4, 0.7 and estimated) and two assumptions pertaining to natural mortality (0.148* and estimated) (**Table 4**). The trends in SSB are shown in **Figure 17**.

Management considerations.

Blue Marlin is primarily a bycatch species as opposed a species that is targeted directly. As such, even a "no retention" management action will result in a quantity of Blue Marlin being killed by encountering hooks. At haulback mortality of Blue Marlin from the US longline fleet is shown in (**Figure 18**). The observer data suggests that between 25 and 30% of Blue Marlin are dead at haul back from longlines. Thus, forecasts assuming zero TAC are not useful to depict the management action of "no retention" and do not reflect the potential population increase under a no retention policy. Similarly, any TAC that is low enough to be met within the year such that the fishery stops retaining however keeps fishing does will under estimate the actual fishing mortality and thus possible future stock trajectory.

4. Discussion

Model diagnostics suggest that there are conflicts between the signals in population trend amongst the various CPUE times series as well as between the CPUEs in general and the length compositions. These conflicts could be due to length sampling error/bias, inadequate CPUE standardization (i.e. observational uncertainties) and/or population model misspecification (i.e. structural uncertainties).

Model_8 (both natural mortality and steepness are estimated) likely best depicts the true uncertainty in the overall model solution. Model_8 had the highest coefficient of variation in the derived quantities, the expected result when more parameters are estimated. Although the estimates of female natural mortality ($M = 0.098$) are at the lower end of the "expected" range, the estimated steepness ($h = 0.627$) is well within the range considered in this assessment. For these reasons, Model_8 may be the best choice to move forward with should a single model be used to estimate stock status.

Acknowledgements

The author wishes to acknowledge the efforts of the ICCAT Secretariat (especially A. Kimoto, M. Ortiz and C. Palma) for their support in the development of this assessment.

References

- Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R. and Maunder, M.N., 2021. A cookbook for using model diagnostics in integrated stock assessments. *Fisheries Research*, 240, p.105959.
- Francis R.I.C. Chris. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*. 68(6): 1124-1138. <https://doi.org/10.1139/f2011-025>
- Goodyear, C.P., Simulated responses of the size distribution of blue marlin to fishing mortality. *Col.Vol.Sci.Pap. ICCAT*, 54 (3): 869-881. (2002)
- Hoolihan, J.P., Luo, J. and Arocha, F., 2019. Age and growth of blue marlin *Makaira nigricans* from the central western Atlantic Ocean. *Fisheries research*, 220, p.105346.
- 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., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *Ices J. Mar. Sci.* 72, 99–110. <https://doi.org/10.1093/icesjms/fsu198>.
- Method, R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fish. Res.* 142, 86–99. <https://doi.org/10.1016/j.fishres.2012.10.012>.
- Orbesen, E.S., Snodgrass, D., Hoolihan, J.P. and Prince, E.D., 2011. Updated US Conventional Tagging Data Base for Atlantic Blue Marlin (1955-2008), with Comments on Potential Stock Structure. *Collect. Vol. Sci. Pap. ICCAT*, 66(4), pp.1685-1691.
- Shimose, T., Yokawa, K. and Tachihara, K., 2015. Age determination and growth estimation from otolith micro-increments and fin spine sections of blue marlin (*Makaira nigricans*) in the western North Pacific. *Marine and Freshwater Research*, 66(12), pp.1116-1127.
- Wilson, C. A., J. M. Dean, E. D. Prince, and D. W. Lee. 1991. An examination of sexual dimorphism in Atlantic and Pacific Blue Marlin using body weight, sagittae weight, and age estimates. *Journal of Experimental Marine Biology and Ecology* 151:209–225.

Table 1. Estimated parameters, values and standard deviations of the Atlantic Blue Marlin Stock Synthesis model_1.2.

Label	Value	Phase	Min	Max	Init	Status	Parm_StDev	Pr_type	Prior	Pr_SD
L_at_Amax_Fem_GP_1	301.182	2	241	361	301.171	OK	4.271	Normal	290.238	9.740
VonBert_K_Fem_GP_1	0.144	4	0.115	0.173	0.144	OK	0.026	Normal	0.114	0.050
Richards_Fem_GP_1	2.128	4	1.704	2.557	2.131	OK	0.612	Normal	2.909	1.370
L_at_Amax_Mal_GP_1	210.729	2	168.6	252.8	210.707	OK	1.815	Normal	202.748	2.371
VonBert_K_Mal_GP_1	0.425	4	0.34	0.51	0.425	OK	0.028	Normal	0.535	0.106
SR_LN(R0)	4.951	1	4	6	4.955	OK	0.041	No_prior	NA	NA
Q_extraSD_US_LL_6(6)	-0.029	4	-2	2	-0.029	OK	0.040	No_prior	NA	NA
Q_extraSD_Ven_LL_8(8)	-0.186	4	-2	2	-0.186	OK	0.038	No_prior	NA	NA
Q_extraSD_Ven_Art_9(9)	-0.161	4	-2	2	-0.161	OK	0.070	No_prior	NA	NA
Q_extraSD_Ven_RR_10(10)	0.141	4	-2	2	0.141	OK	0.061	No_prior	NA	NA
LnQ_base_Chi-Tip_early_11(11)	-11.402	1	-25	25	-11.459	OK	0.127	No_prior	NA	NA
Q_extraSD_Chi-Tip_mid_12(12)	-0.051	4	-2	2	-0.051	OK	0.069	No_prior	NA	NA
Q_extraSD_Chi-Tip_late_13(13)	0.157	4	-2	2	0.158	OK	0.072	No_prior	NA	NA
Q_extraSD_Brazil_LL_14(14)	0.187	4	-2	2	0.187	OK	0.072	No_prior	NA	NA
Q_extraSD_JPN_LL_15(15)	-0.001	4	-2	2	-0.001	OK	0.044	No_prior	NA	NA
Q_extraSD_Ghana_Gill_16(16)	0.548	4	-2	2	0.548	OK	0.210	No_prior	NA	NA
LnQ_base_Japan_00_17(17)	-11.018	1	-25	25	-11.226	OK	0.083	No_prior	NA	NA
LnQ_base_Chi-Tip_early_11(11)_ENV_mult	-0.286	1	-1	1	-0.293	OK	0.034	No_prior	NA	NA
LnQ_base_Japan_00_17(17)_ENV_mult	-0.089	1	-1	1	-0.119	OK	0.014	No_prior	NA	NA
Size_DblN_peak_Art_Gillnet_1(1)	223.598	2	178.89	268.33	223.607	OK	3.087	No_prior	NA	NA
Size_DblN_top_logit_Art_Gillnet_1(1)	-11.720	3	-14.06	-9.38	-11.720	OK	2.198	Normal	-11.720	2.200
Size_DblN_ascend_se_Art_Gillnet_1(1)	7.576	4	6.06	9.09	7.576	OK	0.108	No_prior	NA	NA
Size_DblN_end_logit_Art_Gillnet_1(1)	10.171	2	8.14	12.21	10.171	OK	1.977	Normal	10.150	2.000
Size_DblN_peak_LongLine_2(2)	87.458	2	69.97	104.95	87.458	OK	16.272	Normal	91.264	18.000
Size_DblN_ascend_se_LongLine_2(2)	10.583	3	8.46	12.7	10.581	OK	1.971	Normal	10.555	2.000
Size_DblN_peak_RR_4(4)	242.053	2	194.08	291.12	242.603	OK	9.314	Normal	220.000	44.000
Size_DblN_top_logit_RR_4(4)	0.000	3	-1	1	0.000	OK	0.020	Normal	0.000	0.020
Size_DblN_ascend_se_RR_4(4)	8.701	4	6.97	10.45	8.707	OK	0.259	Normal	8.220	1.600
Size_DblN_descend_se_RR_4(4)	2.000	5	1.6	2.4	2.000	OK	3.651	Normal	2.000	4.000
Retain_L_width_RR_4(4)_BLK1repl_1987	32.463	4	26.87	40.31	33.591	OK	4.793	Normal	33.591	60.000
Retain_L_width_RR_4(4)_BLK1repl_1994	4.780	4	3.85	5.78	4.813	OK	0.711	Normal	4.813	60.000
Retain_L_width_RR_4(4)_BLK1repl_1999	11.161	4	8.95	13.42	11.184	OK	0.588	Normal	11.184	90.000
Retain_L_asymptote_logit_RR_4(4)_BLK2repl_1999	-0.038	6	-0.05	-0.03	-0.038	OK	0.008	Normal	-0.038	0.008
Retain_L_asymptote_logit_RR_4(4)_BLK2repl_2005	-3.429	6	-4.12	-2.74	-3.429	OK	0.678	Normal	-3.429	0.686

Table 2. Summary statistics runs tests, retrospective analysis, and hindcast cross-validation (HCxval) model diagnostics applied to the Atlantic Blue Marlin Stock Synthesis model_1.2.

Diagnostic	Index	Type	Statistic	Value	Result
Runs Test	US_LL_6	cpue	p-value	0.033	Failed
Runs Test	Ven_LL_8	cpue	p-value	0.228	Passed
Runs Test	Ven_Art_9	cpue	p-value	0.000	Failed
Runs Test	Ven_RR_10	cpue	p-value	0.006	Failed
Runs Test	Chi-Tip_early_11	cpue	p-value	0.100	Passed
Runs Test	Chi-Tip_mid_12	cpue	p-value	0.012	Failed
Runs Test	Chi-Tip_late_13	cpue	p-value	0.022	Failed
Runs Test	Brazil_LL_14	cpue	p-value	0.028	Failed
Runs Test	JPN_LL_15	cpue	p-value	0.048	Failed
Runs Test	Ghana_Gill_16	cpue	p-value	0.090	Passed
Runs Test	Japan_00_17	cpue	p-value	0.000	Failed
Retrospective		SSB	Hurtado Rho	0.130	Passed
Retrospective		Rec	Hurtado Rho	0.186	Passed
Retrospective		F	Hurtado Rho	-0.149	Passed
Retrospective		Bratio	Hurtado Rho	0.130	Passed
Hcxval	US_LL_6	cpue	MASE	1.124	
Hcxval	Ven_LL_8	cpue	MASE	0.478	Passed
Hcxval	Ven_Art_9	cpue	MASE	1.558	Failed
Hcxval	Chi-Tip_late_13	cpue	MASE	2.154	Failed
Hcxval	JPN_LL_15	cpue	MASE	1.213	

Table 3. Derived quantities from for the Atlantic Blue Marlin SS3 model_1.2.

Quantity	Value	SD	CV
F / FMSY 2022	0.776	0.118	15.2%
SSB / SSBMSY 2023	0.762	0.107	14.1%
SSB_unfished	59,773	2,465	4.1%
Totbio_unfished	98,985	3,396	3.4%
SmryBio_unfished	93,981	3,194	3.4%
Recr_unfished	141.297	5.82604	4.1%
SSB_MS Y	20,666	873	4.2%
SPR_MS Y	0.509	0.001	0.1%
annF_MS Y	0.078	0.000	0.3%
Dead_Catch_MS Y	2,899	99	3.4%
Ret_Catch_MS Y	2,899	99	3.4%
B_MS Y/SSB_unfished	0.346	0.001	0.3%

Table 4. Model names, data used fixed values and estimated values (in italics) for the eight model configurations considered.

Model	Growth	Catch	M	Steepness
1 *	New + old otolith *	Landings + reported dead discards *	0.148*	0.5*
2	New + old otolith	Landings + reported dead discards	0.148	0.4
3	New + old otolith	Landings + reported dead discards	0.148	0.7
4	New + old otolith	Landings + reported dead discards	0.148	<i>0.457</i>
5	New + old otolith	Landings + reported dead discards	<i>0.115</i>	0.4
6	New + old otolith	Landings + reported dead discards	<i>0.107</i>	0.5
7	New + old otolith	Landings + reported dead discards	<i>0.095</i>	0.7
8	New + old otolith	Landings + reported dead discards	<i>0.098</i>	<i>0.627</i>

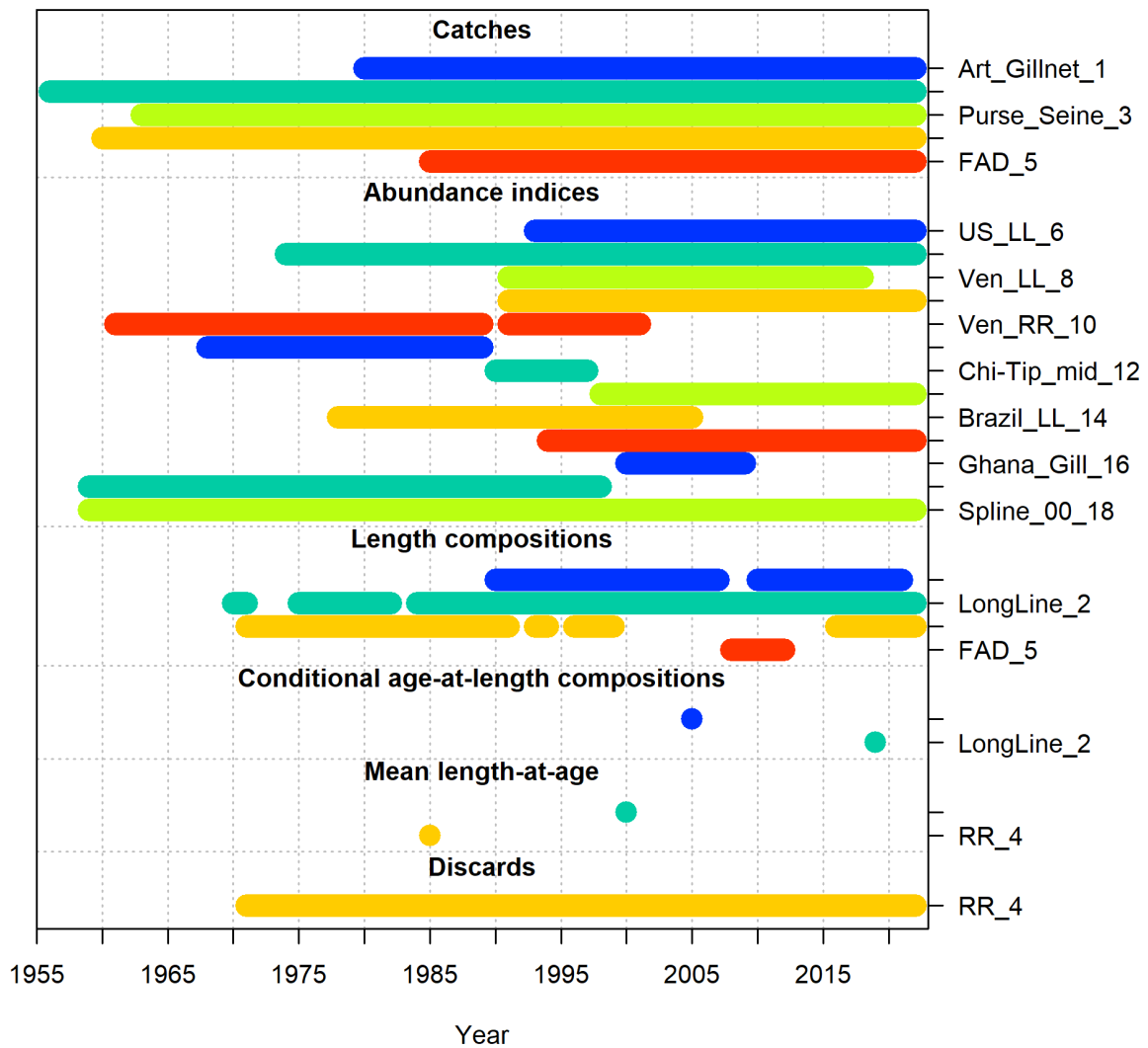


Figure 1. Observational data sources by year used in the Stock Synthesis models for Atlantic Blue Marlin, 2024.

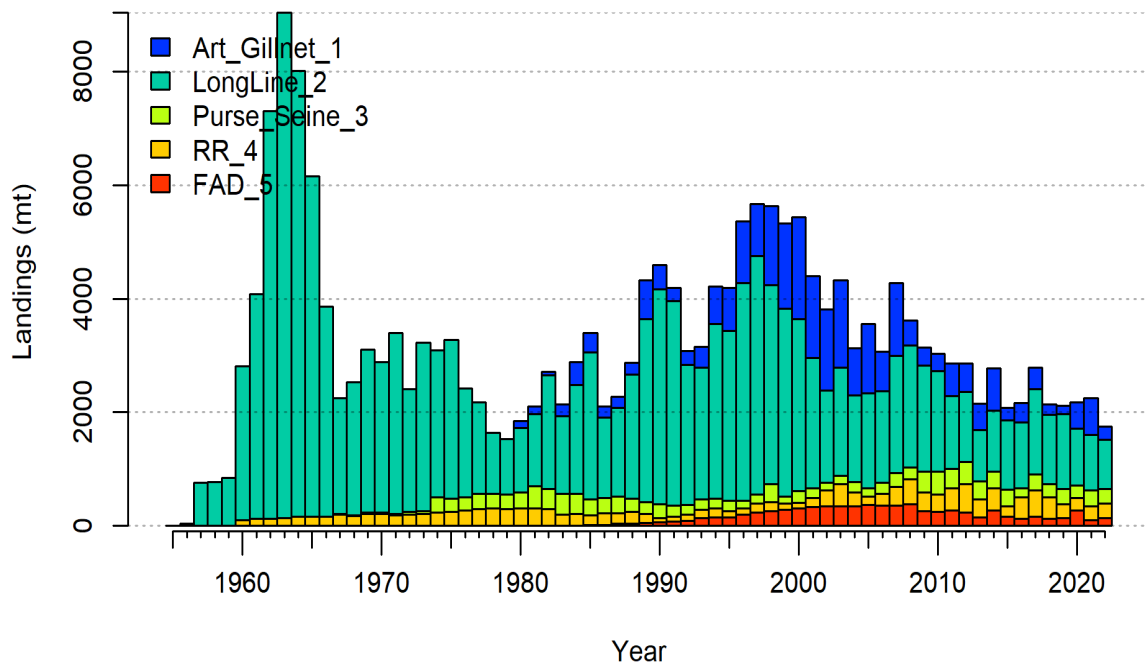


Figure 2. Total landings of Atlantic Blue Marlin year and gear, 1950-2022.

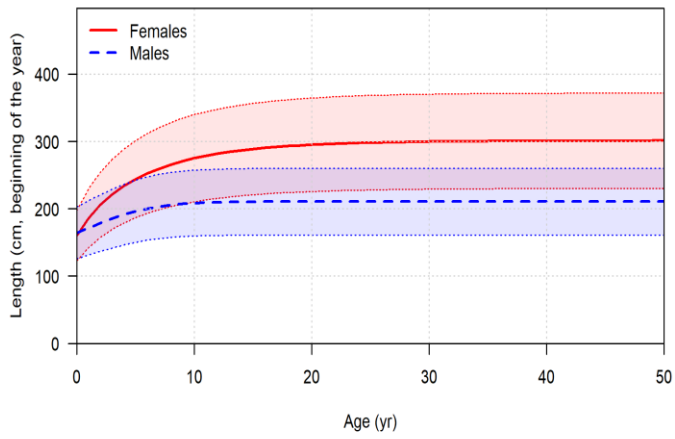
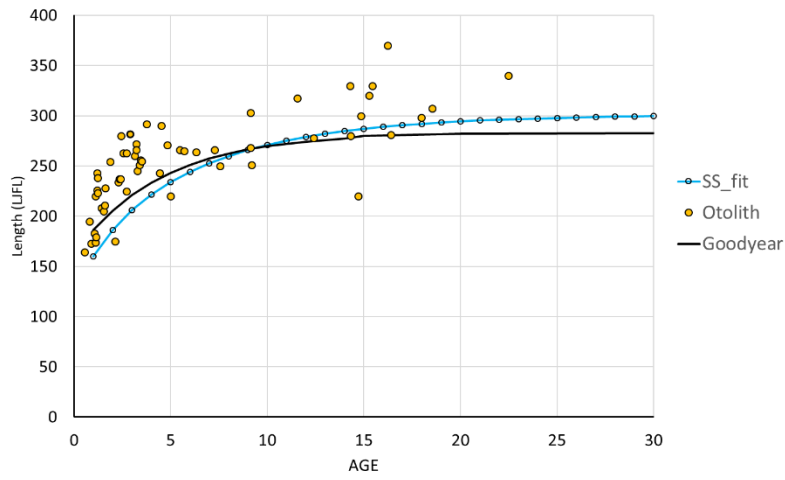


Figure 3. (Top) SS fit to female growth information, and (bottom) resulting sex specific growth curve for Atlantic Blue Marlin.

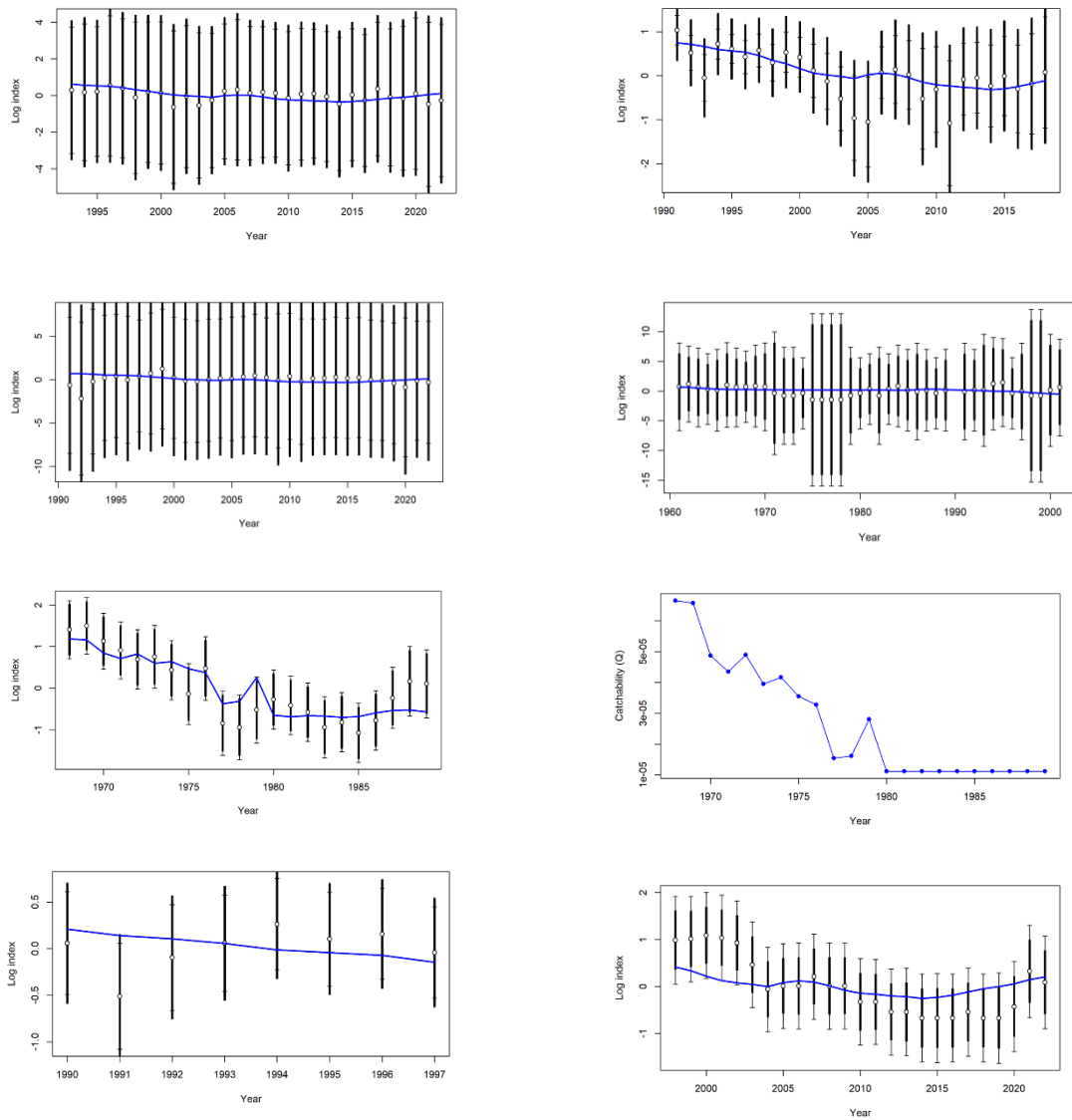


Figure 4. Fit to Atlantic Blue Marlin (log) indices of abundance for the SS3 model. From left to right US_LL, Venezuelan_LL, Venezuelan_ART, Venezuelan_RR, Chinese Taipei_LL-EARLY and Chinese Taipei_LL-EARLY catchability, Chinese Taipei_LL-MID and Chinese Taipei_LL-LATE.

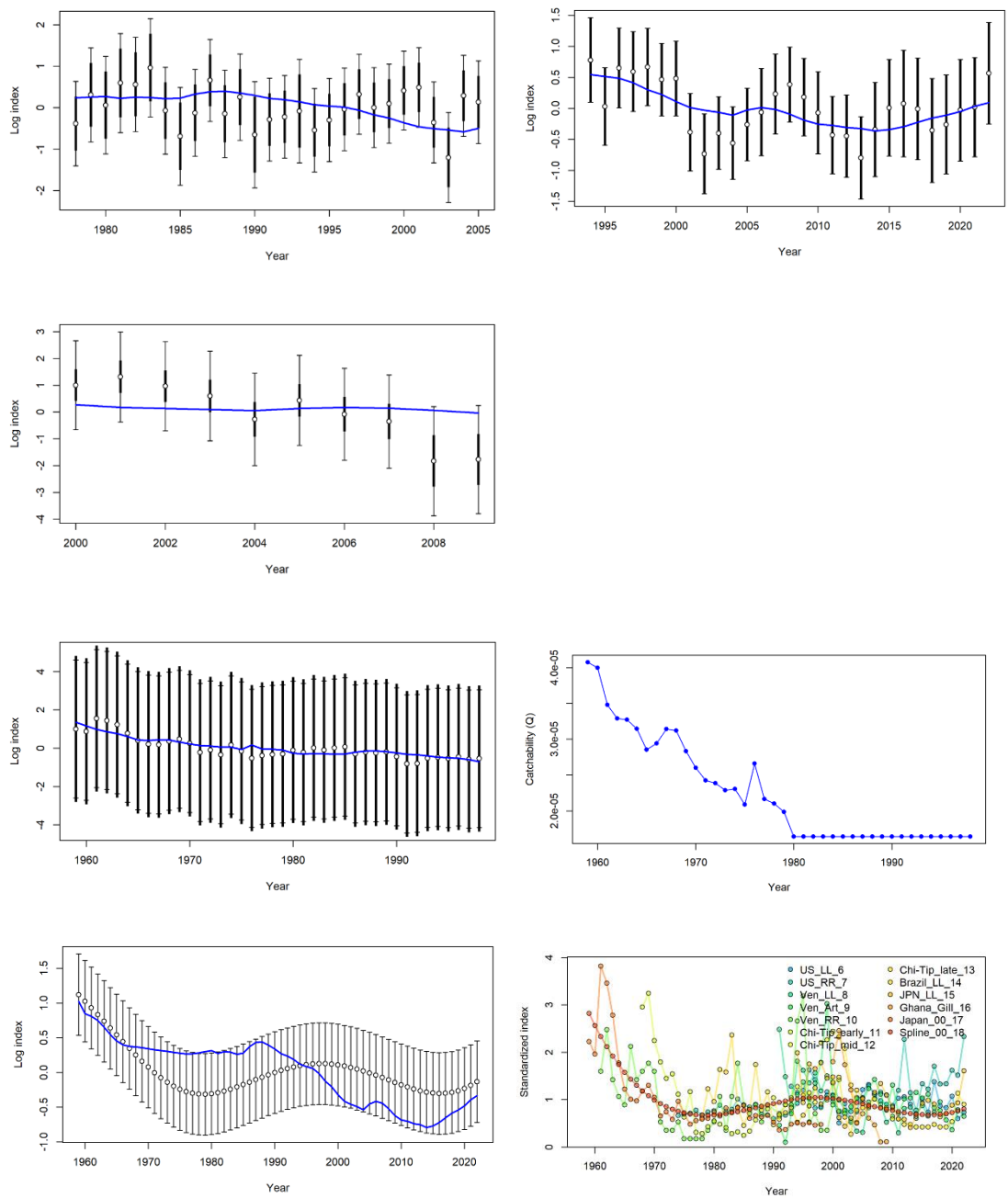


Figure 5. Fit to Atlantic Blue Marlin indices of abundance for the SS3 model. From left to right Brazilian_LL, Japan_LL, Ghana_GN, Japan_LL historical, Japan_LL historical catchability, spline fit to all CPUE (given zero weight), and standardized indices overlaid.

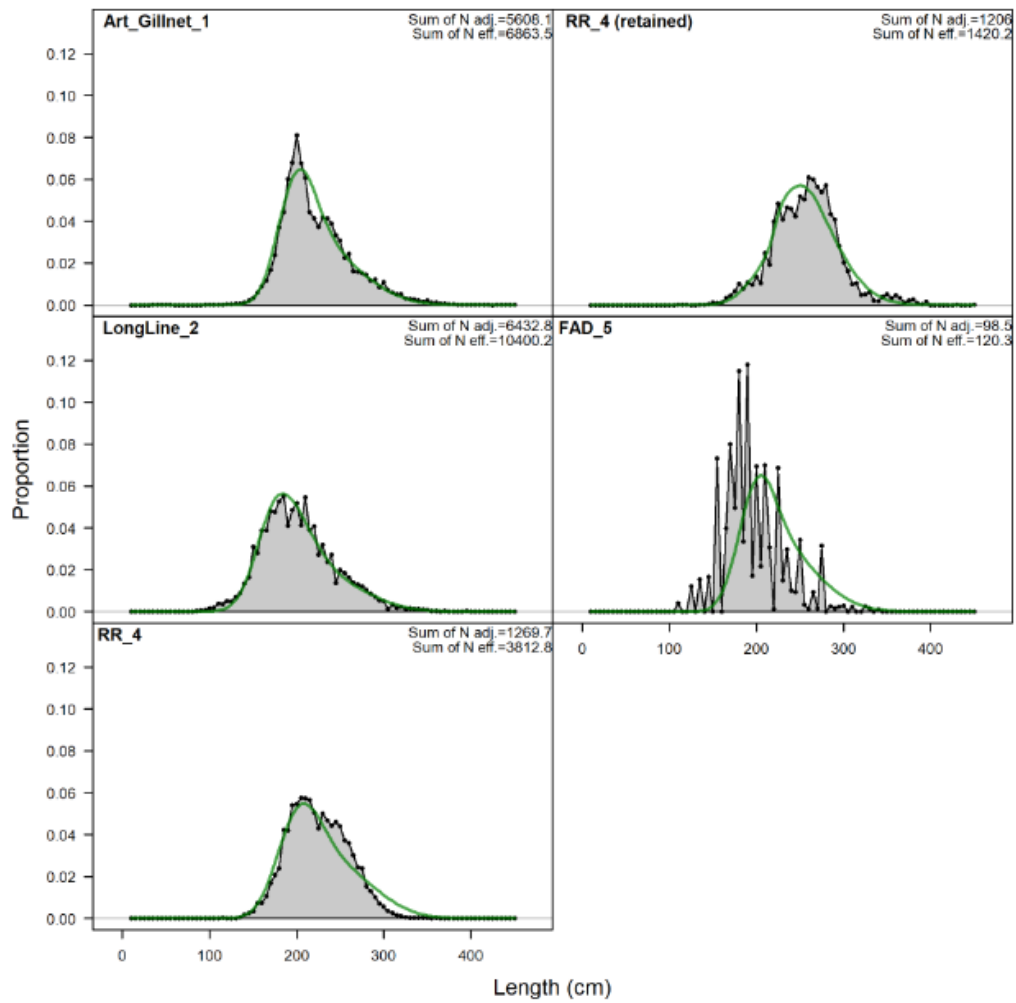


Figure 6. Length frequencies for Atlantic Blue Marlin and SS3 model fit by fishery across all years.

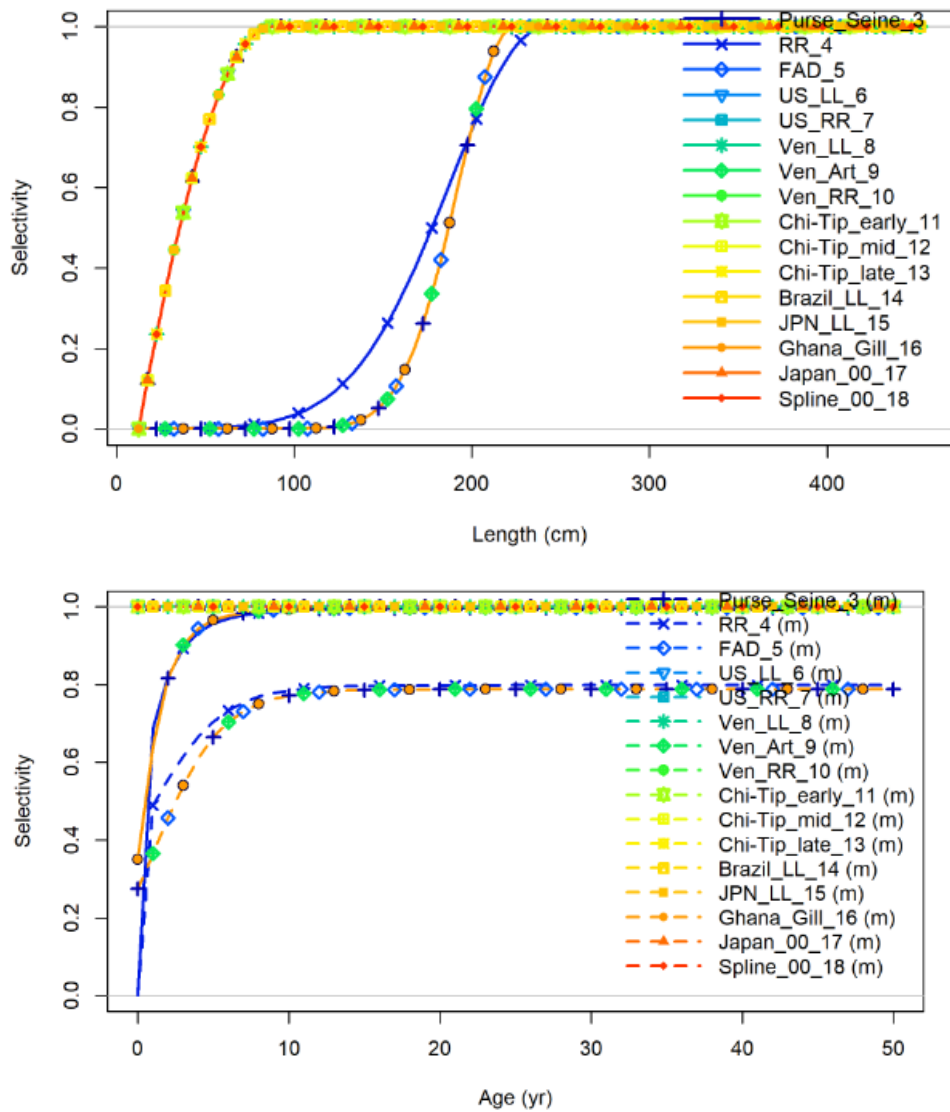


Figure 7. Estimated ending year (top) length and (bottom) age-length selectivities by fleet for Atlantic blue marlin from the SS3 model.

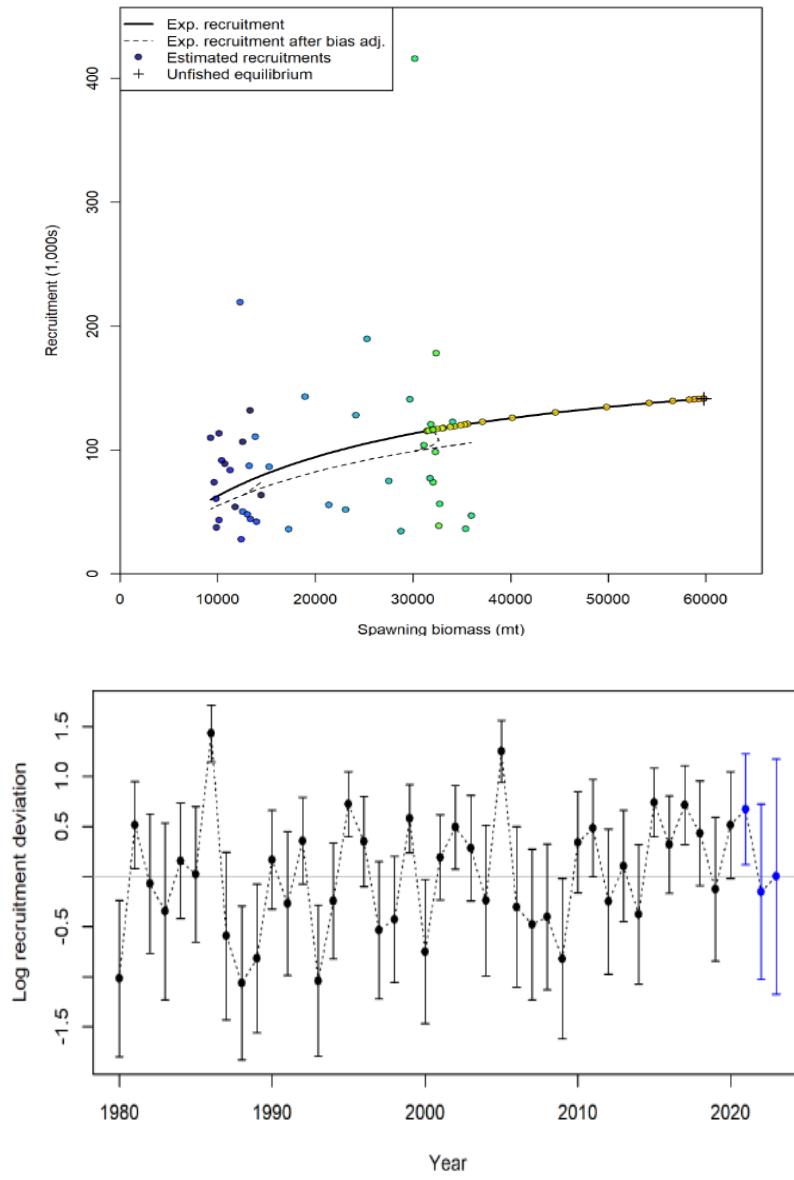


Figure 8. Estimated (top) stock-recruitment function and (bottom) recruitment deviations from SS3 model for Atlantic Blue Marlin

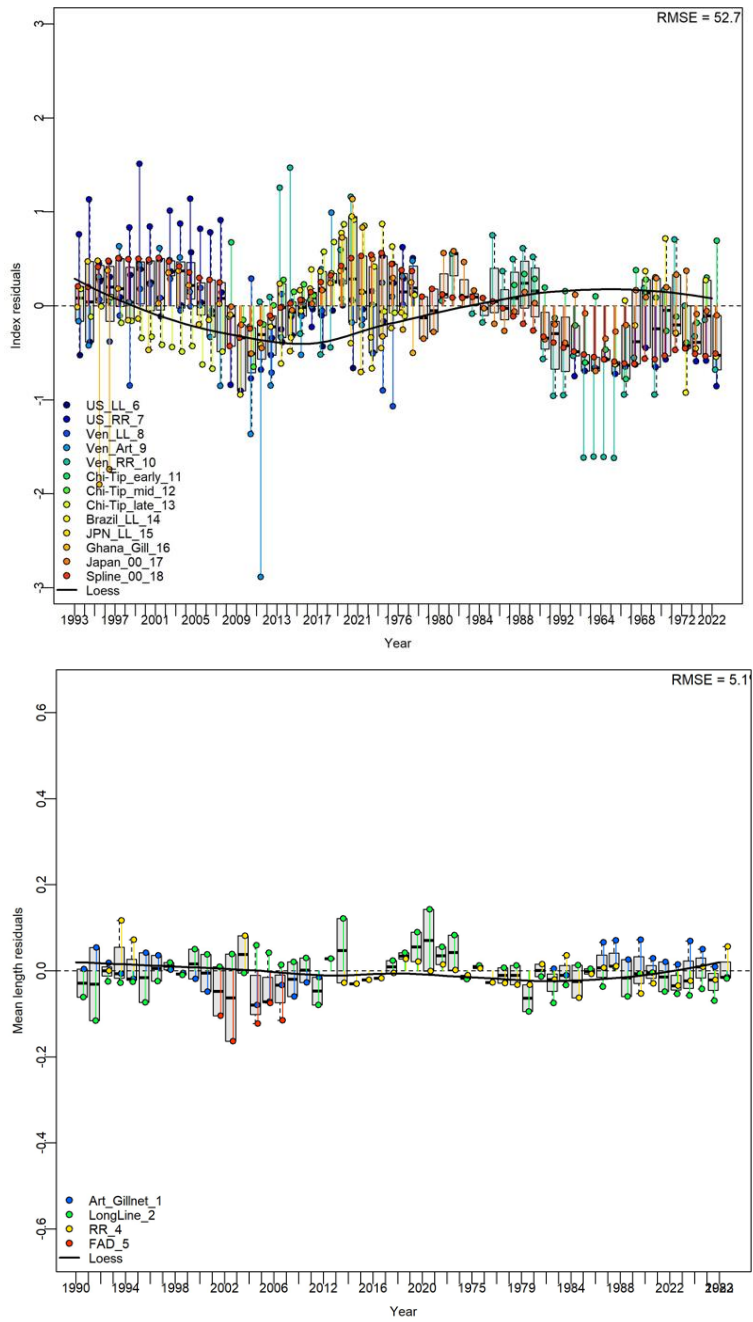


Figure 9. Joint residual plots for (top) multiple CPUE fits from different fishing fleets, (bottom) (c) annual mean length estimates for multiple fishing fleets from the Atlantic Blue Marlin.

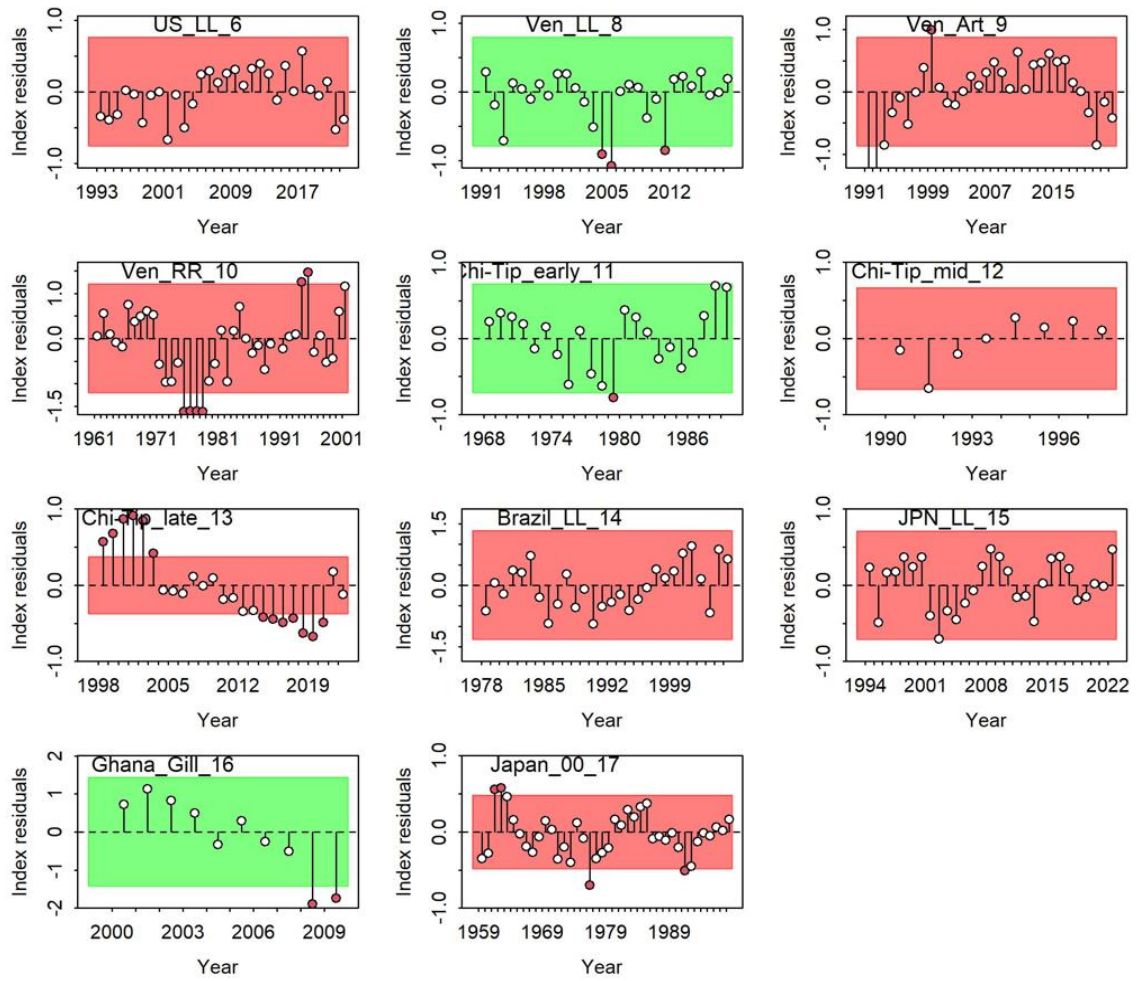
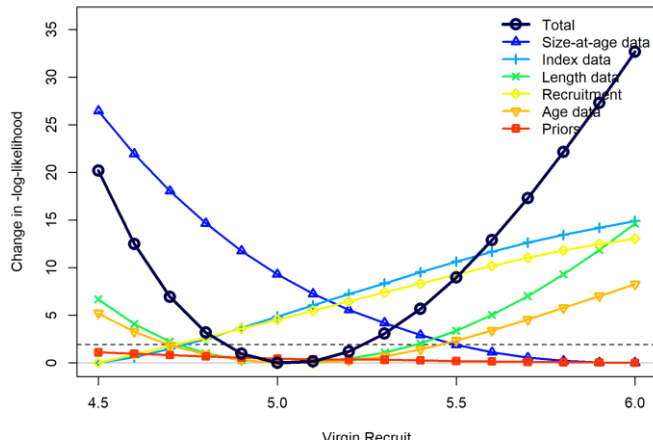
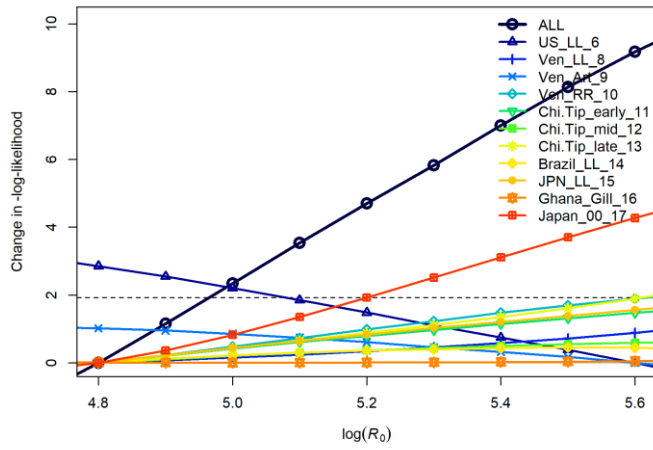


Figure 10. Runs test results illustrated for the indices of abundance for the SS model for North Atlantic swordfish. SPN_1 and Combined_CPUE were not used in the SS3 model.



Changes in CPUE by fleet



Changes in Length by fleet

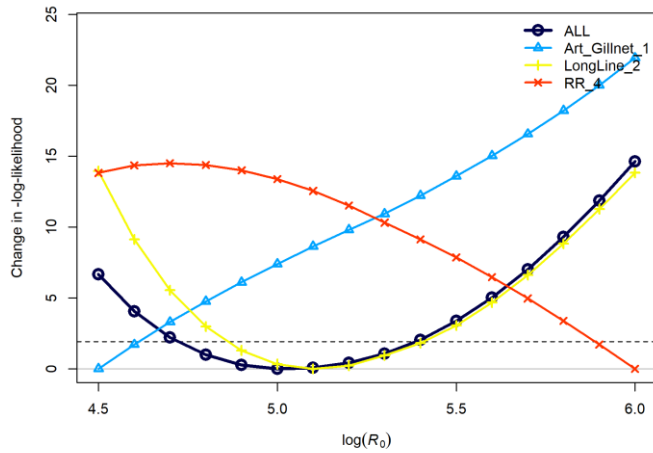


Figure 11. Log-likelihood profiles for R_0 for the various data components included in the North Atlantic Blue Marlin Stock Synthesis models, showing the contribution of (top) all data likelihood components, (center) among abundance indices and (bottom) among length composition data.

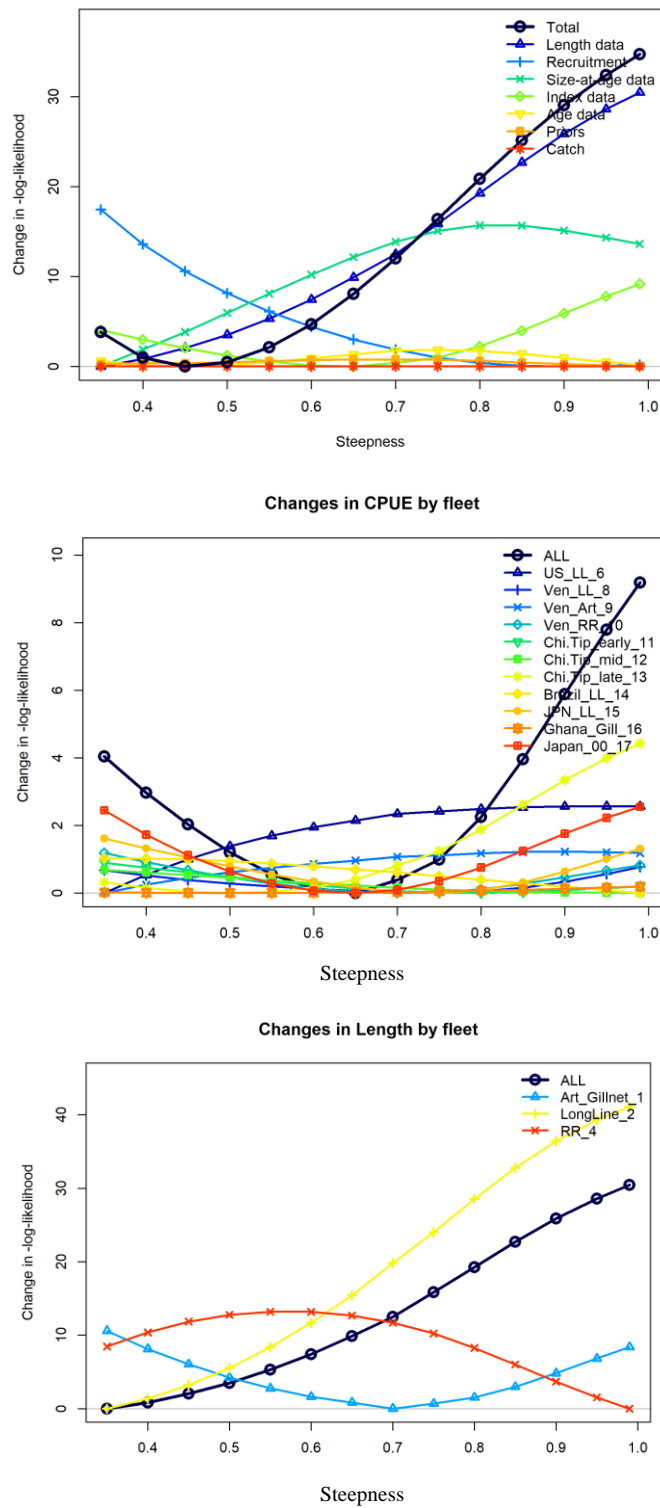


Figure 12. Log-likelihood profiles for steepness for the various data components included in the Atlantic Blue Marlin Stock Synthesis models, showing the contribution of (top) all data likelihood components, (center) among abundance indices and (bottom) among length.

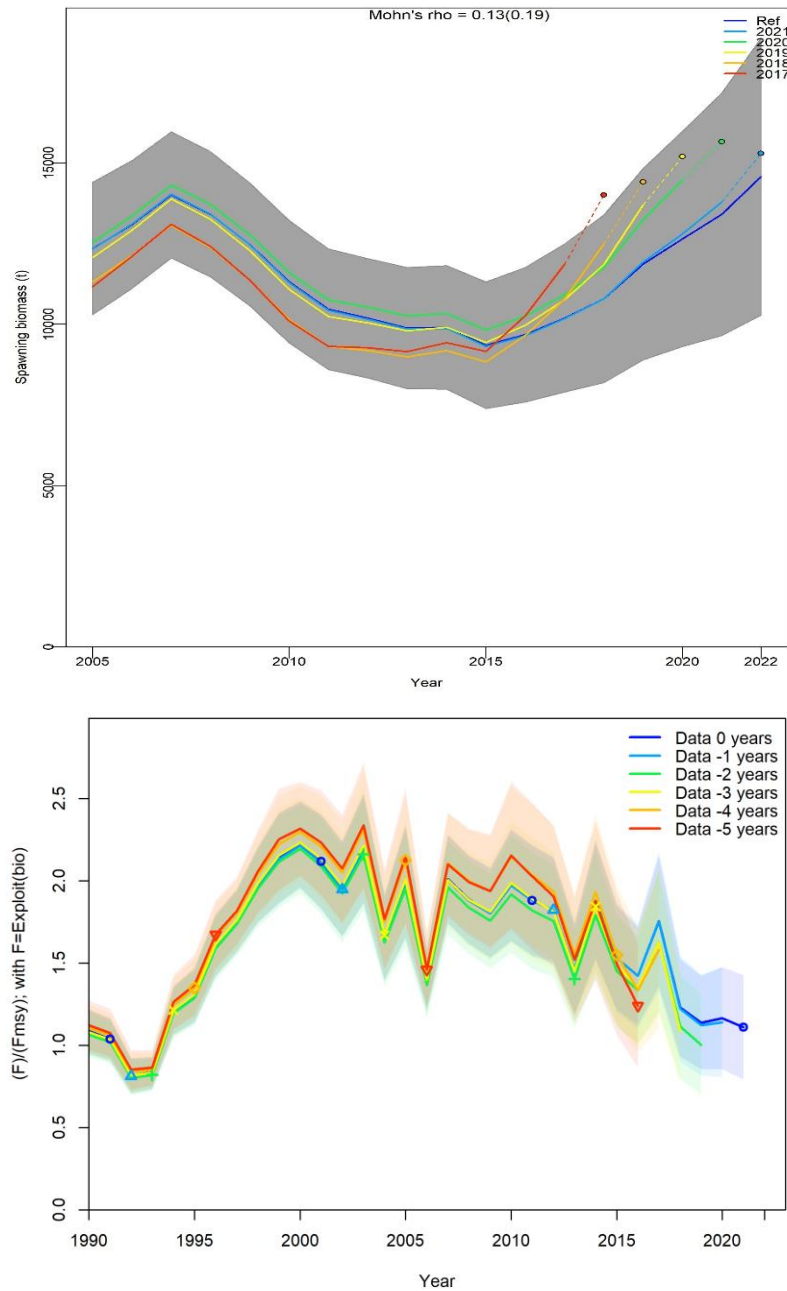


Figure 13. Retrospective analysis of spawning stock biomass (SSB) estimates for Atlantic Blue Marlin SS3 models conducted by re-fitting the reference model (Ref) after removing five years of observations, one year at a time sequentially. Mohn's rho statistic and the corresponding 'hindcast rho' values (in brackets) are printed at the top of the panels. One-year-ahead projections denoted by color-coded dashed lines with terminal points are shown for each model. Grey shaded areas are the 95 % confidence intervals from the reference model

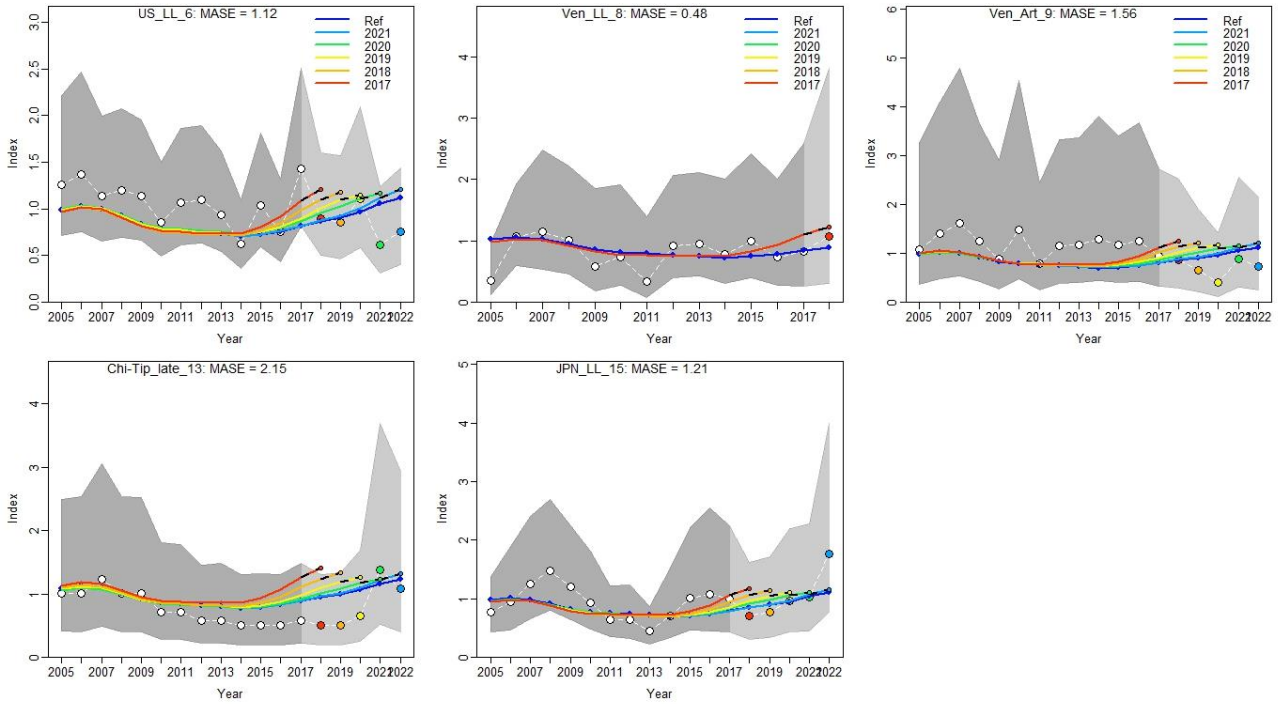


Figure 14. Hindcasting cross-validation (HCxval) results for catch-per-unit-effort (CPUE) fits from the Atlantic Blue Marlin SS3 model, showing observed (large points connected with dashed line), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (Ref) and five hindcast model runs (solid lines) relative to the expected catch-per-unit effort (CPUE). The observations used for cross validation are highlighted as color-coded solid circles with associated 95 % confidence intervals (light-gray shading). The model reference year refers to the endpoints of each one-year-ahead forecast and the corresponding observation (i.e., year of peel + 1). The mean absolute scaled error (MASE) score associated with each CPUE time series is denoted in each panel.

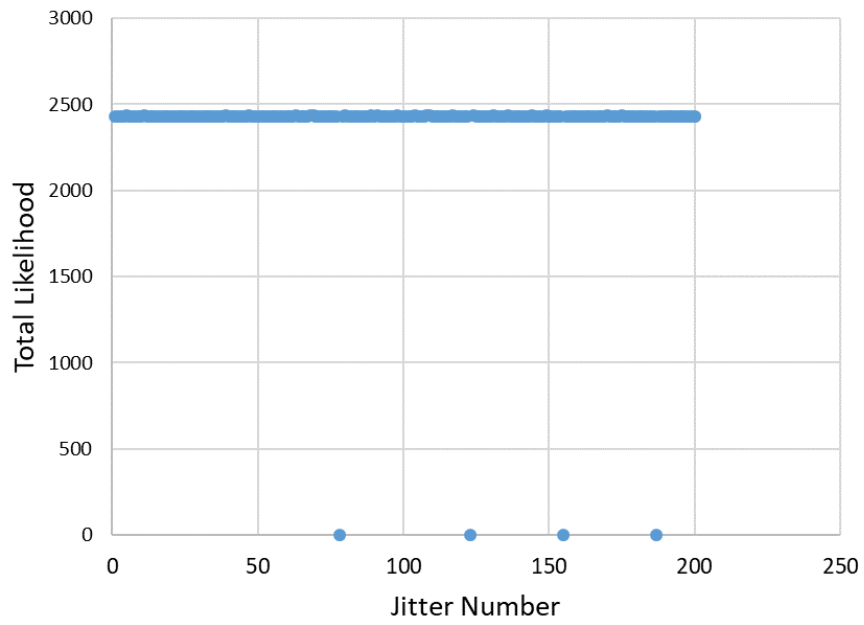


Figure 15. The jitter diagnostic for global convergence conducted for the Atlantic Blue Marlin SS3 models.

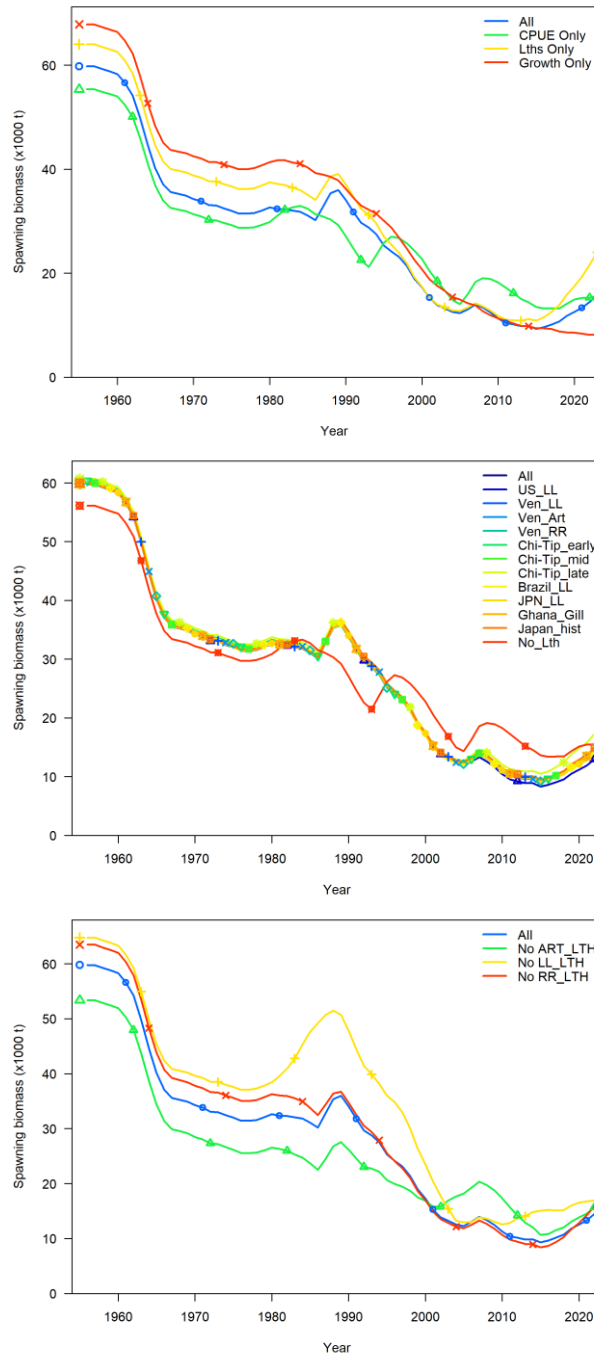


Figure 16. Jack-knife analysis (i.e. removal of one fleet at a time) of (top) data group; (middle) CPUE; and (bottom) length composition data for Atlantic Blue Marlin using the SS3 model.

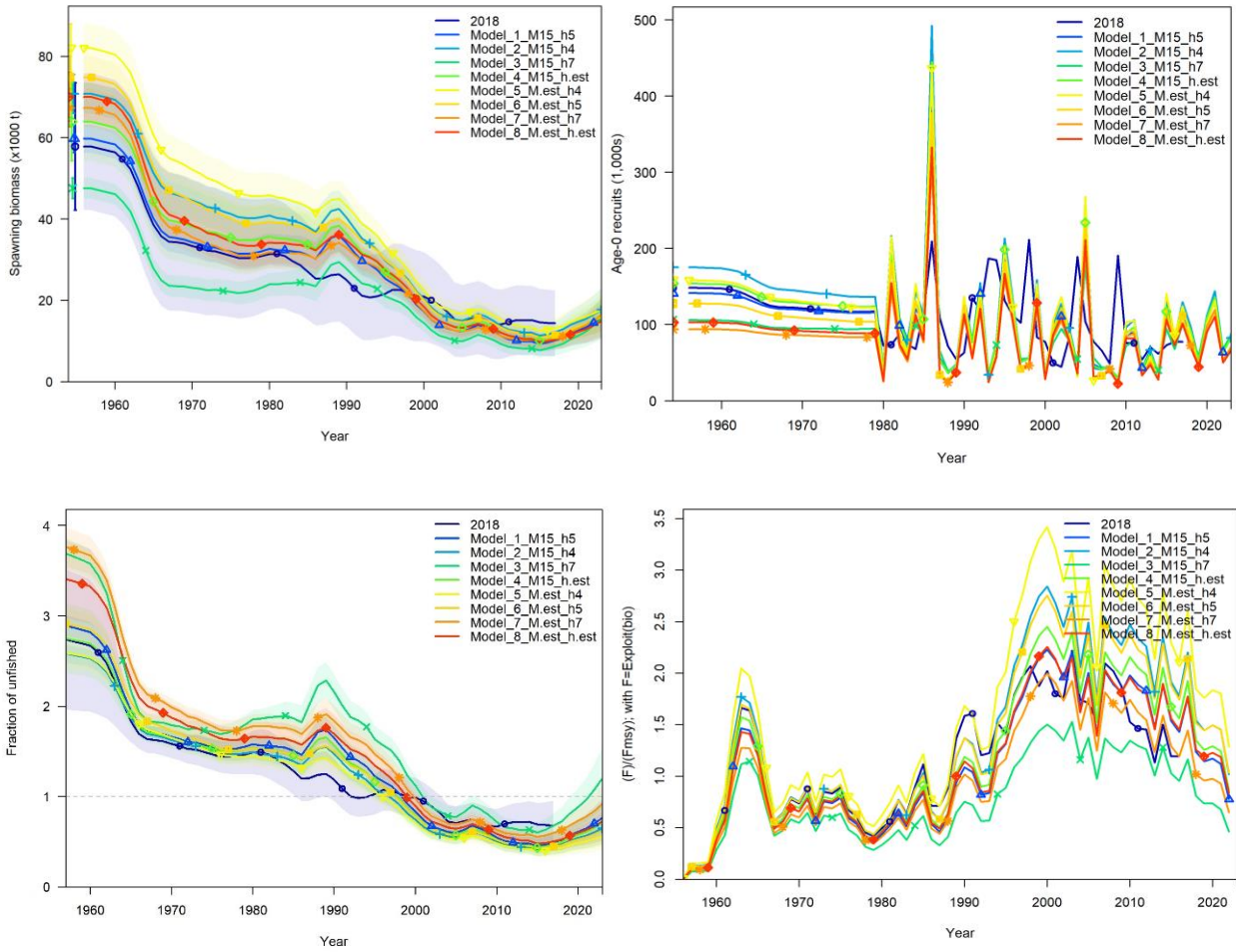


Figure 17. Model uncertainty as expressed via different assumptions for steepness and natural mortality for Atlantic Blue Marlin SS3 model (top) spawning stock biomass; (bottom) F/F_{msy} .

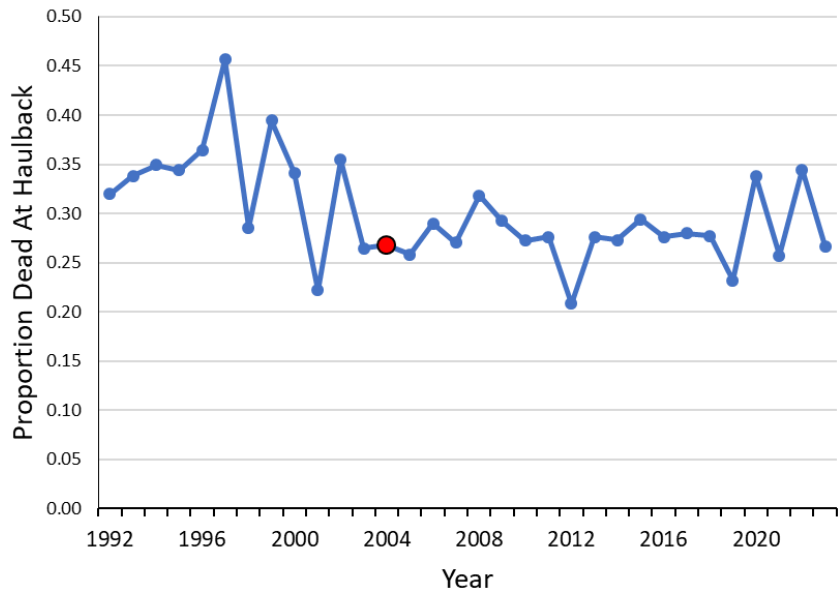


Figure 18. Proportion of Blue Marlin dead-at-haulback from the US longline fishery 1992-2023. Red dot in 2004 denotes the year circle hooks began to be used.