Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTIETH REGULAR SESSION

Manila, Philippines
14 – 21 August 2024

ISC's Western and Central North Pacific Striped Marlin Assessment Consensus Peer Review

WCPFC-SC20-2024/GN-WP-04

ISC[1]

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

International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) Western and Central North Pacific Striped Marlin Assessment Consensus Peer Review

April 15-19, 2024
Institute of Oceanography, National Taiwan University (IONTU), Chinese Taipei

Reviewers

Hiromu Fukuda fukuda_hiromu57@fra.go.jp
Simon Hoyle simon.hoyle@gmail.com
Ian Stewart Ian.Stewart@iphc.int

Chair

Robert Ahrens robert.ahrens@noaa.gov

Table of contents

List of acronyms

BILLWG - Billfish working group
CPUE - catch per unit effort
MLS - Striped marlin
IBBS - International billfish biological sampling program
ISC - The International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean
SS3 - Stock Synthesis 3
TOR - Terms of reference
WCNPO - Western and central north Pacific Ocean
WG - Working group

Introduction

This document is a consensus report for review of the International Scientific Committee for tuna and tuna-like species (ISC) western and central North Pacific striped marlin 2023 assessment. This report was prepared by Dr. Hiromu Fukuda, Dr. Simon Hoyle, and Dr. Ian Stewart with support from the review chair Dr. Robert Ahrens. The in-person review meeting took place April 15-19, 2024 at the Institute of Oceanography, National Taiwan University (IONTU), Chinese Taipei. The review meeting was structured according to terms of reference (TOR) provided by the ISC billfish working group (BILLWG). Presentations by BILLWG members were followed by discussion with the review panel members. The BILLWG conducted additional analyses during the meeting when requested, which helped the panel to understand the assessment and its input data. Materials provided to the panel included the 2023 assessment base case files, the 2023 stock assessment report, and documents describing the development of model inputs. The final day of the review involved further discussion of each of the panel's recommendations to ensure comprehensiveness and clarity. This process naturally led to consensus in the recommendations and as a result the review panel and the chair felt that a single consensus report was the most effective approach to communicate their recommendations to the ISC and BILLWG.

The panel and the review chair thank the members of the ISC BILLWG who attended the review meeting (Dr. Michelle Sculley, Dr. Yi-Jay Chang, Dr. Chi-Lu Sun, Dr. Wei-Chuan Chiang, Dr. Mikihiko Kai, Dr. Marko Jusup, and Dr. Jon Brodziak), for their hard work and exceptional support. They also thank the members of the IONTU for their generous hospitality during the review meeting. Finally, the panel and the review chair would like to express their deepest sympathies for all the people affected by the large earthquakes in the eastern Taiwan island in April 2024.

This document has three sections. The short introduction provides background outlining the motivations and history that led to this review. This is followed by a discussion section which is organized by TOR and provides a narrative of the main topics discussed relevant to the panel's recommendations. The final section provides the panel's specific recommendations and is also organized by TOR. Recommendations have been separated into “short-term” and “long-term”, the former are intended to be addressed for the next stock assessment and the latter are larger efforts that may take longer to reach resolution.

Background

The 2019 Western and Central North Pacific Ocean striped marlin (WCNPO MLS) assessment was conducted by the ISC BILLWG using the Stock Synthesis platform to fit a length-based age-structured population dynamics model to data on catches, size compositions, and indices of abundance based on catch per unit of effort (CPUE). The ISC indicated that the WCNPO MLS stock was overfished and that overfishing was occurring, which prompted the efforts of managers to better understand how the stock would respond to fisheries management actions. In the 2019 assessment, the BILLWG highlighted many important uncertainties that should be considered when evaluating the results of the assessment including: stock structure, driftnet catch, life history parameters (especially growth, maturity, and natural mortality), initial equilibrium conditions, input CPUE index standardization and spatio-temporal structure, and diagnostic patterns. Subsequently, requests have been made to the ISC BILLWG to provide additional information, including a rebuilding target based upon 20% of a dynamic SSB0 and a rebuilding plan based upon constant catch quotas. While the BILLWG attempted to address many of these concerns in the 2023 stock assessment prior to the development of a rebuilding plan, concerns associated with these uncertainties remain. The ISC has agreed that the assessment requires further investigation and expert advice in the form of an external peer review, in order to improve confidence in future WCNPO MLS stock assessments.

The terms of reference developed by the ISC outline the objectives and scope of an external peer review of the 2023 WCNPO MLS stock assessment in the ISC BILLWG. They are based upon the peer review process approved by the ISC Plenary in 2023. It is expected that the outcomes and recommendations from this peer review would be incorporated, to the extent possible, in the next WCNPO MLS assessment, tentatively scheduled for 2027.

Objectives

1. Undertake, in consultation with the ISC BILLWG and following the guidelines described in Process for the Independent Review of stock assessments, a peer review of the 2023 MLS stock assessment in WCNPO.
2. Based on the review work, provide recommendations for improving the assessment, including data inputs and related analyses, modeling approaches and treatment of uncertainty.
3. In conjunction with the BILLWG members, identify improvement options that are feasible for application to the next WCNPO MLS assessment.

Key motivations and discussion related to recommendations

This section of the report contains noteworthy discussion points and key factors that lead to the recommendations related to the sub components for each term of reference.

The panel noted two key features associated with uncertainty in the 2023 WCNPO MLS assessment. These were 1) an apparent change in the level of estimated recruitment (population scale) before and after the mid-1990's, and 2) insufficient age at length data for older fish to estimate the growth curve. The change in the estimated population scale occurring in the mid-1990's coincided with the moratorium on the high sea driftnet fishery (1992), the year in which the Japanese longline CPUE index was split (1994), the year in which size composition data was first included in the model (1994), and the first year for which the recruitment deviation from the stock recruitment relationship was estimated with a sum-to-zero constraint (1994). The growth curve in this assessment strongly affects the estimation of population scale through the relatively impactful model residuals regarding the size composition, given the current selectivity and data weighting assumptions. Thus, each of the broad topics in the TOR regarding fishing practice, input data, biological assumptions and model assumptions is concerned with the above-mentioned two features of the current assessment. The panel considers that a consistent approach will be necessary for each broad topic in the TOR in order to develop an internally consistent model. For this reason, there may be some duplication of the description in the sub-sections below.

In this section, we highlight the issues, discussions, and findings under each of the TORs so that the reader has some context for each recommendation.

TOR 1 - Review the information available on Pacific MLS stock structure and conceptual model and provide any recommendations for changing WCNPO MLS stock boundaries or to the fleet structure.

Stock boundaries

The BILLWG presented the most recent understanding of the stock structure of Pacific striped marlin. The most recent genetic work indicates at least 3 genetically distinct spawning populations. Two involve spawning within the western central Pacific (in the north and south) and the third in the eastern Pacific (Mamoozadeh et al., 2020). Currently, these populations are observed to have different life history characteristics (Farley et al., 2021). Under the WG's current conceptual model there is insignificant interaction between these populations and the assessment, consistent with the 2019 assessment, covers the North Pacific population in waters west of 150W and north of the equator.

The review panel noted that recent work around the Hawai'ian archipelago, based on genetics and popup satellite archival tags (Martinez 2021, Lam et al. 2021), indicates regional connectivity, particularly for the northern and southern striped marlin populations. A WG member also noted that fishermen in Hawai'i identify two physically different morphs of striped marlin. The review panel noted that the evidence for mixing adds a degree of complication to the current conceptual model. To the extent that these two populations mix across the western and central Pacific, unknown proportions of catches and lengths are misclassified. If the mixing is isolated to the Hawai'an archipelago the effect is likely not large because of the lower contribution of the US fishery to the total removals. However, given the apparent differences in growth between the populations, there is the potential for the length frequency distribution to be impacted and for estimates of fishing mortality, derived from fits to these data, to be impacted. Since the US fishery is interacting with mainly younger individuals this effect may be ameliorated.

The panel also noted that, in comparison with the rest of the Pacific, there are much larger fish at higher latitudes in the south Pacific, with a progressive increase in mean size with latitude, but there is no such trend observed in the northern hemisphere (Ijima and Kanaiwa 2019). The cause of this pattern in striped marlin has not been determined but hypotheses include spatially varying growth, discard patterns, and ontogenetic movement in the SWP stock.

Fleet Structure (this is also under model configuration)

The BILLWG presented the fleet structure and its rationale. The fleets in the 2023 assessment were mainly defined by the flag (the US and Taiwanese fleets), fishing gear (Japanese drift net fleets), and in some cases by the combination of the area and season (Japanese longline fleets). Significant effort has gone into the specification of fleets to conform to the areas-as-fleets assumption in the assessment model. There was a lot of discussion related to the challenges of defining fleets for such an approach with a highly mobile species that appears to have clear spatial ontogeny within the assessment region. As presented in the recommendation section, the review panel thought that more work was still needed on this aspect of the assessment model. There were concerns that the analysis used to develop the Japanese fleet categorization resulted in a fleet structure that was too complicated (ISC 2023). For all fleets the panel suggested that more complete investigations of the spatio-temporal distribution of size would be helpful in conjunction with a length frequency standardization approach. These efforts would help to highlight seasonal and spatial differences to inform fleet designations.

There was discussion of the potential benefits of the finite mixture model approach adopted for the current Japanese fleet designation (Ijima and Kanaiwa 2019). Most notable was the potential to also generate standardized CPUE for each fleet identified as part of the same analysis. It was noted that ultimately the structure generated by the approach was spatially complex and should be simplified. There was also discussion of the use of CPUE in the finite mixture model as the original intent of the approach was to identify potential areas of spawning individuals. The panel felt that for the purpose of fleet identification, CPUE was less appropriate than other operational level information and consistency in the length frequency distributions of the proposed fleets (in order to allow temporally constant selectivity).

There was discussion of the potential to disaggregate Taiwanese data (Lee et al. 2021) by season in a similar way to the Japanese fleets. However, it was considered that, at least at present, this would overly complicate the assessment. The strong residuals associated with the Taiwanese length data may be addressed by reducing effective sample sizes and by switching to the use of weight data, which is likely to be more reliable. The generalized size composition function of the Stock Synthesis software could be useful when applying the converted length composition data from the weight composition since weight measurement may be conducted at lower resolutions (e.g., in units of 1 kg and with a degree of rounding) than the length measurement (e.g., in units of cm) in particular for small MLS.

The panel discussed the potential to develop an index fishery approach (Maunder et al 2020, Xu et al. 2023) based on the joint analysis of data from multiple fleets. This would lead to a simplified fleet structure at least for the index fisheries, which could be based on data from the majority of the range of the stock.

TOR 2 - Model inputs, commenting on the adequacy and appropriateness of data sources and data inputs to the stock assessment.

Growth: review the approach to estimation of growth parameters and consider the implications of potential regional variations in growth.

The discussion around growth in the model was pervasive throughout the review. The assessment obtains most of its information about population scaling from the relationship between the population growth curve, fishery selectivity, and observed fish sizes. Decisions about the growth curve have a notable effect on the assessment results, and more data to inform the growth curve are required to reach consensus on these results.

There was general agreement that understanding growth is the most important area of research for this stock. The current IBSS program was commended by the review panel and should provide results that will enhance the assessment model.

A number of aspects of data acquisition and growth model fitting were revisited during discussion. The BILLWG presented information on the data used to estimate growth, selection of the functional form of the growth model, and the method used to estimate parameters. The BILLWG noted that the current growth curve was based on fin spine samples that were estimated to be aged from 0.5 to 6. Recent work comparing spine and otolith aging for the south Pacific Ocean MLS and albacore (Farley et al 2021[1,2]) indicates that ages from spines have a small positive bias at young ages, a large negative bias at older ages, and considerable uncertainty at all ages. Similar biases and uncertainty are likely for the WCNPO MLS.

A considerable improvement of this data situation is anticipated due to the currently ongoing IBSS. The panel recommended comparison of results between the otolith ageing and spine ageing from the same individuals obtained by the IBSS. The panel considered that, if a growth curve from the IBSS program is not available for the next assessment, then additional work will be needed to account for potential bias and uncertainty associated with use of the currently available spine age data.

The panel also highlighted the importance of using fractional ages, ensuring an understanding of the genetic origin of the samples, and of obtaining samples from a wide range of sizes. Given the sensitivity of the model to asymptotic mean length (Linf) estimates, they emphasized the need to avoid positively biasing Linf through over-representation of fish that are large for their age.

The panel also noted the value of estimating growth inside the model by fitting to age-at-length data (Lee et al. 2024). This approach addresses the biases associated with the strongly size-dependent sampling of longline fisheries, from which length at age for the youngest age classes is overestimated, and the size-dependent nature of the sampling to obtain age data.

Catch: review the treatment of the catch data, especially with regards to catch prior to 1993, when driftnet catch total amount is highly uncertain due to unspecified species attribution and spatial extent.

The panel noted that a lot of effort has gone into understanding the catch for this population but stressed the importance of continuing to work towards a comprehensive time series of data spanning the longest possible time period even if the full time period of information is not used within the assessment. The panel did note that model sensitivity to some of the uncertainty in the catch could be enhanced by the discontinuity in some of the CPUE time series. The panel stressed that further investigation into underreported or unreported catch should be conducted. This can be accomplished by comparing observed / training catch rates to logbook-based catch rates to more fully understand discarding rates and to ensure the catches reported to the WCPFC reflect fishing activity.

Size composition: review the approach for pre-treatment of size composition data (i.e., reweighting), how size composition is weighted for the likelihood function, and how decisions are made to determine which size data are included.

The panel noted that there was a data gap in the size composition before 1994 when the observed size data became available for most of the fleets. Because of the perception of poor data quality in the model prior to 1994, the length selectivity for the removals in the assessment early period was assumed to be similar to those of the same or similar fisheries in the recent period. Since there should be a difference in the age structure of MLS between the early and recent periods due to the temporal trend in the removals, this data gap should ideally be addressed by efforts to obtain size data for the early period. This also should help to estimate the initial age structure of the assessment (e.g. early recruitments and initial fishing mortality).

The review panel contributed a number of technical recommendations related to the size composition data. Of particular importance is the preliminary weighting of length composition sample sizes external to the assessment model and the rebalancing of fleet-specific residuals internal to the model. This recommendation applies to all fleets but most notably the Taiwanese fleet and the US fleet (Brodziak and Sculley 2022) which had very large standardized residuals in the 2023 assessment (ISC 2023).

There were some concerns related to the large residuals associated with fits to the length composition data from the Taiwanese longline fleet (Lee et al. 2021). The most reliable source of composition data for this fleet is likely to be weight composition data, as length data may be estimated rather than directly measured in some fishing operations. The panel determined that it would be better to convert these measured weights to lengths outside the model instead of using the more uncertain length data.

The panel identified a need for more information about factors affecting the sizes of fish observed by all fleets. These may include variation in the sizes of the fish themselves (spatial, seasonal, interannual, time of day, depth, and their interactions), and factors associated with fishing methods (e.g., set time and set depth). A simple standardization approach (length ~ location + year + set type) may help by separating different effects (Hoyle et al 2021).

CPUE: review the standardization methods and spatio-temporal structure of the CPUE data for each fleet, and the decision process for data weighting and exclusion of indices from the model.

The discussion around CPUE was wide ranging and covered a number of technical details as well as best practices. A primary focus of the panel was the discontinuity in CPUE in 1994 (Ijima and Koike 2021). Some of the requested additional model runs were focused on more fully understanding the impact of this break in the time series. In general, the panel felt that this break in the time series was critically important as it occurred at a time where the high seas driftnet fishery had ended and the CPUE transitioned from stable to declining. This break created a discontinuity in the assessment model as well as a loss of information in the model scaling with the early time period gaining information about scaling from catch and CPUE and the later period from size composition data. A number of the panel's recommendations are focused on resolving this conflict and the exploratory runs with these changes resulted in notably different model scaling and reference point outcomes.

Concern as to the quality of the Taiwanese logbook data prior to 2006 was raised given experience in other RFMO assessments. Particular attention was paid to the 2003-2005 time period which produces a CPUE pattern that is unlike that of other time series. They also noted the low logbook coverage early in the time series, and the potential for this to affect the index.

The final main challenge addressed in the reviewers' recommendations related to the spatial coverage of the fleets. The contraction of some of the fleets has resulted in a decline in spatial coverage and there is an opportunity for collaboration across nations to develop time series from multinational data which would have more comprehensive spatial coverage. This recommendation also opens the possibility of adopting an index fishery approach (Maunder et al 2020, Xu et al. 2023).

Data inputs: identify and provide recommendations on the key areas for improvement in data collection (both fishery data and biological information).

The review panel strongly supported the continuous improvement of the data inputs. Particular attention was given to ensuring that a complete time series of catch and discards continued to be developed even if the full time series was not used in the assessment model. More detailed work related to investigating the reporting of discards and the validation of yearbook reported catches was encouraged. The panel reinforced in this section the need to share information across fleets and nations to ensure that inputs intended to represent the entire spatial extent of the population are derived from spatially comprehensive data.

Other life history parameters: review the other life history parameters used (weight-length, maturity, natural mortality, stock-recruitment, etc.) for internal consistency and appropriateness for the WCNPO stock.

The panel noted the interdependency of a number of the life history parameters in the assessment model and noted that consistency between the parameters is needed. This reevaluation will also be needed as growth and maturity information is updated from the IBBS program. For example, currently the estimate of steepness depends on growth, maturity, and mortality. Mortality has some dependency on growth and longevity. There are some instances where the values used to generate estimates are not the same as the values used in the model and this should be rectified. The review panel was very supportive of the IBBS program and was encouraged by the consensus understanding that this program can bring to life-history information used in the assessment as well as to the credibility of model outputs. The panel also noted that there is the potential for the IBBS program to reveal a more complete picture of the spatial structure of life history characteristics of the stock. How such complexities are incorporated into the assessment will need careful consideration and affirming simulation work.

TOR 3 - Model configuration, assumptions and settings

Fleet structure: review fleet definitions and spatio-temporal structure of catch, CPUE, and size composition inputs.

See discussion above

Selectivity: review selectivity assumptions and settings.

The review panel contributed a number of technical recommendations related to the selectivity assumptions for the removal and index fleets. Particularly, the panel found that there were multiple fleets assumed to have an asymptotic selectivity based on the knowledge of the fishing practice (i.e., fishing gear selectivity). The panel noted that selectivity in an areas-as-fleets model is based on a combination of gear selectivity and spatial availability at size. The panel and WG confirmed the aggregated size composition data for each fleet and found that the largest sizes of MLS were caught by the Taiwanese distant water longline fleet (Fleet 18). Because of the residual pattern in the Japanese driftnet fleets, a recommendation was made to have asymptotic selectivity only in the Taiwanese distant water longline fleet (subject to review after changing the data source to weight frequency data) and that dome-shaped selectivity should be assumed for other fleets.

There were also some concerns related to the very large standardized residuals in the fit to the US and Taiwanese longline fleet length compositions. Given the complexity of the stock structure, a simple area/gear as fleet approach might not serve to reconcile the observed multimodal distribution of the observed size composition data. The panel recommended that analysts explore more flexible selectivity shapes (e.g. non-parametric age selectivity or cubic spline), particularly for the U.S. longline fleet. The panel also noted that temporal flexibility of the selectivity (e.g. time-blocks) could be considered, when based on background information from the fishery and/or the migration of the stock. However, for model stability and convergence it would be preferable to assume constant selectivity unless model fit strongly suggested this was infeasible. It was suggested that using finer population length bins, which allows for a smoother approximation to the population dynamics, could help model convergence.

Initial equilibrium conditions: review the estimation of initial equilibrium catch and fishing mortality, recommend if the BILLWG should be estimating the equilibrium conditions (as in the 2023 model) or fixing them and running sensitivity runs to evaluate the sensitivity of these conditions (as in the 2019 model).

In the 2023 assessment, the age structure of the assessment start year (1977) was determined by estimating the 14 recruitment cohorts (1964-1976) prior to the assessment start year as well as the initial equilibrium fishing mortality for Fleet 2 (Japanese longline fishery). Although the initial fishing mortality had been fixed at a certain value in the 2019 assessment, the review panel confirmed that the method used in the 2023 assessment was more appropriate, and that flexible initial conditions should be favored unless the model start year occurs prior to all appreciable catch and length information.

Uncertainty: review the approach used to represent uncertainty in model-derived management quantities, considering structural, model and input data uncertainty.

In the 2023 assessment, the uncertainties associated with alternative assumptions about life history parameters, initial conditions, stock structure, treatment of Japanese driftnet catch and the modeled production function were demonstrated through a series of sensitivity analyses.

Given the complexity of the sources of uncertainty and the inability to model these factors in a single base case model via parameter uncertainty, the panel recommends the adoption of an ensemble model approach. The ensemble should consider: growth, steepness, catch uncertainty, and conflicting time series. In addition, the assessment start year could be considered as one dimension of the ensemble, if an abundance index spanning the entire assessment period (1977 to the terminal year) cannot be developed.

Start year: review the suitability of the current start year (1977) and suggest potential alternatives, such as 1994 (the start of the high seas driftnet moratorium).

Acknowledging that the current base case model estimated the population scale during the early period (1977-1993) largely independently of the late period (1994-) due to splitting the Japanese longline CPUE index before and after 1994, the panel recommended developing a single CPUE index covering the assessment period. The WG provided an example sensitivity analysis that connected the early and late Japanese longline CPUE (S1 and S5), which resulted in a reduced systematic trend in the estimated recruitment deviations. This implied that a single time series of the abundance index throughout the assessment period could improve the obvious discrepancies between the early and late periods. In case there is a technical difficulty in developing an abundance index for the entire assessment period, the panel recommended an alternative model starting in 1994 for inclusion in an ensemble approach.

Alternative models: review the use of SS3 as the modeling software and determine if it is an adequate tool for the assessment.

The panel found no issues with the use of the SS3 modeling software given the current structure of the assessment.

TOR 4 - Model diagnostics

Review the suitability of the diagnostics used and reported for the assessment.

The panel found that the diagnostics used for this stock assessment were thorough and appropriate for this type of model. In the 2023 stock assessment, it was noted that several of the sensitivity analyses may not have converged, based on large gradients and/or the lack of a positive definite Hessian. The review panel recommends that all model runs for which results are reported should be checked for apparent convergence and that the results of unconverged models should not be reported or used.

The model fit to average length relative to the specified sample size for length composition data was used to evaluate whether weighting was consistent with model fit (‘Francis weighting’; Francis, R.I.C.C. 2011). In addition to this approach, the review panel also recommends considering the absolute scale of the standardized residuals, and further down-weighting where these values are large (e.g. >4).

The diagnostics presented included the mean absolute scaled error statistic (MASE), measuring the predictive skill of the model 1-year ahead. The review panel noted that the WCNPO MLS assessment is not conducted annually, and that rebuilding analyses extend several years into the future. Therefore, the panel recommended that a longer-term hind-cast evaluation be conducted that aligns with the assessment cycle and/or time period over which the information from an assessment is likely to be used.

Consider the diagnostics provided for the 2023 WCNPO MLS assessment and provide guidance on follow-up work where the diagnostics suggest issues, i.e., data conflicts.

The panel reviewed the suite of diagnostics provided for the assessment and found them very helpful, representing current best practices. The panel recognized that there is no single diagnostic that can be used to determine if a model is suitable, and that consideration of patterns in residuals, scale of residuals, retrospective patterns, predictive skill, and basic convergence diagnostics are all important, and should be used in tandem to determine the relative merits among contending models.

The panel discussed how to proceed when there is an apparent conflict between data sources (e.g., one CPUE series; see CPUE section above) even after efforts have been made to address the conflict in the model. In this case, the panel recommends that conflicting data be fit in alternative models and the results included in an ensemble (Maunder & Piner 2017, Schnute & Hilborn 1993).

The driver of the pattern of higher fishing mortality after the high-seas driftnet fishery was banned in 1993.

The 2023 stock assessment showed a strong pattern of higher fishing mortality after 1993, despite decreasing catch. The panel noted that because the CPUE series were separated at this part of the time-series, the two periods were scaled separately in the model and that this could be causing the patterns in both recruitment and fishing mortality which could be an artifact caused by the model misspecification. The panel requested alternative models that allowed the Japanese driftnet selectivity to be dome-shaped and included a single CPUE series (1977-2020) for the Japanese longline fleets. Results of these exploratory models suggested that these changes were promising avenues for addressing the trend and scale of fishing mortality in future assessments (see ‘Requested assessment runs’ below).

Evaluate the adequacy of the sensitivity analyses in regard to completeness and incorporation of results. Recommend improvements to the communication of the sensitivity run results (plots, tables, and/or text)

The review panel found the sensitivity analyses in the stock assessment document and presented by the working group to be comprehensive and to have addressed the major sources of uncertainty that had been identified. The presentation of the time-series' of spawning biomass, average fishing mortality (ages 3-12) and the Majuro plot showing the relative values compared to the reference points were all important in understanding the differences among the sensitivities. The panel noted that the sensitivity using the 2022 growth curve and not including a sum-to-zero constraint on recruitment deviations produced spawning biomass and fishing mortality time-series that were similar to other sensitivity runs, but a very different relative biomass. After further investigation during the review, it was identified that for model runs that do not include the sum-to-zero constraint on recruitment, the central tendency of the time-series may differ from that used to calculate the reference point. The panel recommends that in such cases the reference point calculation should be done outside the model based on the net result of the estimated recruitment deviations, or the internal calculation would need to be corrected for the average of the log recruitment deviations relative to zero. This was only a concern for that single sensitivity analysis.

TOR 5 - Comment on the proposed reference points and management parameters (e.g., MSY, F_{MSY}, SSB_{MSY}, 20%$SSB_{F=0}$); if possible and feasible, estimate values for alternative reference points or alternative methods of determining the appropriate reference years for the dynamic B0 calculations.

An overview of the methods used to determine stock status and fishing intensity was presented; which included the rationale for using a 20-year moving window for averaging the dynamic B0 as 20%$SSB_{F=0}$. The panel found that the reference point calculations were reasonable, recognizing that the choice of reference points is partially a management decision. The review panel noted that a moving 20-year average may not reflect the most current conditions when stock productivity (recruitment) is trending over time. Furthermore, the WG needs to be careful that the recent 20-year average of the dynamic B0 included the cohorts recruited prior to 1994, which were not estimated with a sum-to-zero constraint. Therefore, the panel recommended calculating and reporting both the 20-year average and the annual dynamic B0 values and associated 20% reference points for comparison. The choice of a 20-year average for the dynamic SB0 calculation was motivated by mean generation time, apparent changes in productivity in approximately 2000 and the desire for a reference point that did not change rapidly over time. The panel noted that each subsequent stock assessment conducted in the future will produce a new estimate, as the 20-year period moves forward and that it will be important to explain and delineate between the changes in the estimated spawning biomass as well as the changes in the dynamic B0. The panel suggests that the working group review the standard practice outlined by the WCPFC and consider adopting a similar approach to facilitate easier interpretation by those evaluating the results across multiple assessments.

The 2023 stock assessment used the terminal year's relative fishing mortality by fishing fleet to estimate F_{MSY} and $F_{SSB20\%}$. As the terminal year's data can sometimes be incomplete and relative Fs may vary by year, the panel suggested using an average of the last 3-5 years (not including the terminal year) to provide a more stable and reliable estimate of fishing intensity reference points. The panel discussed the relative merits of F vs. SPR to characterize fishing intensity and the potential sensitivity of F-based metrics to the distribution of fishing mortality across fleets, especially when there are different selectivity curves among fleets. Further, the choice of ages over which to consider F and whether to use the apical F or an average can affect the interpretation of the results. The panel suggested reporting both the currently calculated average F across ages 3 to 12 as well as SPR and 1-SPR so that relative scale and trends can be compared for both.

TOR 6 - Suggest research priorities to improve our understanding of essential population and fishery dynamics, necessary to formulate best management practice, with the identification of priorities to improve future assessments.

The WCNPO striped marlin stock assessment has been shown to be extremely sensitive to the specification of the growth curve. During the review there was considerable discussion about fin-ray vs. otolith ageing methods and the potential for bias in ageing methods to have a large effect on the scale and trend of the stock assessment results. The current growth curve is informed by fish that only range from 0.5 to 6 years of age, but ages 7+ are very influential in the calculation of reference points. In order to resolve these uncertainties, it is critically important to continue the broad geographic sampling of age and length that is underway in order to further refine the estimated growth curve to represent the entire stock. Further, given uncertainties and evidence from other species of large differences between ages derived from fin-rays and otoliths, age validation via temporal trends in bomb radiocarbon or other methods is needed to validate the estimated growth curve.

Sensitivity and alternative model runs conducted for this stock assessment reflect the very low information content of the CPUE time-series and length data about the scale of the striped marlin population. It is likely that additional years of relatively flat CPUE and length data will not improve this precision. Therefore, the review panel strongly recommends consideration of an external approach for estimating the scale of the population. Close-Kin Mark-Recapture (CKMR), which has been used to provide estimates of the population scale for similarly uncertain stock assessments (e.g., southern bluefin tuna; Bravington et al 2016), should be investigated. The panel recommends a power analysis in the short term to determine the magnitude of sampling that may be needed and the resources that this sampling and analysis might entail. In the longer term, this approach should be implemented, if it proves tractable for striped marlin. Sampling of genetic material across the entire geographical range of the stock should be continued while this approach is being considered. Including age-at-length data in the assessment is also likely to provide information about population age structure and scaling.

The review also discussed the complex nature of the striped marlin stock, with geographical patterns in life history characteristics, evidence for several genetic stocks across the Pacific, and the potential for mixing of some individuals among the stocks and various fisheries based on tagging projects to date. The panel recommends that additional population genetics work be pursued in order to better define stock boundaries, identify where catches could contain fish from the other assessment areas, and allow for estimation of life-history traits that are specific to the stock being assessed.

Finally, the panel recognizes that the challenges of estimating population scale and trend from CPUE and length frequency data alone given the likelihood of spatial heterogeneity in growth are considerable. The current stock assessment approach should be simulation tested to better understand how reliable population estimates may be given this type of information and different hypotheses about spatial and temporal variation in growth and life history traits.

TOR 7 - Comment on whether the stock assessment methods, results, and assessment decision process are clearly and accurately presented in the detailed report of the stock assessment.

The review panel appreciated the extensive documentation provided prior to the meeting, including the historical assessment documents, working papers associated with each of the data sources and biological analyses, and the 2023 stock assessment. The assessment provided a comprehensive suite of data summary, model fit and diagnostics. The sensitivity analyses providing the time-series of spawning biomass, fishing mortality and status (Majuro plot) were very helpful in identifying the key sources of uncertainty and their relative magnitude. The review panel thanks the working group for excellent presentations supporting the documents during the review.

Many of the data sources and biological inputs to this stock assessment rely on extensive analyses conducted externally and documented in separate working papers (e.g., CPUE by country, maturity, growth, etc.). The review panel was unable to provide a detailed review of all aspects of these analyses due to time constraints and the level of detail reported in the working documents. The panel suggests that additional detail be provided in future working papers that includes the rationale for the decisions made in developing and comparing models and sufficient diagnostics to evaluate model selection choices and their effects on the results. This is especially important for the model-based approaches used for CPUE and size data standardization. The panel encourages the use of standard practices in data preparation and encourages coordinated efforts among working group teams such that approaches are consistent for data sets used in the assessment. The panel appreciated those instances where supporting analyses had been published in peer-reviewed journals.

In the long term, the panel recommends that it is important for the institutions involved with this assessment to afford sufficient time and resources for data analyses, coordination among working group members and comparison of methods and results prior to their use in the stock assessment.

Requested assessment runs

The following is a list of additional assessment runs requested by the review panel to help them more fully understand the model behavior and refine their recommendations. All additional assessment runs showed notable differences in population scale compared to the base case.

- Change the Japanese drift net fishery to a double normal selectivity. This was requested to see if there was an improvement in the fit to the length frequency data as it was noted that larger sizes are over-predicted in the residual plots.
- Use a Japanese longline CPUE time series for fleet 01 that spans the entire time period. This was requested to link the early and late parts of the assessment model because of the termination of the driftnet fishery on the high seas.
- Due to conflicting information in the Taiwan DWLL CPUE time series, particularly the increase from 2003-2005, exclude it from the likelihood.
- Apply a 2% effort creep correction to the early part of the Japan CPUE time series. This was done to explore the potential conflict in the model between a flat CPUE time series and the higher driftnet removals.
- Use a time series of removals starting in 1950. This was requested to explore the sensitivity of the assessment scaling to the length frequency data.
- Combine the driftnet selectivity change, 1950s start, Taiwan CPUE time series removal, and full-time span of Japan fleet 01 CPUE.
- Run the age-structured production model with these changes: Combine the driftnet selectivity change, 1950s start, Taiwan CPUE time series removal, and full time span of Japan fleet 01 CPUE.

The result of these runs suggest that the working group should work toward a new base case model that considered in priority order 1) the addition of dome shaped selectivity for the Japan driftnet fishery, 2) CPUE time series for the Japan longline fleet (Survey 01 & Survey 05) that spans the entire time period should the model start in 1977, 3) the truncation of the Taiwan longline CPUE to 2006+ until potential data reporting uncertainties can be resolved.

Recommendations

TOR 1 - Review the information available on Pacific MLS stock structure and conceptual model and provide any recommendations for changing WCNPO MLS stock boundaries or to the fleet structure.

Short-term

- No short-term recommendation related to stock structure and boundaries.
- Simplify the Japanese fleet structure. The panel recommends not using the CPUE component in the current finite mixture model as a primary determinant for fleet structure. More focus should be given to operation level information to define the fleet structure as the primary objective should be to ensure a consistent fleet-specific selectivity. The key determinant of the fleet structure should be the size structure of the catch. Any fleet categorization should not simply be determined by the model output.
- For all fleets, a better understanding of the spatial and temporal structure should be explored to more fully understand the patterns of CPUE, size, targeting, fleet structure and vessel turnover.

Long-term

- Continue to pursue genetic research to more fully understand mixing between the genetically distinct population. This research should encompass the whole Pacific. There is an opportunity to utilize the genetic samples from the current IBSS program. This may require some modifications to the current sampling programs to ensure that genetic samples meet the requirements for the questions being pursued.
- The WG should explore the use of an index fishery approach which would link to an exploration of a unified CPUE analysis (Maunder et al 2020, Xu et al 2023).

TOR 2 - Model inputs, commenting on the adequacy and appropriateness of data sources and data inputs to the stock assessment.

Growth: review the approach to estimation of growth parameters and consider the implications of potential regional variations in growth

Short-term

- Understanding the interaction between growth and selectivity is perhaps the most critical aspect of improving the assessment. The review panel recognized that there is a deficiency of samples at older ages and this has consequences in the estimation of a growth model.
- The WG should explore the possibility of fitting the growth model within the assessment, together with the inclusion of conditional length-at-age data. There is the option to use a fleet currently in the assessment for selectivity or to create another fleet for this purpose.
- The panel notes that the current IBSS project is a significant step in resolving growth uncertainty and this project should continue to be well supported. The spatial evaluation of growth is supported by the panel and careful consideration to how spatial growth applies to fleet structure and the interpretation of length data is needed.
- Should there be delays in the production of a new growth curve then effort needs to be put into resolving the potential age bias of using spines to produce the growth curve.

Long-term

- There needs to be continuing effort to resolve growth curves across space, sex, time, and genetic origin.

Catch: review the treatment of the catch data, especially with regard to catch prior to 1993, when the driftnet catch level is highly uncertain due to unspecified species attribution and spatial extent.

Short-term

- Efforts should be made to continue to improve the full historical time series of catch and associated uncertainty, even if the full catch time series is not used in the model. This would include an evaluation of the reported catch by other nations and whether or not reported catches make sense given fleet effort and area specific catch rates of reference nations. As an example, given the fishing effort of one nation in specific locations, if the CPUE of another nation was used would you estimate the same scale of catch.
- Efforts to improve the characterization (spatial location) and associated uncertainty in the driftnet data would increase confidence in these data and should be pursued.
- Efforts should be taken to more fully understand discarding in all fleets. A potential starting point is the comparison of discarding between observed / training vessels and logbooks.
- The reporting of MLS may not have been consistent across vessels and fleets. Efforts should be made to more fully understand if there are reporting biases in all

fleets. A potential starting point is the comparison of MLS reporting rates between observed / training vessels and logbooks, at a reasonable spatial and temporal resolution.

Long-term

- None

Size composition: review the approach for pre-treatment of size composition data (i.e., reweighting), how size composition is weighted for the likelihood function, and how decisions are made to determine which size data are included.

Short-term

- Length frequency should be explored in detail at the national level for full historical time series. Methodologically this would include the application of a standardization method to understand the factors that cause spatial and temporal variability in the size data. This would allow a more complete understanding of the spatial and temporal nature of the data, how it relates to population size structure and fleet structure, and the standardization of size data should an index fishery be used.
- The US shallow set longline length data should be removed from the length composition data to resolve the discontinuity in the selectivity as a result of the fishery changes that occurred in the early 2000s.
- Weight data should be used for the Taiwanese fleet instead of the length data since the length data may not always be measured, but can be estimated. The conversion from weight to length should be done outside the model using the fleet-specific length weight relationship. The panel acknowledged that the length-weight relationship in the model is the Taiwanese fleet's relationship but still recommended doing this conversion external to the assessment model. Consider inputting these data as generalized compositions to allow for bin sizes that can span potentially rounded weights.
- Bootstrapping or a model-based approach should be used to establish initial sample sizes external to the model to account for the properties of the underlying data (Thorson et al. 2017, Stewart et al. 2014)

Long-term

- None

CPUE: review the standardization methods and spatio-temporal structure of the CPUE data for each fleet, and the decision process for data weighting and exclusion of indices from the model.

Short-term

- Make every effort to not split CPUE series (see discussion in Hoyle et al. 2024).
- If the time period of the assessment includes the period with high seas driftnet fishing as early longline fishing, effort should be made to standardize Japanese longline CPUE without splitting its time series since this could be a data source of the population dynamics bridging before and after 1994.
- Investigate the potential for monofilament branchlines and other technological advances that may have affected Japanese CPUE during the 1980s.
- Use the annual variance estimates from the CPUE standardization as a starting point for model inputs. These can be rescaled but should reflect the differences in precision among years within each series. Where necessary, either iterate or estimate additional variance such that the model fit (RMSE - root mean squared error) is consistent with the average input standard error by fleet.
- For the Taiwanese CPUE series, omit data before 2006 until there is confidence that these data can be used to provide reliable information. Explore the data for the period prior to 2006 to identify factors that may have caused the jump in CPUE 2003-2005.
- When there are conflicting CPUE series covering the same period, and these conflicts cannot be resolved, they should be included in alternative model scenarios rather than combining them in the same model (Francis 2011, Schnute and Hilborn 1993, Hoyle et al 2024). This assumes that consideration has been given to the nature of the conflict and that the conflict is not due to a model misspecification (e.g., an error in the fleet designation)
- Provide full diagnostics for all CPUE series that may be included in the assessment, including residual plots, effect plots (i.e., the effect of the covariate on the expected CPUE), and influence (i.e., the impact of the covariate on the index over time) plots.

Long-term

- Develop joint CPUE series across nations and multiple fleets, to address the following issues:
    - Provide indices that cover the majority of the stock across the whole time series. Such an approach would help to limit the effect of contracting spatial coverage in some of the fleets. It is unclear what impact declining spatial coverage has on the estimation of annual random fields even when only a portion of the random fields are used in standardization.

  - Provide a single series using consistent methods for data cleaning and model fitting, rather than multiple series that may conflict.
  - Develop a shared understanding among the collaborators.

Data inputs: identify and provide recommendations on the key areas for improvement in data collection (both fishery data and biological information).

Short-term

- As previously mentioned, continue to improve the full historical time series of catch, regardless of whether the full catch is used in the model, more fully understand and characterize discarding in all fleets, more fully understand and account for reporting discrepancies in all fleets.

Long-term

- The range contraction of some of the fleets highlights the importance of information sharing across fleets and nations. Consideration is needed to ensure that biological information used to make inferences at the population level has the appropriate spatial coverage.
- Age sampling for a fleet could provide better estimates of population scale.

Other life history parameters: review the other life history parameters used weight-length, maturity, natural mortality, stock-recruitment, etc.) for internal consistency and appropriateness for the WCNPO stock.

Short-term

- There have been developments in how to estimate life-history parameters. New life history data has been and is being collected. Revisiting the life history values used in the model in light of this new information and approaches should be considered, with the goal of internal consistency across the development of the full suite of parameters used in the model.
- There is evidence to suggest a west to east difference in maturity. Note also that samples from spawning areas may not be representative of the population-level maturity ogive. Consideration should be given to addressing these differences by establishing a CPUE-weighted maturity ogive (Farley et al. 2014).
- The IBBS project collecting information on growth, maturity, and genetics is a significant step to helping understand the spatial distribution of the underlying life history characteristics and needs continued support to ensure success.

- The steepness prior should be updated to take into account changes in the input parameters. It is important to allow for uncertainty in the values of all input parameters and uncertainty about the structure of the stock recruitment relationship, which may result in a flatter prior.
- New information on growth and maturity will require natural mortality to be updated. Given the potential for spatially varying life history values, spatial consideration will need to be given to how to weight such information appropriately in the estimation of these values.

Long-term

- None.

TOR 3 - Model configuration, assumptions and settings

Fleet structure: review fleet definitions and spatio-temporal structure of catch, CPUE, and size composition inputs.

Short-term / long-term

- See recommendation in previous sections

Selectivity: review selectivity assumptions and settings.

Short-term

- The population level size bins should be reduced from 5 cm (to 1 or 2 cm) as this has the potential to provide a smoother likelihood surface where selectivity is highly dome-shaped. This does not mean that the bins for the fleet specific length frequency data need to be reduced.
- Consider setting the parameter defining the width of the top of domed selectivity curves to span at least 2 population size bins.
- Given the structure of the area-implicit assessment model, a fleet assumed as an asymptotic selectivity should be chosen based on the observed data (i.e. empirical selectivity method). The panel recommends fleet 18, the TWN DWLL, as this fleet has the largest observed fish, though this may be revisited after the data source has been changed from length frequency to weight frequency observations.
- Aim to remove time blocks from the selectivity parameterization of the Japanese and US fleets.
- Explore more flexible selectivity for the US longline fleet in order to better fit the bimodal size-composition information. There is also the potential to use an age-

based selectivity of this fleet given the apparent length-based modal progression seen in the data.
- Review parameter and asymptotic variance estimates from all selectivity curves and reparametrize where there is no apparent information (e.g. Size_DblN_descend_se_F16_US_LL(16) in the 2023 base case model).
- Consider reducing the range of the lower and higher bounds for selectivity parameters (especially those that are logistic transformed) and adjusting the phasing to achieve more reliable convergence. Ideally this would reduce/remove the need for a .par file and assist profile and jitter analysis convergence.

Long-term

- None

Initial equilibrium conditions: review the estimation of initial equilibrium catch and fishing mortality, recommend if the BILLWG should be estimating the equilibrium conditions (as in the 2023 model) or fixing them and running sensitivity runs to evaluate the sensitivity of these conditions (as in the 2019 model).

Short-term

- As in the 2023 assessment, the initial conditions should be estimated. This includes recruitment deviations for the initial age structure and initial F.

Long-term

- None

Uncertainty: review the approach used to represent uncertainty in model-derived management quantities, considering structural, model and input data uncertainty.

Short-term

- The review panel felt that the workgroup did a good job of presenting the sensitivity of the results.
- The panel recommends the adoption of an ensemble model approach. The ensemble should consider: growth, assessment start year, steepness, catch uncertainty, and conflicting time series. Growth should be ensembled based on observed spatial differences in growth if they are identified to exist. Model start date should be ensembled if there is no possibility of linking the early and late parts

of the CPUE time series. (A 1977 + and a 1994+ model). Steepness could be ensembled by selecting three steepness values that represent a plausible range of steepness for the species. Uncertainty in catch should be incorporated by including ensembles with high/low, and best estimates. Should conflicting (e.g., JPN vs. TWN) CPUE time series not be resolved the assessment should be ensembled over these conflicting time series.

- The full time series (a 1952+ model) of catch should be included at least as a sensitivity run.

Long-term

- Some simulation work would be required to understand the details of how, if it is present, spatial differences in life history characteristics should be accounted for within an assessment model. Best practices have not been determined.

Start year: review the suitability of the current start year (1977) and suggest potential alternatives, such as 1994 (the start of the high seas driftnet moratorium).

Short-term

- The panel would prefer the model to begin in 1977, combined with improved long-term catch time series, and with a CPUE series the same length as the assessment period. If discrepancies are not resolved between early and late assessment periods then a model such as the one starting in 1994 was recommended for inclusion in an ensemble approach.

Long-term

- None

Alternative models: review the use of SS3 as the modeling software and determine if it is an adequate tool for the assessment.

Short-term

- The review panel supports the use of an age-structured production model to provide a good diagnostic tool for the current assessment framework.

Long-term

- There is a notable reality that the reliance of the assessment on length composition data and the nature of the growth pattern results in challenges in determining the scale of the population. Consideration should be given to the use of close-kin approaches to estimate population scale.

TOR 4 - Model diagnostics

Review the suitability of the diagnostics used and reported for the assessment.

Short-term

- Check convergence of all models, to avoid incorrect inferences from models that have not converged. The results of models that have not converged should not be reported, or included as sensitivity runs.
- Given that stock rebuilding has been evaluated based on future projection, longer term (i.e., an assessment cycle or generation time) hind-casting could be conducted to determine the prediction skill of the model.
- Continue to use a broad suite of metrics to characterize model suitability.
- Consider the absolute scale of the residuals from diagnostic plots and reweight data sets accordingly (e.g., some of the size data residuals were notably larger than others).

Long-term

- None

Consider the diagnostics provided for the 2023 WCNPO MLS assessment and provide guidance on follow-up work where the diagnostics suggest issues, i.e., data conflicts.

Short-term

- The panel felt that the diagnostics provided were helpful.
- As noted earlier, data sets that are in conflict such as CPUE time series should not be included in the same model but accounted for in an ensemble approach.

Long-term

- None

The driver of the pattern of higher fishing mortality after the high-seas driftnet fishery was banned in 1993.

- The flat CPUE through the high catch period is scaled independently of the later period where length frequency data is fit. This appears to have forced the model into a domain where it needs to be highly responsive to recruitment deviations as well as fishery removals to fit the data. The adoption of the review panel's recommendations related to selectivity and continuity of time series should change this pattern. The response of the model to these changes as demonstrated by the additional runs requested suggest these are productive areas of exploration.

Evaluate the adequacy of the sensitivity analyzes in regard to completeness and incorporation of results. Recommend improvements to the communication of the sensitivity run results (plots, tables, and/or text)

Short-term

- The panel found the sensitivity analysis presented by the workgroup to be quite comprehensive.
- The use of non-zero sum recruitment deviations is a reasonable sensitivity check. However, the use of non-zero sum recruitment deviation can cause a difference between the time-series and the reference point and therefore the reference point would need to be calculated based on the net result of the recruitment deviations.

Long-term

- None

TOR 5 - Comment on the proposed reference points and management parameters (e.g., MSY, F_{MSY}, SSB_{MSY}, 20%$SSB_{F=0}$); if possible and feasible, estimate values for alternative reference points or alternative methods of determining the appropriate reference years for the dynamic B0 calculations.

Short-term

- Recommend calculating and reporting both the 20-year moving average as well as the annual dynamic B0 so that the trends can be compared.
- Recommend averaging relative Fs over the last 3-5 years but not including the terminal year for the calculation of $F_{SSB20\%}$ rather than using the terminal year.
- The panel suggests continued reporting of additional status metrics such as %SPR or 1-SPR.

- The panel recommends reviewing the standards outlined by the WCPFC and considering the adoption of the same approach.

Long-term

- None

TOR 6 - Suggest research priorities to improve our understanding of essential population and fishery dynamics, necessary to formulate best management practice, with the identification of priorities to improve future assessments.

Short-term

- The development of an age validated growth curve is essential to improve the reliability of the assessment model.
- Consider exploring requirements for CKMR.

Long-term

- Continue to develop a more comprehensive understanding of the genetic structure of the entire Pacific as well as the genetic composition of the removals.
- Implement CKMR approaches should they prove to be tractable for the population.
- Simulation work to understand the best assessment approaches to deal with a complex fishery and life history spatial structure.

TOR 7 - Comment on whether the stock assessment methods, results, and assessment decision process are clearly and accurately presented in the detailed report of the stock assessment.

Short-term

- The review panel found the reporting of the process to be well documented, appreciated the extensive supporting material, and was highly appreciative of the effort.
- Some of the supporting documentation in the working group papers would benefit from greater detail in the decisions made and well as the diagnostics used. It would be helpful to have this information within these documents. This is important for the development of both CPUE time series and size data. Encourage analysts to follow standard guidelines for documenting these analyses, and development of

standards for data areas without them. We also encourage coordination across groups so that they follow similar approaches.

Long-term

- Recommend working with the institutions involved with the assessment and reporting process to ensure that personnel are afforded the time to fully explore data analyses and report comprehensively on the findings.

Literature cited

Bravington, M., Grewe, P. & Davies, C. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat Commun* **7**, 13162 (2016). https://doi.org/10.1038/ncomms13162

Brodziak J and Sculley M. Observed length composition data for striped *marlin, Kajikia audax*, in the Hawaii longline fishery, 1994-2020. 2022. Working paper submitted to the ISC Billfish Working Group Intersessional Workshop, March 2022. ISC22/BILLWG-01/01

Ijima H and Kanaiwa M. Size-dependent distribution of Pacific Striped Marlin (*Kajikia audax*) : The analysis of Japanese longline fishery logbook data using the finite mixture model. 2019. Working paper submitted to the ISC Billfish Working Group Intersessional Workshop, January 2019. ISC19/BILLWG-01/09

Ijima H and Koike H. CPUE Standardization for Striped Marlin (*Kajikia audax*) using Spatio-Temporal Model using INLA. 2021. Working paper submitted to the ISC Billfish Working Group Intersessional Workshop, December 2021. ISC21/BILLWG-03/01

ISC. Stock assessment report for striped marlin (*Kajikia audax*) in the western and central north Pacific Ocean through 2020. Report of 23rd meeting of the International Scientific Committee for Tuna and Tuna-like species in the North Pacific Ocean, Annex 14.

Farley JH, Eveson JP, Krusic-Golub K, Kopf K. Southwest Pacific striped marlin population biology (Project 99). Information paper submitted to the 17th regular session of the scientific committee of the Western and Central Pacific Fishery Committee. 2021: WCPFC-SC17-2021/SA-IP-11

Farley JH, Hoyle SD, Eveson JP, Williams AJ, Davies CR, Nicol SJ. 2014. 'Maturity ogives for south Pacific albacore tuna (*Thunnus alalunga*) that account for spatial and seasonal variation in the distributions of mature and immature fish'. PLoS One. DOI: 10.1371/journal.pone.0083017.]

Farley JH, Krusic-Golub K, Eveson JP. Updating age and growth parameters for South Pacific albacore (Project 106). Information paper submitted to the 17th regular session of the scientific committee of the Western and Central Pacific Fishery Committee. 2021: WCPFC-SC17-2021/SA-IP-10

Francis R.I.C.C. Data weighting in statistical fisheries stock assessment models. 2021. Canadian Journal of Fisheries and Aquatic Sciences 68: 1124-1138

Hoyle SD, Campbell RA, Ducharme-Barth ND, Grüss A, Moore BR, Thorson JT, Tremblay-Boyer L, Winker H, Zhou S, Maunder MN. Catch per unit effort modelling for stock assessment: a summary of good practices. 2024. Fisheries Research. https://doi.org/10.1016/j.fishres.2023.106860

Hoyle SD, Chang ST, Fu D, Itoh T, Lee SI, Lucas J, Matsumoto T, Yeh YM, Wu RF, Lee MK. Review of size data from Indian Ocean longline fleets, and its utility for stock assessment. 2021. Working paper submitted to the 23rd session of the Indian Ocean Tuna Commission Working party on Tropical Tuna. IOTC-2021-WPTT23-07.

Lam CH, Tam C, Lutcavage M. Connectivity of Striped Marlin from the Central North Pacific Ocean. Frontiers in Marine Science, 2022; 9. DOI=10.3389/fmars.2022.879463

Lee HH, Maunder MN, Piner KR. Good Practices for estimating and using length-at-age in integrated stock assessments. Fisheries Research, 270, 2024. ISSN 0165-7836, https://doi.org/10.1016/j.fishres.2023.106883.

Lee K, CH Yi, Wang WJ, Lu CY, and Chang YJ. Catch and size data of striped marlin (*Kajikia audax*) caught by the Taiwanese fisheries in the Western and Central North Pacific Ocean during 1958-2020. Working paper submitted to the ISC Billfish Working Group Intersessional Workshop, December 2021. ISC/21/BILLWG-03/05

Mamoozadeh NR, Graves JE, McDowell JR. Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species. Evol Appl. 2020; 13: 678–699. https://doi.org/10.1111/eva.12892

Martinez J. Stock Composition Of Striped Marlin (*Kajikia Audax*) In The Central North Pacific Ocean Inferred By Analyses Of Genome-Wide Molecular Markers. Dissertations, Theses, and Masters Projects. 2021; William & Mary. Paper 1638386006. https://doi.org/10.25773/wn5a-5r23

Maunder MN, Piner KR. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research. 2017 Aug 1;192:16-27.

Maunder MN, Thorson JT, Xu H, Oliveros-Ramos R, Hoyle SD, Tremblay-Boyer L, Lee HH, Kai M, Chang SK, Kitakado T, Albertsen CM. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fisheries Research. 2020 Sep 1;229:105594.

Schnute, J.T. and Hilborn, R. Analysis of contradictory data sources in fish stock assessment. 1993, Canadian Journal of Fisheries and Aquatic Sciences 50: 1916-1923.

Stewart, I.J., and Hamel, O.S. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. 2014. Canadian Journal of Fisheries and Aquatic Sciences 71(4): 581-588. doi:10.1139/cjfas-2013-0289.

Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. 2017. Fisheries Research 192: 84-93. doi:10.1016/j.fishres.2016.06.005.

Xu H, Maunder MN, Minte-Vera C, and Lennert-Cody C. Exploratory analysis for the stock assessment of Bigeye tuna in the Eastern Pacific Ocean. The meeting document submitted to the 14th meeting of the Inter-American Tropical Tuna Commission Scientific Advisory Committee, 2023. Document SAC-14-05