SCRS/2019/110 Collect. Vol. Sci. Pap. ICCAT, 76(4): 265-292 (2020)

CURRENT STATUS OF THE WHITE MARLIN (*KAJIKIA ALBIDA*) STOCK IN THE ATLANTIC OCEAN 2019: PREDECISIONAL STOCK ASSESSEMENT MODEL

Michael Schirripa[1]

SUMMARY

*Pre-decisional stock assessment configurations, diagnostics and results are described for the 2019 fully integrated assessment model for Atlantic white marlin (*Kajikia albida*). Three alternative models were studied, each with progressively more complexity. Diagnostics included profile analysis, run tests on CPUE fits, examination of residual trends, and retrospective analysis. Of the three models considered Model_3 (estimated catch multiplier and variance reweighting used on CPUEs) performed the best with regard to diagnostics. Estimates of maximum sustainable ranged from 1355 t – 1397 t. Estimates of F/Fmsy for 2017 ranged from 0.768 to 0.990. Estimates of SSB/SSBmsy for 2017 ranged from 0.411 to 0.512. All three models indicated that the stock is overfished but that overfishing is not occurring.*

RÉSUMÉ

Les configurations, les diagnostics et les résultats de l'évaluation des stocks avant la prise de décision sont décrits pour le modèle d'évaluation entièrement intégré du makaire blanc de l'Atlantique (Kajikia albida) *de 2019. Trois modèles alternatifs ont été étudiés, chacun de plus en plus complexe. Les diagnostics comprenaient une analyse de profil, des tests sur les ajustements de CPUE, l'examen des tendances résiduelles et une analyse rétrospective. Sur les trois modèles considérés, le modèle_3 (multiplicateur de capture estimé et repondération de la variance utilisée sur les CPUE) a donné les meilleurs résultats en ce qui concerne les diagnostics. Les estimations de la production maximale équilibrée allaient de 1.355 t à 1.397 t. Les estimations de* F/F_{PME} *pour 2017 allaient de 0,768 à 0,990. Les estimations de la* SSB/SSB_{PME} *pour 2017 variaient de 0,411 à 0,512. Les trois modèles indiquaient que le stock est surexploité, mais qu'il n'y a pas de surpêche.*

RESUMEN

*Se describen las configuraciones, diagnósticos y resultados de la evaluación predecisiva de stock para el modelo de evaluación totalmente integrado de aguja blanca del Atlántico (*Kajikia albida*) de 2019. Se estudiaron tres modelos alternativos, cada uno con una complejidad cada vez mayor. El diagnóstico incluyó el análisis del perfil, la realización de pruebas de ajustes de CPUE, el examen de tendencias residuales y el análisis retrospectivo. De los tres modelos considerados, el Modelo_3 (multiplicador de captura estimado y reponderación de la varianza utilizada en las CPUE) fue el que mejor funcionó con respecto al diagnóstico. Las estimaciones de la sostenibilidad máxima oscilaron entre 1355 t y 1397 t. Las estimaciones de* F/F_{RMS} *para 2017 oscilaron entre 0,768 y 0,990. Las estimaciones de la* SSB/SSB_{RMS} *para 2017 oscilaron entre 0,411 y 0,512. Los tres modelos indicaron que el stock está sobrepescado pero que no se está produciendo sobrepesca.*

KEYWORDS

Stock Synthesis, white 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

Atlantic white marlin (Kajikia albida) are part of the overall ICCAT billfish species group. It was last assessed in 2012 with a terminal year of 2010. Estimates of the status of the stock were derived and reported from two different model types, a stock production model (ASPIC) and a fully integrated model (SS, Stock Synthesis). From this assessment (anonymous 2012), using the SS model, the stock was declared to be overfished (B/Bmsy = 0.32 [0.23-0.41]) with overfishing "not likely" (F/Fmsy = 0.72 [0.51-0.93]).

In 2012, the Commission implemented Rec. 12-04, intended to reduce the total harvest to 400 t in 2013, 2014, and 2015 to allow the rebuilding of the white marlin stock from the overfished condition. In 2015, the Commission extended the 400 t annual catch limit to 2016, 2017, 2018 (Rec. 15-05), and 2019 (Rec. 18-04). This value of annual catch was based on the 2012 projection that showed that in order to keep the stock from experiencing overfishing with at least a 50% probability the 2013-2022 catch would need to remain at no greater that 400 t. The Committee noted in 2018 that if catches continue to exceed this TAC, as was the case for 2015 and 2016, the rebuilding of the stock will proceed more slowly. Landings of white marlin/spearfish have exceeded 400 t for all years 2013-2017 (**Table 1**). This report is a predecisional update of the 2012 stock assessment.

2. Methods

With a few notable exceptions, the methods and model configuration used for this assessment intentionally follow closely to those used in the 2012 assessment. Layout of the data available is shown in **Figure 1**.

Growth. Life history information for white marlin remains scarce and very little new information has been published since 2012. Growth estimates were derived from Drew et al. (2006a), Drew et al. (2006b), Drew et al. (2007), and Die et al. (2008). The results of these works indicated that rings had been lost in older fish, and that ring radii differed between males and females. The fitted growth model for white marlin also showed sexual dimorphism in growth parameters. Very few of the sampled fish were age one or two, suggesting that the majority of the fish available to the longline and artisanal fleets in our sample locations are older than two years of age. It should be noted that some members of the Group felt that the L-infinity for females ($L\infty$ = 172 cm LJFL) was small, especially in relation to the Length at 50% maturity (L50 = 160 cm LJFL). This may be explained by the fast early life stage growth of white marlin that can reach $L\infty$ quite early but still be relatively young. Following these results the SS model was configured as a two sex model.

Reproduction. Estimates of maturity and reproductive aspects were taken from the ICCAT manual, which was supported by work reported by Oliveira (2007) and Arocha. Spawning biomass was made proportional to weight and length at 50% was L50 = 160 cm.

Natural mortality. As with most all highly migratory fish species, natural mortality estimates from actual data are rare and mostly based on oldest observed ages from growth studies. However, these studies are conducted the population well after fishing has occurred, which decreases the probability of observing the naturally occurring oldest age fish. The estimated age of fish observed in the above-mentioned age-and-growth studies ranged 1-13 years, though most were aged 3-8 years. These assigned ages were consistent with a subset of mark-recaptured fish sampled. Mark-recapture records suggest that white marlin are capable of living 15+ years (Orbesen et al., 2008; Ortiz et al., 2003). At the 2018 blue marlin data preparatory meeting, a species thought to live longer than white marlin, were assigned a natural mortality of 0.15 based on oldest observed fish. Following these results the SS model configured with a natural mortality of M = 0.20 maximum age of 20 years.

Stock-recruitment relation. A Beverton-Holt stock-recruitment relation was used. The ICCAT 2018 blue marlin data preparatory meeting had lengthy discussions with regard to values of steepness for billfish. Based on these discussions the Group decided to use the three possible values for steepness of 0.4, 0.5 and 0.6. The lower bound was selected based on the value estimated in the last blue marlin assessment. The upper bound was based on the informed decision that white marlin are more productive than blue marlin. The ICCAT 2012 white marlin assessment estimated value of steepness of approximately 0.6. The base case model in this work allowed steepness to be estimated with a Bayesian prior of 0.50 and a standard deviation of 0.05 and a normal distribution. Models were also configured with fixed steepness of 0.40, 0.5 and 0.60 for sensitivity analysis. Examination and comparison of the annual standard deviation of recruitment and the residual mean square error of the deviations suggested a sigma-r of 0.20 was appropriate. Too large of a value of sigma-r could artificially account for the most recent decreasing trends in spawning stock biomass to be explained solely by negative recruitment deviations during the same period. Furthermore, some indices of abundance for this stock tended to have suspiciously sporadic year to year deviations, which could also have an undue influence on recruitment deviations with larger values of sigma-r.

Landings. Historic Task 1 reported landings and dead discards are shown in **Figure 2.** Landings of individual CPC 2013-2017 with catch limits are given in **Table 1**. The first regulation on white marlin, Recommendation 1997-09, took the following steps:

All Contracting Parties and non-contracting parties, entities or fishing entities

1. *Reduce, starting in 1998, blue marlin and white marlin landings by at least 25% for each species from 1996 landings, such reduction to be accomplished by the end of 1999.*
2. *Promote the voluntary release of live blue marlin and white marlin.*

In 2012, Recommendation 2012-04 the ICCAT put CPC specific quotas on white marlin:

1. *An annual limit of 2,000 t for blue marlin and 400 t for white marlin/spearfish is established for these stocks, for 2013, 2014 and 2015. This landings limit shall be implemented as follows:*

In 2015, Recommendation 2015-05, an annual limit of 400 t for white marlin/spearfish was continued for these stocks, for 2016, 2017 and 2018. Since 2013 total landings have exceeded the total allowable catch set forth in Rec. 2012-04 (**Table 1**). Dead discards are not included in the total catch limit. The under reporting of white is thought to have begun in earnest in 1998-1999, when the implementation of management regulations by the Commission began. Therefore, the Group concluded that reported catches after 1998-1999 may not include increased discards at sea particularly from longline fleets as a consequence of the implementation of management recommendations, and thus total removals in Task I statistics may be underestimated since 1998-1999. A complete history of ICCAT white marlin recommendations are given in **Appendix 1**.

Release mortality. Horodysky and Graves (2004) found that post-release survival was significantly (P<0.01) higher for white marlin caught on circle hooks (100%) than for those caught on straight-shank (“J”) hooks (65%). Kerstetter and Graves (2005) estimated post-release survival ranging from 63.0% (assuming all non-transmitting tags were evidence of mortality) to 89.5% (excluding non-transmitting tags from the analysis). Analysis of the US Pelagic Observer Program data revealed that approximately 50% of the white marlin released where released alive **(Figure 3)**. These observations do not account for mortality that may have occurred any time after release. (Graves and Horodysky, 2008). No significant differences in the incidence of deep (internal) hooking, hook-induced trauma, or post-release survival were found among fish caught on the different circle hook models. Incidences of white marlin deep hooking, hook-induced trauma, and post-release mortality were significantly lower for the three circle hook models (combined) than for J hooks evaluated in a previous study.

However, given the large sample size both spatially and temporally, a release mortality of 50% was applied to the estimated live discards from longlines, which were then added to the total catch. Recreational tournaments discarded approximately 70% of their catch 1971-1990, 80% 1991 and trended upwards to nearly 100% 2000 forward (**Figure 4**). Given the high survival of these fish discards were not considered the model.

Live and dead discards. Live and dead discards are known to be under reported to ICCAT. Discards, however, are critical observational data for any stock assessment. Although difficult to account for, the ICCAT Secretariat provided estimates of non-reported live and dead discards from longline fleet 2000-2017 through the use of extrapolation from those CPCs that did report. The estimated dead discards were added to the “catch” (which is standard protocol).

Unaccounted (IUU) landings. Along with discards, unreported landings, either intentional or by error, are also a known issue with reported landings. This under reporting can have unforeseen consequences on both the stability and outcome of the assessment model All methods of estimating IUU landings will have assumptions One was way to do this is for the purposes of this assessment, and a minimal number of assumptions, is to estimate the IUU landings within the stock assessment model itself We can use observational data and make a determination as to whether or not the observed landings are in agreement with the remaining observational data. The assumption made here is no different than those of the assessment itself, that the observational data is representative of the stock and the model is relatively stable and free from overt process error. Reported WHM landings show a decreasing trend 1998-2010, and have been relatively consistent 2010-2017 (presumably due to regulations) and well below estimated MSY However the trend in the consolidated CPUEs has also been flat.

Some possible explanations for this are; (1) B_{MSY} has been over estimated, (2) the stock has reached an equilibrium level below that of the correctly estimated MSY, (3) CPUEs are not indicative of stock abundance trends, (4) recruitment has been below average during that time period, and/or (5) landings are being under reported.

For the first in this assessment, these IUU landings were estimated within the assessment model.

The catch is adjusted by the multiplier is calculated as:

$$C_{adj} = C_{exp} * c_{mult}$$

where C_{exp} is the expected catch from the fishing mortality and c_{mult} is the catch multiplier. Information informing the estimates of these landings would be received from the possible mismatch of observed landings, indices of abundance, and estimated productivity and recruit deviations from the fully integrated assessment model. In the case of white marlin, which presumably has at least some surplus production, the observations of an overall flat population trend via the CPUEs coupled with decreasing landings suggests that landings may be being under reported. This degree of under reporting can then be estimated by the difference between those landings expected from these observations and those observed. Under reported were attributed to the longline fleet only.

Indices of abundance. It should be noted that the U.S. Recreational index for blue marlin (from the same source of tournament observations) was excluded from the final model based on the belief that the increasing trend could have been due to an increase in catchability of the fleet. This increase may have been due to several technological advancement made by the gear and boats used in the tournaments. Communications with fishermen who have participated in the white marlin tournament fishery for many years revealed that they have witnessed a great deal of change in the fishery over the years. These changes include, but not limited to, faster and larger boats that can range further, improvements in gear, and technology reducing the need for exploratory fishing (**Appendix 2**). These concerns were also expressed during the blue marlin assessment. Given the obvious changes in catchability, and to remain consistent with the blue marlin assessment and decision, the US Sport fishery CPUE was omitted from all white marlin models as well.

Also excluded from the blue marlin assessment was the recent Chinese-Taipei. This is because the steep decline in the CPUE was attributed to a lack of reporting discarded blue marlin. The same non-reporting of discarded fish also hold true for the white marlin CPUE. Given that an unknown number of fish that were caught were not included in the CPUE, and to remain consistent with the blue marlin assessment and decision, the Chinese Taipei_late CPUE was omitted from all white marlin models as well.

Each year of each CPUE was assigned a standard error as provided by during the Data Prep meeting. However, some calculated standard errors were so small as to suggest an overly precise estimation. In the cases where any year of any index had a SE less than 0.30, that year was assigned 0.30, and any years greater than 0.30 where assigned their reported values.

During some overlapping time periods the fourteen indices were not in agreement with regard to the trends in the stock size. This created instability within the assessment model due a flat response surface in the fitting. To address this issue for each of the CPUE times an additional parameter was estimated, an additive constant to be added to the input standard deviation of the survey variability (i.e. variance reweighting). The effect of variance reweighting estimates a higher than input standard deviation for those CPUE's that diverged from the overall trends suggested by the rest of the observational data. Those CPUE's that were in agreement with observational data would have a lower estimate of their input variance.

As decided at the WHM Data Preparatory meeting, the Japanese "prior" CPUE (1956-1975) was explored for model sensitivity. Exploratory work with this index indicated that keeping it in the base model was necessary for model convergence. This CPUE is the only CPUE time series for the early years of the assessment and has the contrast necessary for the model to find an adequate solution. As a result, the index was maintained for all subsequent models.

The 2018 blue marlin Group determined that the US Recreational Bill Survey index of abundance was not suitable for inclusion into the assessment. This was based on information presented during the meeting that suggested the catchability of the fleet has significantly and gradually changed over the time period of the CPUE time series. To determine if the same phenomena was true for white marlin several long-time participants of white marlin tournaments were asked for their experiences on this topic. The respondents all agreed that changes in boat speed, power, electronics and gear have increased their catchability. Their responses are given in **Appendix 1**. Given this information, its agreement with the blue marlin information, and consistency between assessments, the RBS CPUE was omitted from all analysis.

The marlin Species Group also made the determination that the third stanza (1998-2017) of the Chinese-Taipei CPUE index was inappropriate for use in the blue marlin assessment. This decision was based on the fact that during this time period, due to regulatory restrictions, discards of marlin (both blue and white) were not included in the CPUE analysis. For this reason, and to maintain consistency between assessments, the third stanza of this index was also omitted from all white marlin assessment models.

Model Sensitivity to the Japan_prior CPUE index. The data preparatory meeting Group had determined that the very early ("prior", 1956-1975) Japanese CPUE time series would be used as a sensitivity analysis. The decision to include the Japanese "prior" (1956-1975) in the base case model was based on the lack of model stability when it was excluded. When starting from the estimated parameter values, the model was able to converge in 50 iterations and invert the hessian in approximately 3 minutes with a convergence level of 0.000244. When the Japanese prior CPUE index was excluded the model was allowed to run for 2500+ iterations and still did not converge, took over 13 minutes with a "convergence" level of 29,340. It was apparent that the prior Japanese CPUE added needed contrast to the observation data in order for the model to find a solution. Based on these results, the base case model included the Japanese 1956-1975 index. Based on these results the Japanese prior CPUE was retained in all subsequent model configurations.

Selectivity. The Gillnet fishery length compositions were clearly different than those of the other three fisheries and clearly not asymptotic. Gillnet landings are primarily from Venezuela which is a fishery that operates closer to shore, an area where large white marlin are less likely to be found. The longline fishery was assumed to have asymptotic selectivity. The initial base model was configured with the recreational fishery as having asymptotic selectivity. Given that recreational fishing data is from tournament, and that tournament fishing generally fish to maximize size, this seemed a reasonable assumption. However, closer examination of the length compositions revealed that longline gear actually caught larger fish than did recreational gear. A profiling exercise on virgin recruitment was conducted on assuming an asymptotic selectivity and a dome-shaped selectivity. The model with asymptotic selectivity seemed to outperform the model with dome-shaped. For the asymptotic model, nearly every convergence level was lower and the model was able to invert the hessian on more runs. Given these diagnostics, along with knowledge of the fishery, the recreational fishery was assigned an asymptotic selectivity.

Model Configurations. Three model configuration were studied, each with progressing complexities.

Model_1 (Base_nocm): Base model without the estimation of a catch multiplier
Model_2 (Base_cm): Base model with the estimation of a catch multiplier for the longline fleet
Model_3 (Base_cmvar): Base model with estimation of a catch multiplier and variance reweighting on CPUEs

3. Results

Landings. The estimate of the catch multiplier for Model_1 was not estimated; for Model_2 was 0.835 (SD = 0.080), and for Model_3 0.870 (SD = 0.095), equaling an average of 157% under reporting of longline landings and dead discards 1998-2017. Applying this value to the fishery as a whole, the reported total landings and dead discards in 2017 of 455 t would actually have been 534 t. It should be noted that the trend in recruit deviations were estimated negative, evidence that the decrease in landings may have also partly attributable to low recruitment. So, the model fit is a trade-off between estimated/fixed parameters such as steepness and sigma-r, and so all of the parameters correlated to them.

Stock recruitment relation. The estimate of steepness for Model_1 was h = 0.55 (SD = 0.05), Model_2 was h = 0.56 (SD = 0.05) and Model_3 was h = 0.55 (SD = 0.04). The estimate of steepness was different from the prior used (h = 0.50) indicating that there was at least some information in the observational data to drive the estimate (**Figure 5**). Annual recruitment deviations are shown in **Figure 6**.

CPUE inclusion/exclusion. The examine model sensitivity to each of the CPUE's each CPUE index was removed one at a time and the trends in spawning stock biomass and B/B0 were compared (**Figure 7**). The two most influential indices for the estimation of SSB where the Brazilian Sport and the ChT_early times series. Removal of the ChT_early resulted in a recent (2002-2017) increase in SSB, as did removal of most other CPUE time series. Only the removal of the Brazilian Sport CPUE index resulted in a flat SSB trend for the same time period. The input variance and input+added are shown in **Figure 8**. The variance of the Venezuelan gillnet CPUE was increased the most, indicating that this data was in the poorest agreement with the remaining data. The variance on Japan_early, Japan_mid and Brazil_Sport had their input variances decreased, indicating an agreement with the rest of the observational data.

Model diagnostics. Several standard diagnostic techniques were investigated and/or applied to the three models to evaluate the stability of the parameter estimates and the model in general.

1. Profiling *R0* (virgin recruitment) and *h* (steepness of S/R function)
2. Run Test on CPUEs
3. Examination of fit-to-CPUE residuals
4. Retrospective analysis

Profiles on the catch multiplier. Profiles on the estimated catch multiplier showed tension between the length data and recruitment with the length data showing a minimum at high values (i.e. low values of reported catch) and the recruitment deviations (**Figure 9**). The CPUE index data also showed minimum amounts at high values. Both Model_1 and Model_2 showed similar trends. Recruitment deviations by catch multiplier are shown in **Figure 10**. The catch multiplier estimates were directly related to the estimated recruitment deviations 1998-2017. This is because the model could either account for increasing CPUEs and constant catches either by adding unreported landings, or by introducing a period of low recruitments coming into the population.

Profiles on R0 and steepness. Profile analysis for R0 by data type is shown in **Figure 11.** For Model_1 profiling on R0 revealed that the index data (CPUE) showed the most change and had a minimum LL at the highest values of *R0* (*R0* = 6.0) and that length data had a minimum LL at lowest level (*R0* = 5.0). The recruitment profile showed less change but fit beset at high levels of *R0*. Model_2 showed similar trends, however, there was a clearer indication of a minimum when examining the total LL. Model_3 showed an even clear minimum but the overall changes in LL's where less than for the other two models. The large drop in LL for Model_2 and Model_3 were investigated and it was found to be the influence of the ChT_early index.

Examination of profiling by individual CPUEs for Model_1 showed that while each of the CPUE's had a minimum LL at the highest value of R0 (**Figure 12**). The Japan_prior showed the largest change and the Cht_mid the least. Model_2 profile showed a clear minimum but still somewhat flat. Nonetheless, minimums did converge at values between the minimum and maximum. Model_3 showed a similar pattern as Model_2 with a clear minimum but a relatively flat profile.

Profiles were also examined by individual length composition data (**Figure 13**). All three models resulted in minimums between the minimum and maximum of the range centered on R0 = 5.3. The longline lengths showed a clear minimum at R0=5.3 however the Sport lengths showed a minimum at the smallest value examined. For Model_2 there were two minimums for the Sport lengths, indicating a poor signal. This is not unexpected given the small sample size for that data. Model_3 showed the relatively best profile with three minimums at values similar to those shown by the R0 profiles.

Profiles on the steepness parameter by data type are shown in **Figure 14**. While each model there was a clear minimum, there was a yet unexplainable sharp drop in the each of the data types except for parameter priors. The more smooth shape of the priors profile is likely due to the prior put on the steepness parameter. However, the models showed a different minimum, indicating that the estimate of the prior was not driven entirely by the value of the prior.

Examination of the profile by individual CPUE showed that for Model_1 and Model_2 the ChT_early hd the most influence and for Model_3 the Japan_early CPUE had the most influence (**Figure 15**). Other CPUE's had little influence on the estimate of steepness which changes in LL of less than five, but a minimum nonetheless.

Profiles on steepness by fleet specific length information showed clear minimums for all three models (**Figure 16)**. Most influential was the Sport length data and least informative was the gillnet length data. Not all profiles showed the minimums at the same values of steepness, indicating that there was information within the observational data to drive the estimate of the parameter.

Overall results of the profile testing are given in **Table 2**. Criteria for passing the test was a greater than were quantified by the likelihood components showing max change as fraction (0.01) of total change. Failure of the test indicated a flat profile while pass of the test indicated curvature to the profile.

Run Test on CPUEs. The runs test (Carvalho et al. 2017) is used to test for randomness of residuals of the model fit the CPUE time series. 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. For Model_1 and Model_2 three of the CPUE indices used in the fitting failed the runs test: Brazil sport, ChT_early, and ChT_mid (**Figure 17**). The runs test also failed for ChT_late and US sport but these indices were not used in the fitting process. Identical results were found for Model_2. Model_3, the variance reweighting model, only failed to the fit to ChT_early and ChT_mid. In this regard Model_3 passed the most number of tests.

Examination of fit-to-CPUE residuals. The examination of the CPUE residual fits showed no differences between the three models (**Figure 18**). However, Model_3 had a RMSE of 66.3% while Model_1 and Model-2 had 64.8 percent. Normally this would indicate that Model_3 had a poorer fit that the other models. But in this case a higher RMSE would be expected from Model_3 as it intentionally down weighted several of the models, resulting in a high RMSE for those CPUE's with increased variancs.

*Retrospective analysis***.** Retrospective analysis was conducted on each of the three models. To quantify and compare the results of the three analyses the Mohan's rho statistic (Carvalho 2017) was calculated on the spawning stock biomass trends using a 5 year "peel". Visual patterns of the retrospective analysis are shown in **Figure 19**. The resulting rho statistic for Model_1, Model_2 and Model_3 were 0.200, 0.131 and -0.060, respectively. The analysis followed the rule of thumb that values of Mohn's rho that fall outside the range (-0.15 to 0.20) can be interpreted as an indication of a retrospective pattern for long-lived species. Although all three models passes the test, Model_3 had the least bias, Model_2 the second least, and Model_1 the most. These results indicate that Model_3 had the least retrospective pattern.

A summary of all diagnostics and performance for each model are shown in **Table 3**. Overall, Model_3 performed the best with regard to diagnostics.

Kobe plots for all models and all models combined are shown in **Figure 20**. Model_1 results show a narrow degree of uncertainty in the estimation of spawning stock biomass. This is a function of the number of parameters that were required to be fixed due to a lack of observational data from which to estimate them from. Model_2 had more uncertainty due to an additional parameter (catch multiplier) being estimated. Model 3 had the most uncertainty of the estimates of the management benchmarks due to not only the catch multiplier being estimated but also increases in most of the CPUE indices. Results of all three models indicate that the white marlin stock is overfished but not experiencing overfishing.

4. Discussion

The results report here are similar to those reported in the 2012 white marlin assessment. Because of the limited amount of observational data and lack of a clear signal within it, the uncertainty around the derived quantities associated with MSY are likely to be under estimated. One approach to capture more of the uncertainty would be to estimate more parameters and, in the face of a lack of data, use informative priors.

Basic biological information is absent for white marlin. Estimates of growth are one of the most influential but least understood life history traits. Because of a lack of age and growth information the longevity of the species, even in a fished population, is highly uncertain. This leads to greater uncertainty in the rate of natural mortality. The assumption of one Atlantic-wide stock is also lacking data.

The consistently declining trend in fishing mortality is an indication that perhaps recent management actions are in fact working as desired. However, the slow increase in spawning biomass for the same period indicates that the stock may not highly productive and may be slow to recover to the spawning stock biomass target. Unaccounted for landings and dead discards have a great influence on the uncertainty in the estimation of stock status and the effectiveness of management actions. The estimated catch multiplier parameter suggests that catch (and dead discards) could be under reported by13-17%. Furthermore, landings have exceeded TAC since 2013. All these factors could also be contributing to the slow recovery rate.

Acknowledgements.

The author would like to acknowledge the help of D. Courtney and H. Winker for sharing R code for diagnostics and discussions of the same.

References

Anonymous. 2018a. Blue Marlin Data Preparatory Meeting, Madrid, Spain, 12-16 March 2018. Document SCRS/2018/001: 39 pp.

Anonymous. 2018b. Blue Marlin Stock Assessment, Miami, United States, 18-22 June 2018. Document SCRS/2018/001: 45 pp.

Anonymous. 2019 White Marlin Data Preparatory Meeting, Madrid, Spain, 12-15 March 2019. Document SCRS/2018/000 (in press): 32 pp.

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

Drew, K., D. Die, and F. Arocha. 2006a. Current efforts to develop an age and growth model of blue marlin (Makaira nigricans) and white marlin (Tetrapturus albidus). Coll. Vol. Sci. Papers ICCAT 59: 274–281.

Drew, K., Die, D.J. and Arocha, F., 2006b. Understanding vascularization in fin spines of white marlin (Tetrapturus albidus). Bulletin of Marine Science, 79(3), pp.847-852.

Drew, K., Die, D.J. and Arocha, F., 2007. Ring counts and timing of ring formation in fin spines of white marlin (Tetrapturus albidus) from the Venezuelan longline and artisanal fisheries. Collect. Vol. Sci. Pap. ICCAT, 60(5), pp.1726-1730.

Die, D.J. and Drew, K., 2008, March. An Atlantic-wide study of age and growth of Atlantic marlins. In Proceedings from the Atlantic Billfish Research Program Symposium (pp. 67-84).

Oliveira, I. M., Hazin, F., Travassos, P., Pinheiro, P. and Hazin, H. G. 2007. Preliminary results on the reproductive biology of the white marlin, Tetrapturus albidus Poey 1960, in the western equatorial Atlantic Ocean. ICCAT Col. Vol. Sci. Pap., 60: 1738-1745.

Orbesen, E. S., Hoolihan, J. P., Serafy, J. E., Snodgrass, D., Peel, E. M. and Prince, E. D. 2008. Transboundary movement of Atlantic istiophorid billfishes among international and U.S. domestic management areas inferred from mark-recapture studies. Marine Fisheries Review, 70: 14-23.

Ortiz, M., Prince, E. D., Serafy, J. E., Holts, D. B., Davy, K. B., Pepperell, J. G., Lowry, M. B. and Holdsworth, J. C. 2003. Global overview of the major constituent-based billfish tagging programs and their results since 1954. Marine and Freshwater Research, 54: 489-507.

Table 1. Reported landings and dead discards of commercially caught white marlin/spearfish by CPC 2013-2017 and landings and dead discards limit for the same years. Note that estimated discards discussed in this document are not included in these landings.

Flag	2013	2014	2015	2016	2017	Limit
Barbados	6	6	10	14	17	10
Brazil	352	102	121	67	47	50
Canada	3	5	3	1	3	10
China PR	2		0	0	3	10
Chinese Taipei	7	7	12	12	7	50
Côte d'Ivoire	1	1	1	1	1	10
Curaçao					1	
El Salvador					0	
EU.España	42	99	125	96	118	
EU.France	0	0	0	1	1	50
EU.Portugal	10	9	7	11	13	
Ghana	1	1	1	1	0	
Guatemala					0	
Japan	24	6	8	9	10	35
Korea Rep.		0			0	20
Maroc					0	
Mexico	31	20	26	20	12	25
Panama					0	
Philippines	2	2				
S. Tomé e Príncipe	41	42	17	15	13	20
St. Vincent and Grena	0				8	
Sta. Lucia	1	1	1	0	1	
Trinidad and Tobago	33	38	32	20		15
U.S.A.	17	18	12	7	7	
UK.Bermuda	0	1	0	0	2	
Vanuatu	0					
Venezuela	77	99	119	187	192	50
Other	4	2	3	2	3	45
Grand Total	**654**	**460**	**495**	**466**	**458**	**400**

Table 2. Summary of pass (green) / fail (red) for profile analysis by model and data type and fleet. Failure of test was a result of the observational data profile have a 1% or less change in the profile shape (i.e. flat)

	MODEL 1				MODEL 2						MODEL 3					
		steepness		R0		steepness		R0		catch mult.		steepness		R0		catch mult.
TYPE																
TOTAL	1.000	TRUE	1.000	TRUE	1.000	TRUE	1.000	TRUE	1.000	TRUE	1.000	TRUE	1.000	TRUE	1.000	TRUE
Catch	0.000	FALSE	0.000	FALSE	0.000	FALSE	0.000	FALSE	0.000	FALSE	0.000	FALSE	0.001	FALSE	0.000	FALSE
Survey	0.447	TRUE	0.881	TRUE	0.462	TRUE	0.800	TRUE	0.605	TRUE	0.365	TRUE	0.589	TRUE	0.342	TRUE
Length_comp	0.345	TRUE	0.160	TRUE	0.392	TRUE	0.123	TRUE	1.413	TRUE	0.501	TRUE	0.399	TRUE	1.259	TRUE
Recruitment	0.210	TRUE	0.063	TRUE	0.274	TRUE	0.110	TRUE	1.079	TRUE	0.259	TRUE	0.238	TRUE	0.677	TRUE
Parm_priors	0.430	TRUE	0.011	TRUE	0.439	TRUE	0.184	TRUE	0.089	TRUE	0.668	TRUE	0.548	TRUE	0.083	TRUE
SURVEY																
Japan_LL_prior	0.270	TRUE	0.369	TRUE	0.449	TRUE	0.320	TRUE	1.219	TRUE	0.167	TRUE	0.193	TRUE	0.343	TRUE
Japan_LL_early	0.179	TRUE	0.142	TRUE	0.159	TRUE	0.114	TRUE	0.124	TRUE	0.479	TRUE	0.421	TRUE	1.340	TRUE
Japan_LL_mid	0.009	FALSE	0.001	FALSE	0.007	FALSE	0.001	FALSE	0.029	TRUE	0.102	TRUE	0.049	TRUE	0.520	TRUE
Japan_LL_late	0.032	TRUE	0.035	TRUE	0.034	TRUE	0.030	TRUE	0.201	TRUE	0.063	TRUE	0.046	TRUE	0.272	TRUE
Ven_LL	0.089	TRUE	0.005	FALSE	0.084	TRUE	0.011	TRUE	0.047	TRUE	0.063	TRUE	0.023	TRUE	0.142	TRUE
Ven_GN	0.060	TRUE	0.000	FALSE	0.064	TRUE	0.012	TRUE	0.020	TRUE	0.136	TRUE	0.054	TRUE	0.122	TRUE
US_LL	0.028	TRUE	0.022	TRUE	0.041	TRUE	0.008	FALSE	0.030	TRUE	0.042	TRUE	0.049	TRUE	0.348	TRUE
Spain_LL	0.209	TRUE	0.101	TRUE	0.278	TRUE	0.126	TRUE	0.472	TRUE	0.039	TRUE	0.034	TRUE	0.129	TRUE
Chin_Tai_LL_early	0.760	TRUE	0.212	TRUE	0.866	TRUE	0.457	TRUE	2.748	TRUE	0.262	TRUE	0.233	TRUE	0.698	TRUE
Chin_Tai_LL_mid	0.064	TRUE	0.014	TRUE	0.062	TRUE	0.007	FALSE	0.056	TRUE	0.022	TRUE	0.014	TRUE	0.030	TRUE
Brazil_Sport	0.017	TRUE	0.011	TRUE	0.040	TRUE	0.010	FALSE	0.169	TRUE	0.129	TRUE	0.119	TRUE	0.142	TRUE
Brazil_LL	0.115	TRUE	0.123	TRUE	0.156	TRUE	0.098	TRUE	0.325	TRUE	0.062	TRUE	0.153	TRUE	0.572	TRUE
LENGTHS																
Gill_Net_1	0.038	TRUE	0.030	TRUE	0.040	TRUE	0.036	TRUE	0.026	TRUE	0.034	TRUE	0.035	TRUE	0.037	TRUE
LongLine_2	0.378	TRUE	1.437	TRUE	0.295	TRUE	0.755	TRUE	0.255	TRUE	0.316	TRUE	0.374	TRUE	0.321	TRUE
Sport_4	0.584	TRUE	0.943	TRUE	0.665	TRUE	0.797	TRUE	0.720	TRUE	0.650	TRUE	0.892	TRUE	0.644	TRUE

Table 3. Overall results of diagnostics for Model_1, Model_2 and Model_3. Red represents either failure of the test or the worst performing model of the three. Yellow represents the model diagnostic was either somewhere in between the other two models or the results were inconclusive. Green represents either a passing of the test or the best performing of the three models.

DIAGNOSTIC	Model 1	Model 2	Model 3	Comments
Profiing				Objective scoring of curvature
Run Test CPUE				Objective counts of pass/fail tests
Residual Analysis				Objective calculation of RMSE
Recruitment Deviations				Not applicable for accept/reject
Retrospective Analysis				Objective statistical test

Table 4. Selected parameter estimates and derived management quantities, standard deviations and lower/upper confidences intervals (+/- 1.96*SD) for three models considered in this study for white marlin.

	MODEL_1					MODEL_2					MODEL_3			
Parameter	Estimate	SD	LCI	UCI		Estimate	SD	LCI	UCI		Estimate	SD	LCI	UCI
R0	5.514	0.080	5.356	5.671		5.503	0.082	5.342	5.663		5.543	0.084	5.378	5.708
steepness	0.550	0.046	0.460	0.641		0.567	0.047	0.475	0.660		0.552	0.046	0.461	0.643
Catch Mult.	1	N/A	N/A	N/A		0.835	0.095	0.649	1.022		0.871	0.129	0.618	1.123
F_{MSY}	0.172	0.02753	0.118	0.226		0.182	0.029	0.125	0.238		0.172	0.027	0.118	0.226
SSB_{MSY}	2090.7	309.8	1483.5	2698.0		2017.9	306.5	1417.1	2618.7		2149.2	325.1	1512.1	2786.3
MSY	1355.4	50.7	1255.9	1454.8		1394.6	51.0	1294.6	1494.5		1397.9	52.7	1294.7	1501.1
SSB_{2017}/SSB_{MSY}	0.411	0.074	0.266	0.555		0.415	0.078	0.261	0.568		0.512	0.112	0.292	0.732
F_{2017}/F_{MSY}	0.937	0.140	0.663	1.210		0.990	0.149	0.698	1.282		0.768	0.112	0.548	0.988

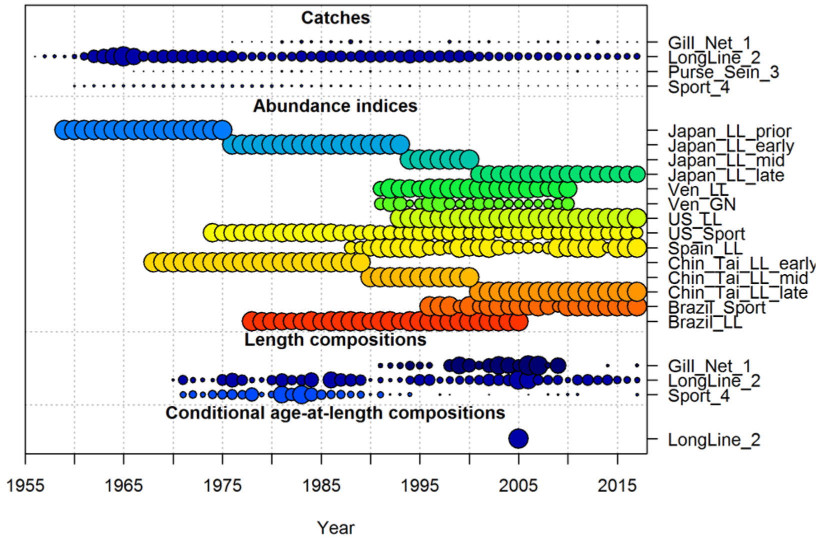

Figure 1. Data sources by year used in the Stock Synthesis models.

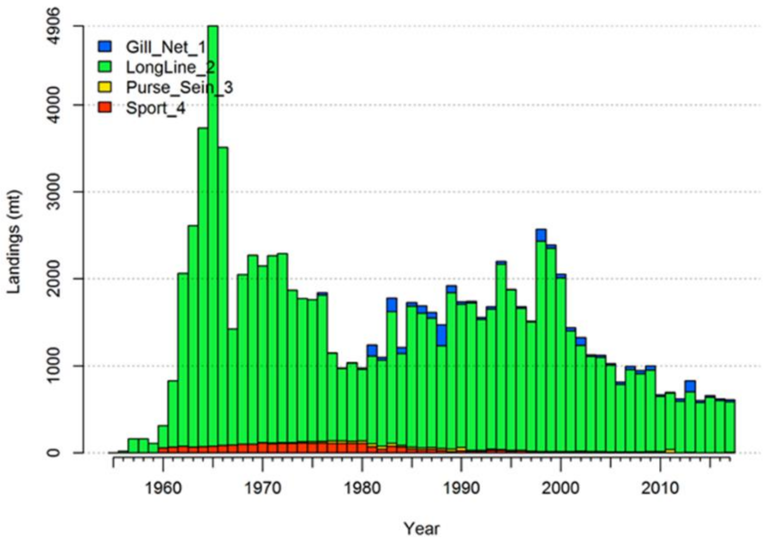

Figure 2. Landings of white marlin and spearfish by year and gear.

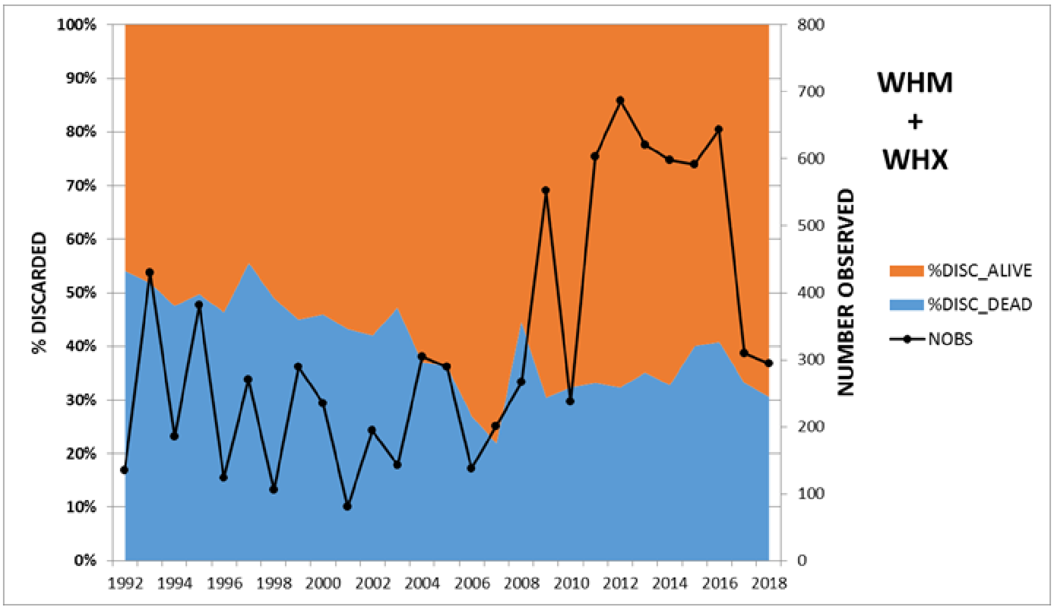

Figure 3. Percent of white marlin and "either white marlin (WHM) or spearfish (WHX) and disposition from US observer program

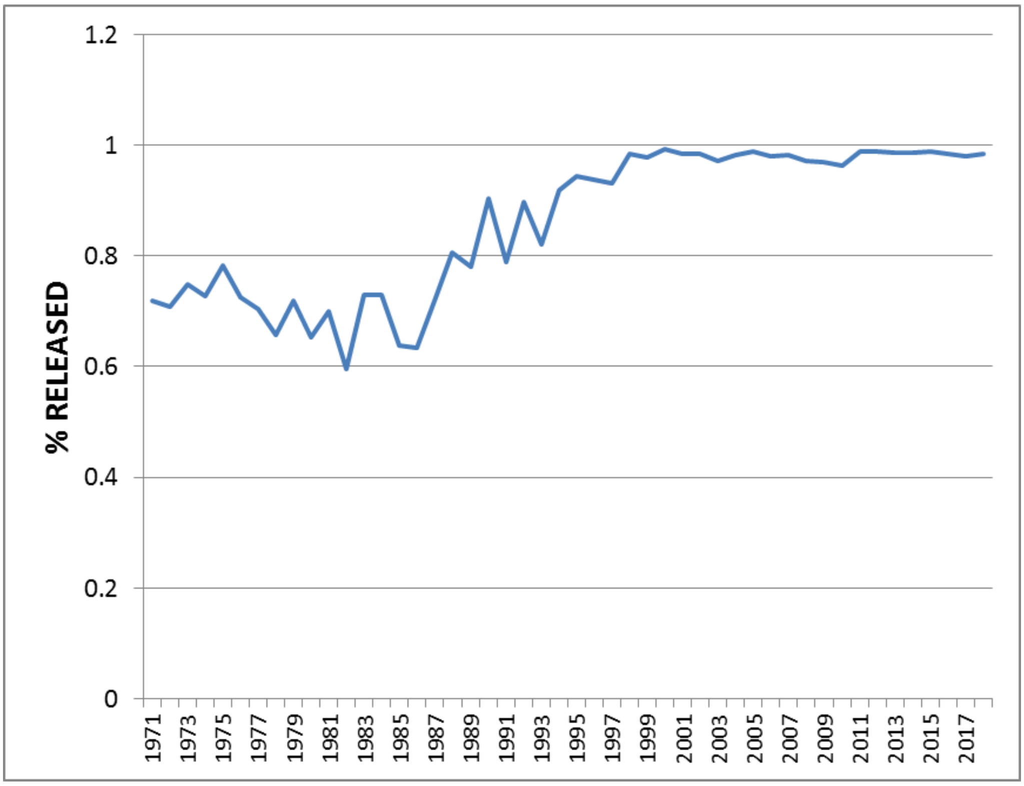

Figure 4. Percent of white marlin released from US tournaments.

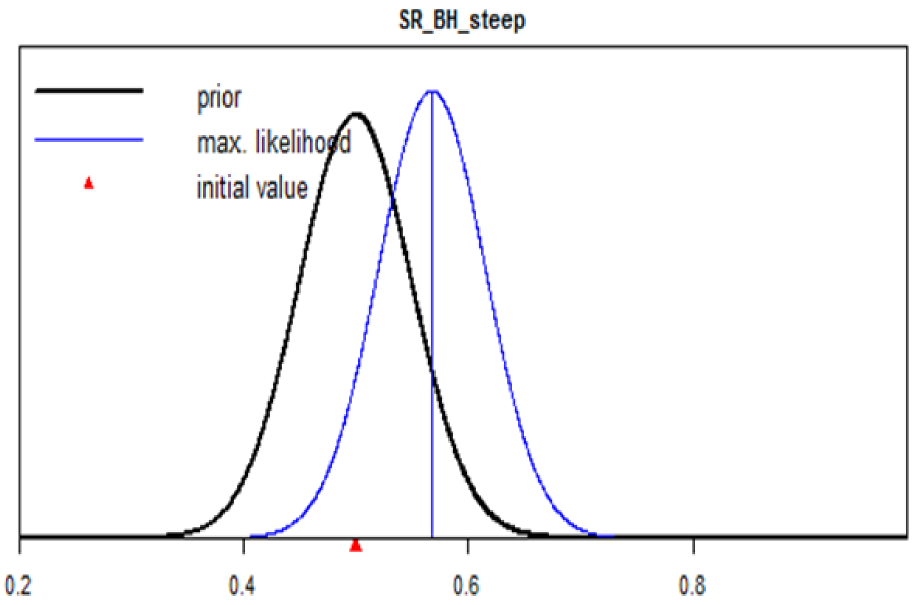

Figure 5. Prior distribution and maximum likelihood for the steepness parameter used in all models.

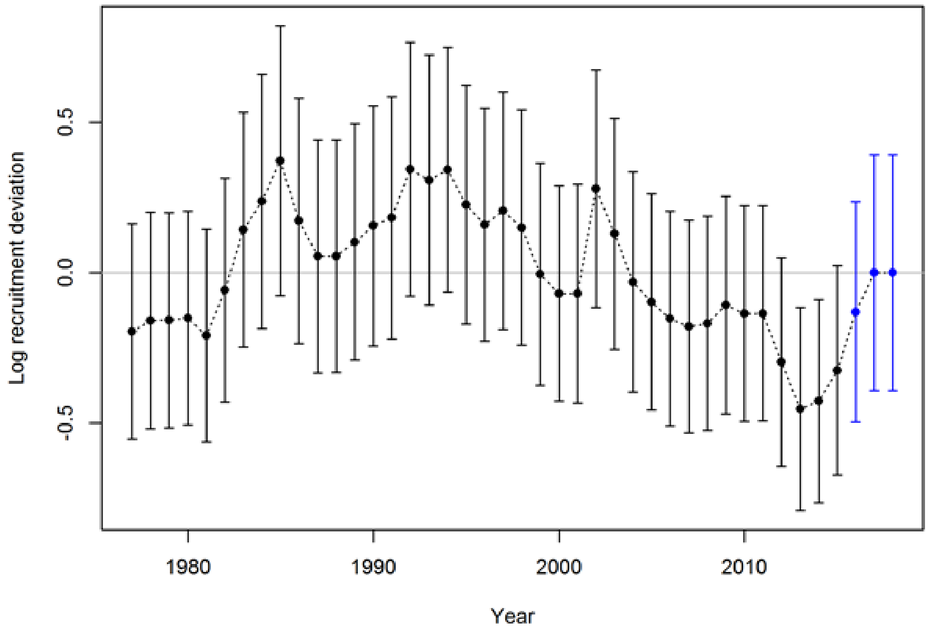

Figure 6. Time series of recruitment deviations for Model_1.

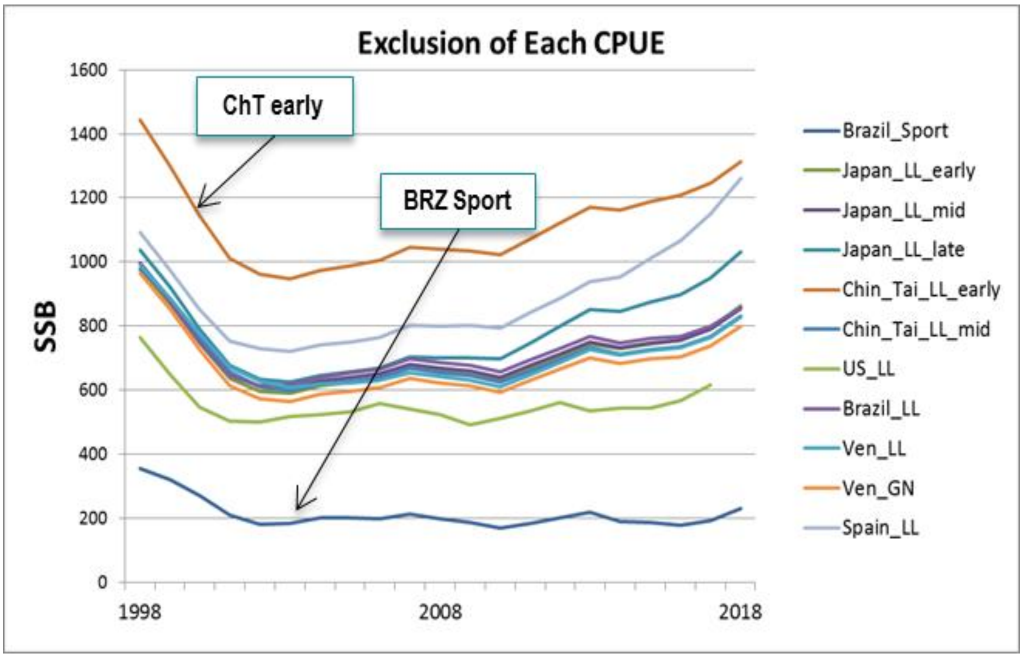

Figure 7. Trends in spawning stock biomass with each of the CPUE's used in Model_1.

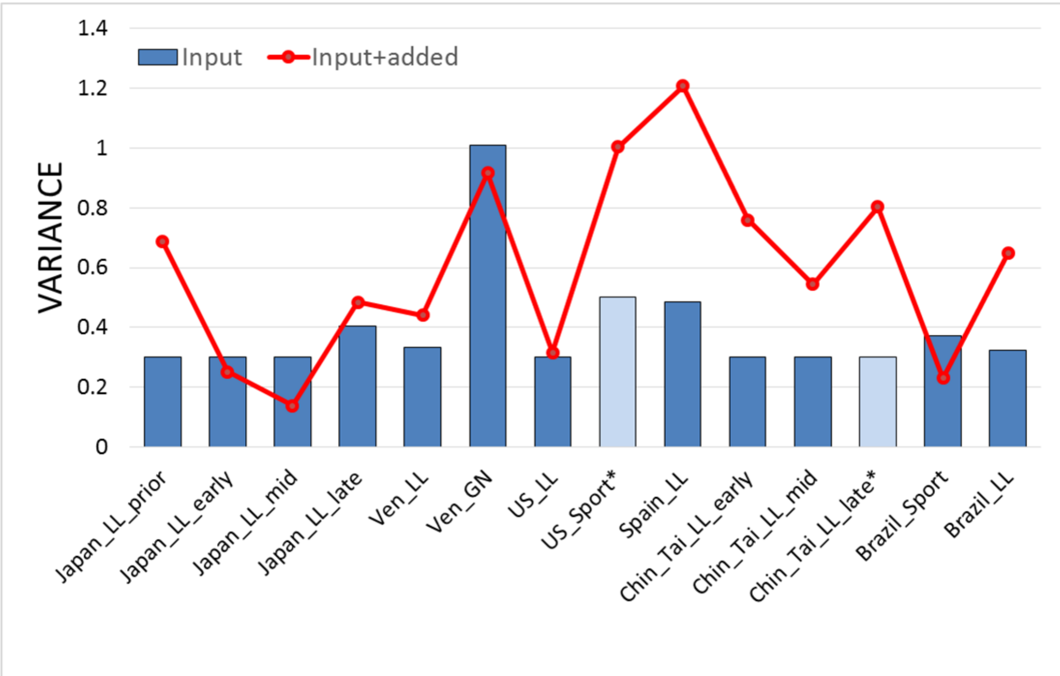

Figure 8. Input and Input + added variance for CPUEs for Model_3

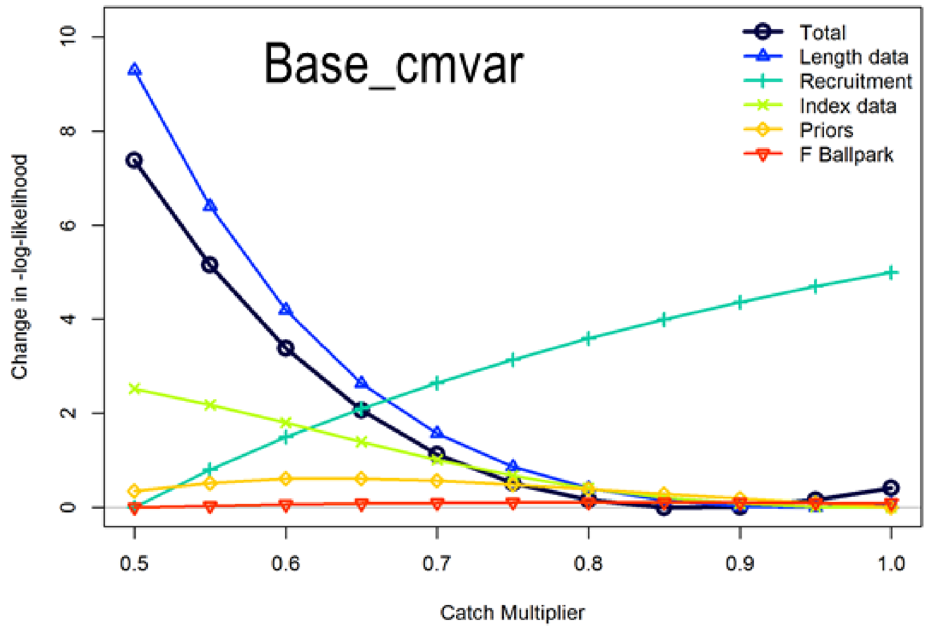

Figure 9. Profile analysis on catch multiplier for Model_3.

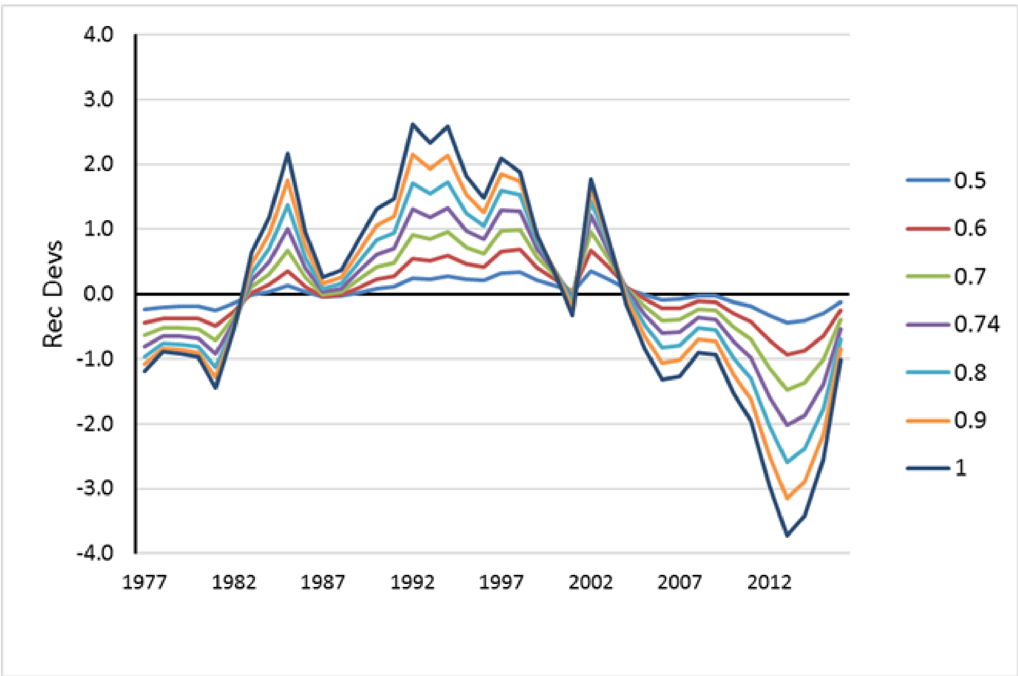

Figure 10. Trends in recruitment deviations for various values of steepness

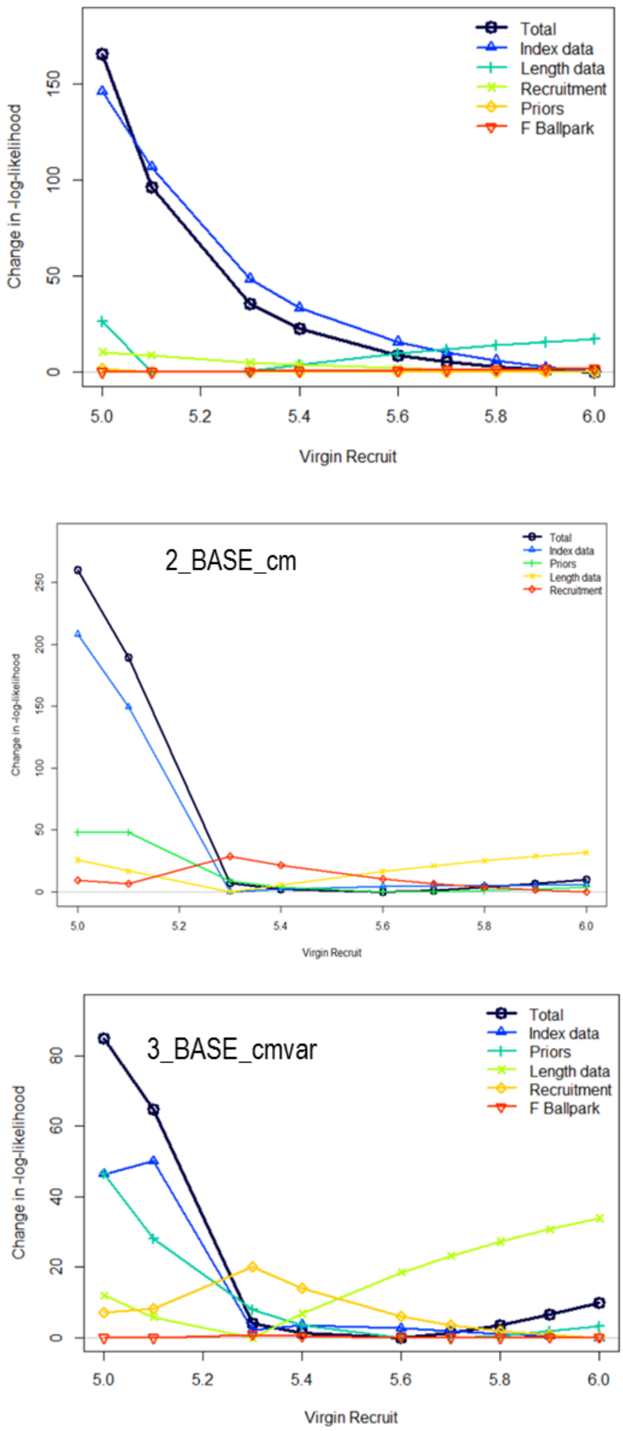

Figure 11. Profile analysis on R0 by data type for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

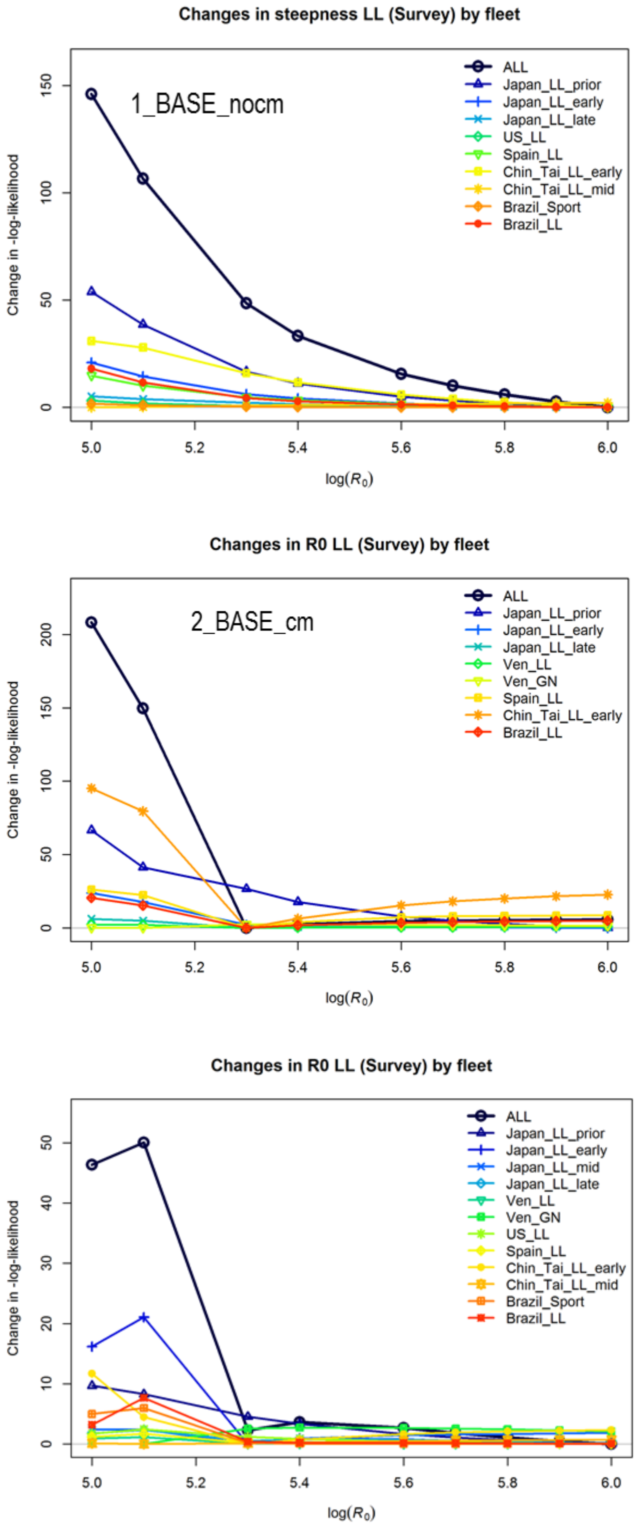

Figure 12. Profile analysis on R0 by CPUE index for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

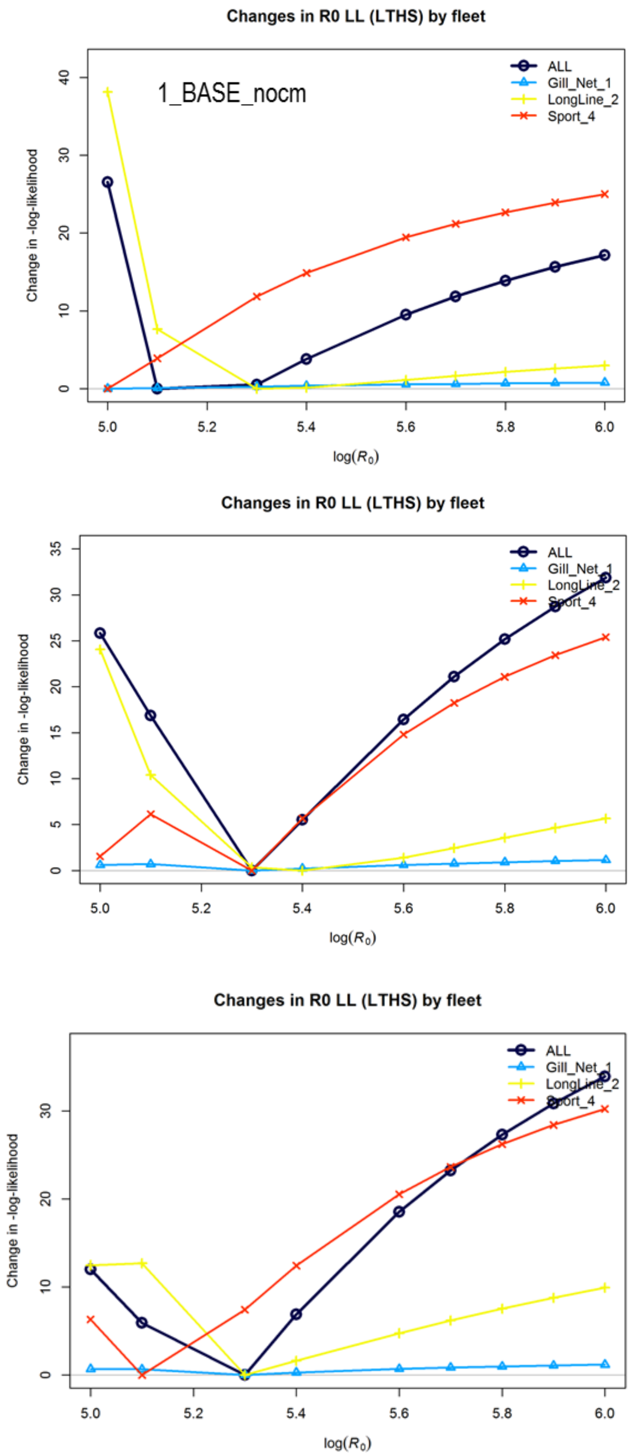

Figure 13. Profile analysis on R0 by length source for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

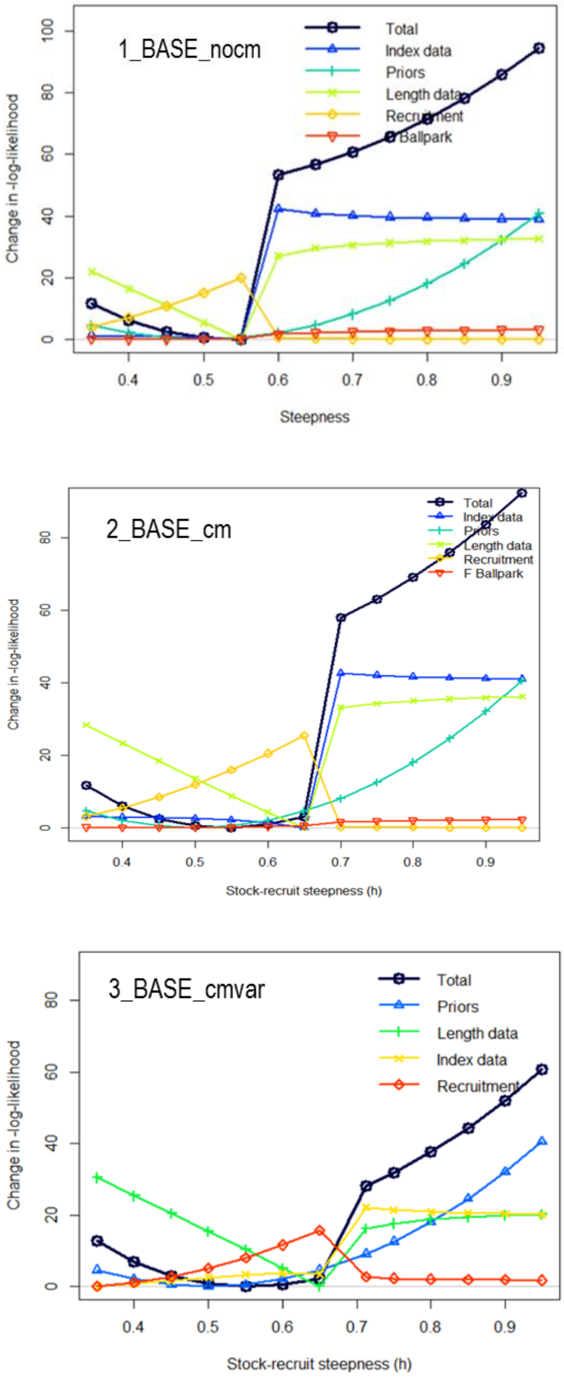

Figure 14. Profile analysis on steepness by model and data type for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

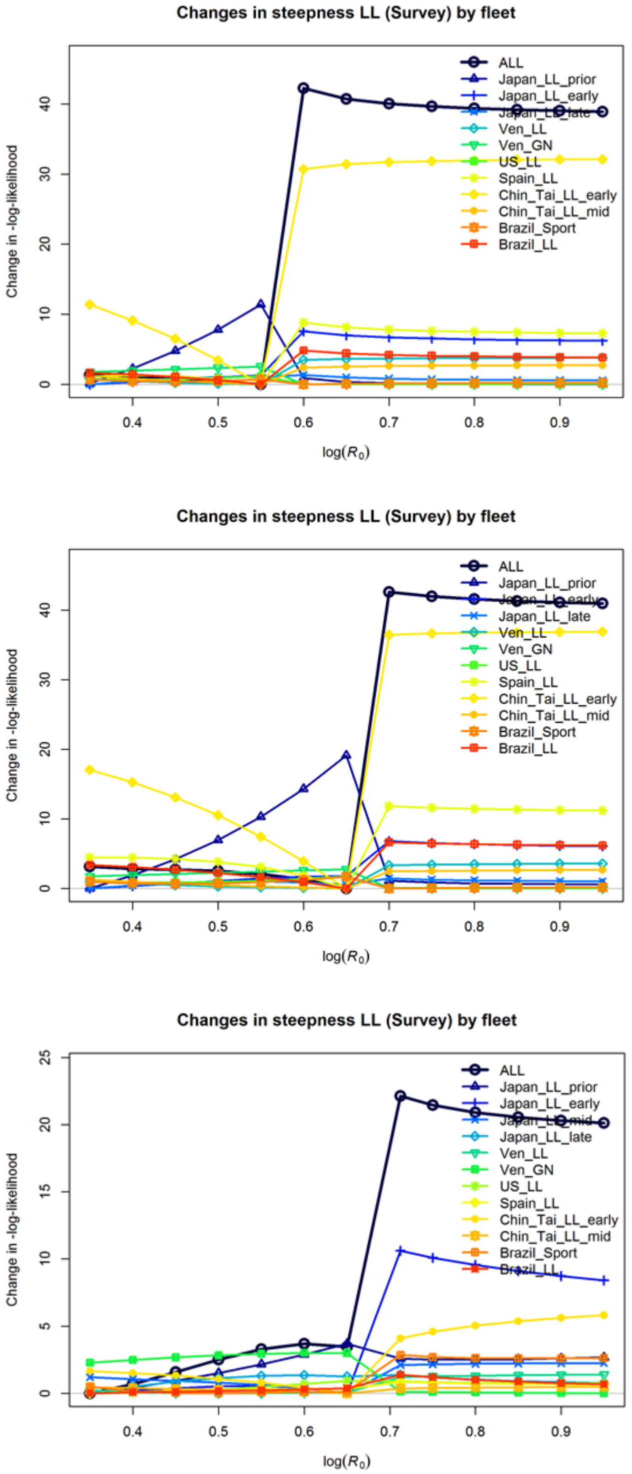

Figure 15. Profile analysis on steepness by CPUE index for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

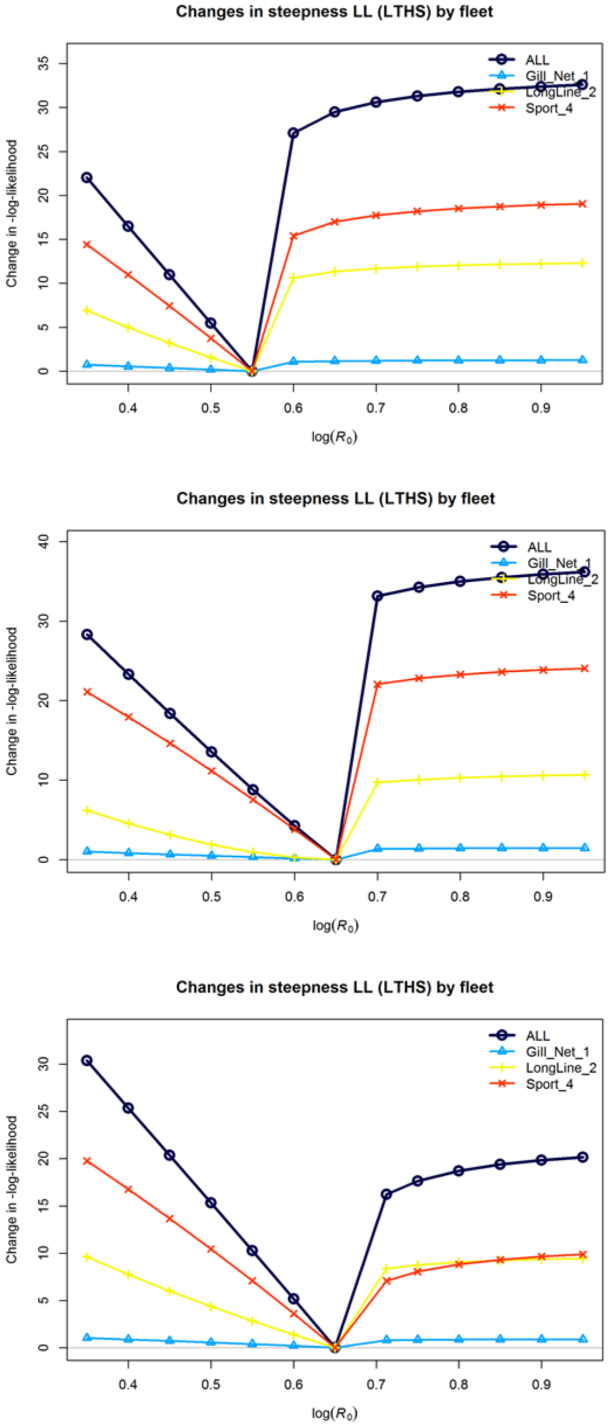

Figure 16. Profile analysis on steepness by length source for Model_1 (top), Model_2 (middle), and Model_3 (bottom).

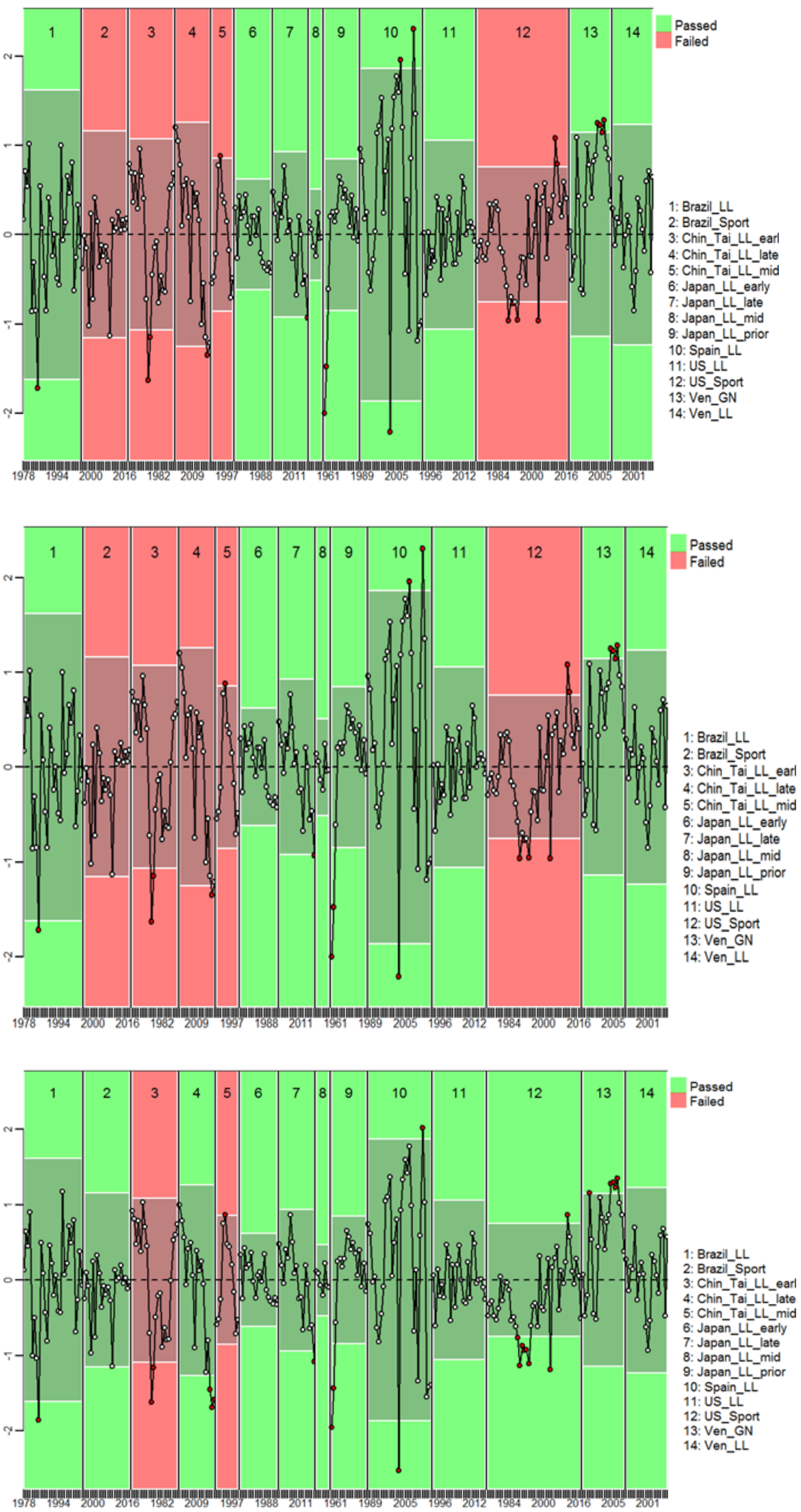

Figure 17. Run tests on fits to CPUE for Model_1 (top), Model_2 (middle) and Model_3 (bottom). Green background indicates no evidence (p > 0.05) to reject the hypothesis of a randomly distributed time-series of residuals. Red background indicates failure of test.

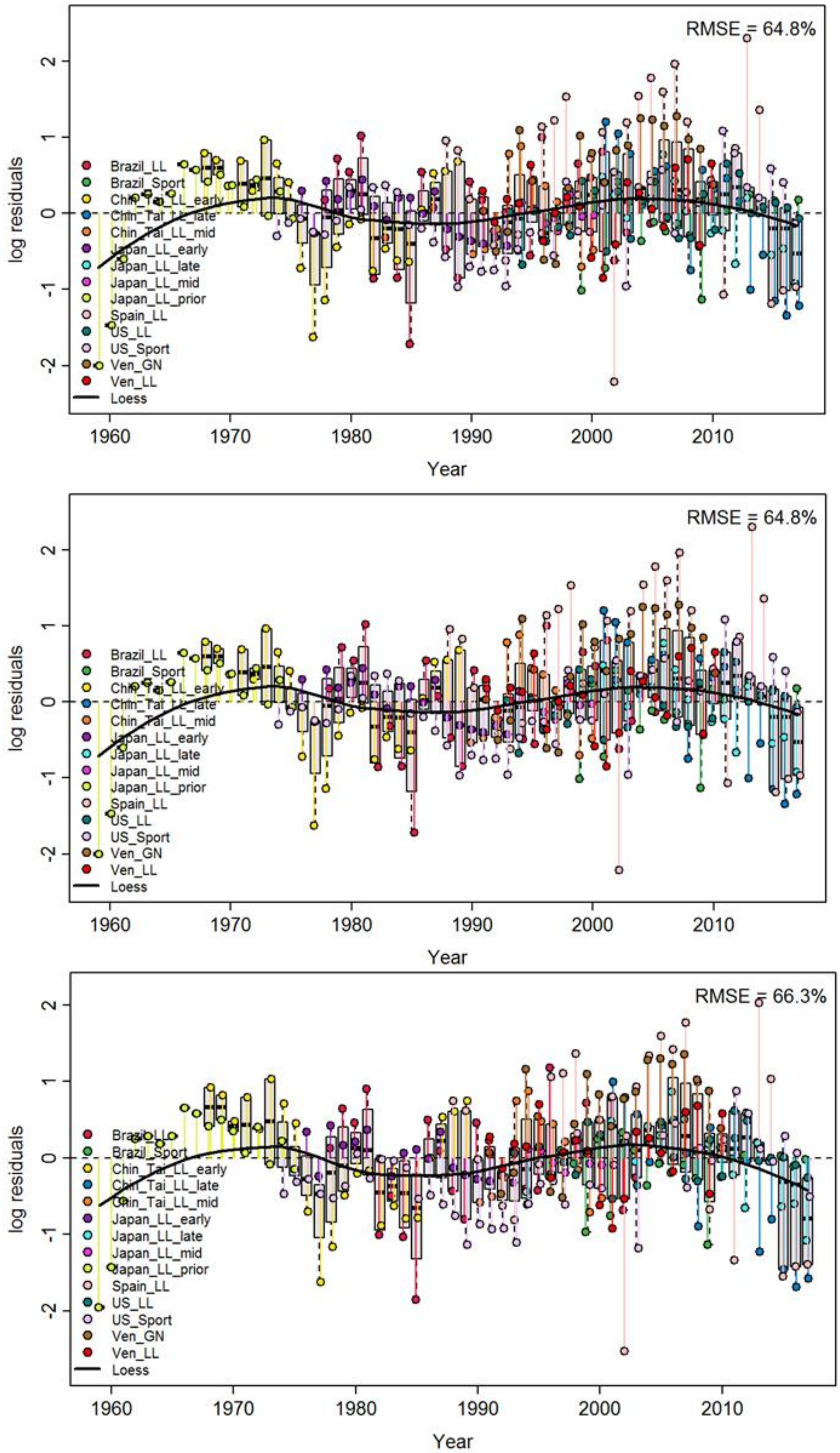

Figure 18. JABBA residual diagnostic plots for Model_1 (top), Model_2 (middle), Model_3 (bottom).

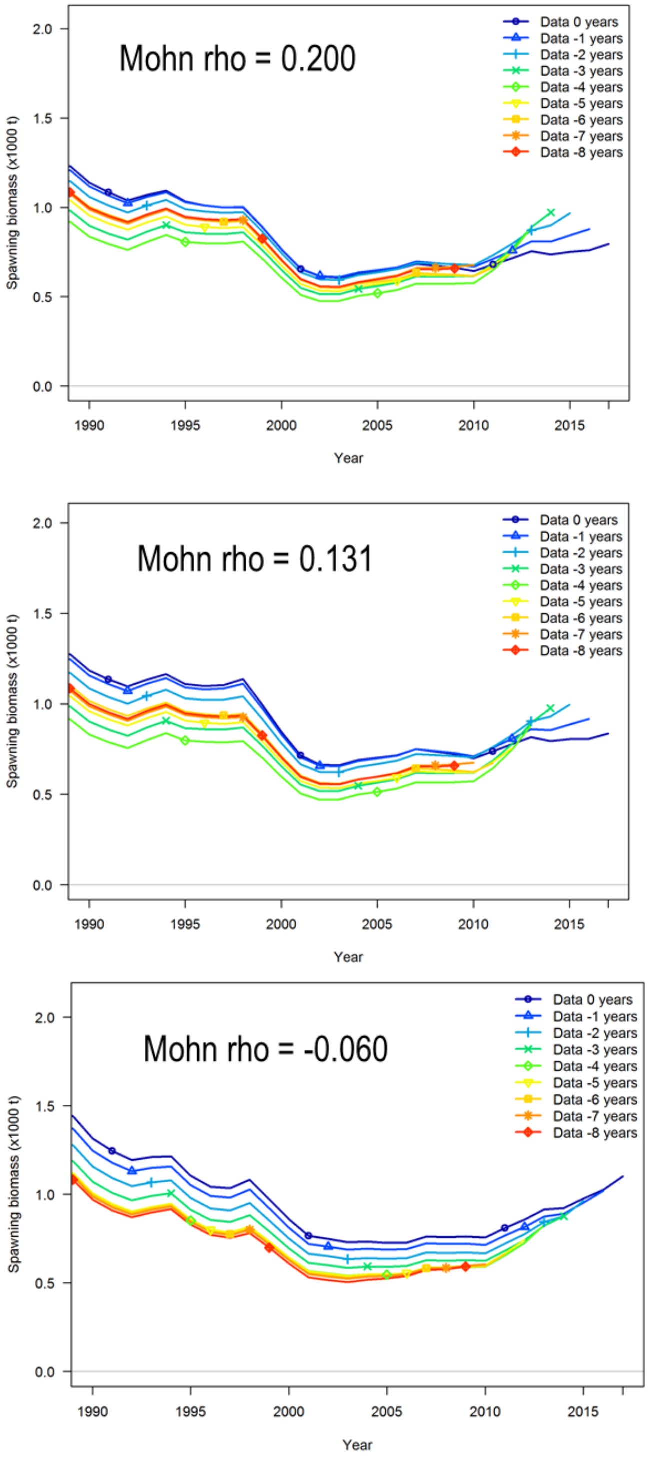

Figure 19. Retrospective analysis for Model_1 (top), Model_2 (middle) and Model_3 (bottom) for spawning stock biomass. The Mohn's rho statistic was calculated for a 5 year "peel".

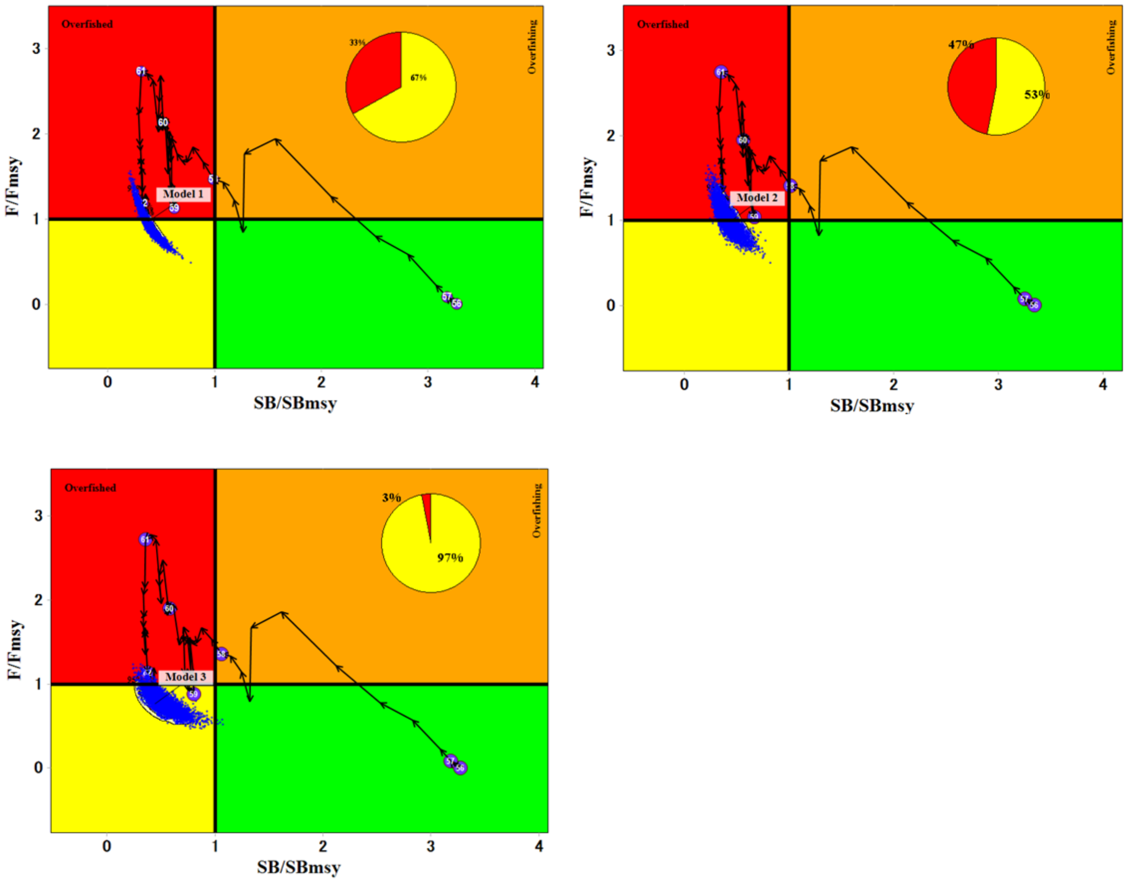

Figure 20. Kobe plots for Model_1 (upper left), Model_2 (upper right), Model_3 (lower left), and all models (lower right).

Appendix 1

ICCAT & SCRS billfish rulings since 1997

1997 - Reduce white and blue marlin landings by 25% of the 1996 total. Promote the release of live white and blue marlin

1998 - Landings of blue and white marlin in 2000 must not be greater than the 1999 total. Current Atlantic blue marlin biomass is at 24% of the MSY level, and that Atlantic white marlin biomass is at 23% of the MSY level; Postpone the 1999 SCRS white and blue marlin stock assessments until 2000.

2000 - Establish a 2 phase Plan to Rebuild Blue Marlin and White Marlin Populations " ...blue marlin has been reduced to a level of 40% and white marlin 15% of that needed to produce maximum sustainable yield and that neither stock is likely to recover if the current levels of mortality continue into the future."

2001 - White marlin stock assessment in 2002. Blue marlin stock assessment postponed until 2003.

2002 - Significant uncertainties with the 2002 white marlin stock assessment. New stock assessments for Atlantic blue marlin and white marlin in 2005. Phase 1 of the Rebuilding Plan for Blue and White Marlin shall remain in effect through 2005 with blue marlin landings 50% of 1999 levels, white marlin 33% of 1999 levels.

2004 - The 2005 stock assessments of Atlantic blue marlin and white marlin postponed until 2006, consistent with the process advised by the SCRS. Phase 1 of the Rebuilding Plan for Blue and White Marlin shall remain in effect through 2006.

2006 - SCRS asks for help with data, logs and observations to make more reliable stock assessments. White and blue marlin assessments in 2010. Phase 1 of the Rebuilding Plan for Blue and White Marlin shall remain in effect through 2010. Also first mention of use of circle hooks and regulating artisanal fleets.

2007 - No specific Atlantic billfish recommendations, but evidence shows many vessels believed to be illegal, unregulated or under-reporting catches (IUU).

2010 - Phase 1 of the Rebuilding Plan for Blue and White Marlin shall remain in effect through 2011. White marlin assessment postponed until 2012.

2011 - New blue marlin assessment indicates that unless the recent catch levels are substantially reduced, the stock will likely continue to decline. Immediately reduce fishing mortality on blue marlin stock, by adopting a TAC of 2,000 mt or less.

2012 - The results of the 2012 white marlin assessment, indicated that the stock was overfished but most likely not undergoing overfishing, while noting significant uncertainty associated with the assessment. White marlin limit of 400 mt. Blue marlin 2000 mt. continued. Length minimums added for recreational anglers.

2015 - An annual limit of 2,000 t for blue marlin and 400 t for white marlin/spearfish is continued for these stocks, for 2016, 2017 and 2018. This landings limit shall be allocated among 13 nations. Work towards assuring compliance with reporting. Efforts to ensure that all blue marlin and white marlin/spearfish that are alive by the time of boarding are released in a manner that maximizes their survival. Recreational fisheries that meet or exceed the following lengths: 251 cm LJFL for blue marlin and 168 cm LJFL for white marlin/spearfish.

Appendix 2

Communications with white marlin tournament fishers

Question: We use billfish tournament data to represent recreational catch and effort to help estimate changes in the quantity of fish available. If the effort has changed in a significant way that would alter the catch per unit of effort then that could have a major impact on the stock assessment results.

For instance, in my history in the fishery for blue marlin the fishing went from slow trolling dead baits, to high speed lures, to slow speed light tackle trolling, to teaser and pitch bait fishing and run and gun styles. Many of which have their own variation depending on what region you are fishing in. To make an overall assumption that the effort has stayed the same would be a mistake when interpreting the catch data so we could get false information from the assessment process.

Do you think that the white marlin fishery has had similar changes in effort? In other words, do you think that the way that fishers pursue them has changed significantly from, say, the 1970s until now?

Now, a second avenue to think about is where the fish are being targeted. Do fishers put the same effort into the same waters that have always been fished?

There are other variables being used in the stock assessment process so it is not critical to answer these questions but it is worthwhile to have a better understanding of shifts that might impact the catch rates.

Answer: Fishermen #1

The White marlin fishery in my opinion has changed drastically. Starting with the implementation of the dredge teaser, the type of tackle being used like lighter leader material and smaller hooks, resulting in a better hook up ratio. Secondly would be the boats. Boats have becoming a lot faster and bigger resulting in waters being fished that were not fished in the 70's. For example the North East Canyons and inside of that were the places that were always fished in the 70's, now boats consistently travel 90-100 miles in search of better conditions, and putting up big numbers sometimes in the 20-30 fish days and more. Technics also with the use of circle hooks has helped the fishery a lot. The mortality rate has to be way down after the release. So to answer your question the effort and technology has changed as well as the waters being fished because of this. The run and gun that is going on in the Bahamas is very new to the White Marlin seen or round scale spear/ hatchet marlin, whatever they are calling them. I have fished the Bahamas all my life and out there where they are fishing and have never seen that type of action out there like they are seeing. Hope this helps and let me know how I can always be of help. I still look at that data you sent me when we tagged those White Marlin in Ocean City that ended up in Trinidad and Tobago.

Answer: Fishermen #2

The short answer is that speed technology and understanding of the habits and instincts of Marlin behavior has changed exponentially over the past forty years. I remember when a good day of White Marlin Fishing out of OC was one or two fish. I won’t even go if I don’t think I can get ten bites now. The number of boats fishing and the range and speed of those boats allows us to find the fish and stay on them. A fast boat in the seventies was eighteen knots today it is forty with almost all boats being able to achieve thirty knots. CHIRP wide angle transducers and now sonar are changing things even more. I would guess that catching ten Marlin today is equivalent to the effort to catch one in the eighties.