Peer review of
Stock assessment of striped marlin in the Western and Central North Pacific Ocean (2012)

Prepared by:

Panayiota Apostolaki

Assignment undertaken for:

Center for Independent Experts

December 2012

Table of Contents

Executive Summary **2**

Background **7**

Description of the Reviewer's Role in the Review Activities **8**

Summary of Findings **8**

TOR 1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data 8

TOR 2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for 9

TOR 3. Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status 13

TOR 4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status 13

TOR 5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices 15

Conclusions and Recommendations **16**

Appendix 1: Bibliography of materials provided for review **20**

Report reviewed 20

Additional material 20

Executive Summary

Striped marlin (*Tetrapturus audax*) is commonly harvested multi-nationally from commercial and recreational fisheries in Pacific Ocean regions. Until recently, striped marlin in the North Pacific was assessed as a single stock. However, the definition of the stock was revised and two stocks have been identified; one in the Western and Central North Pacific Ocean (WCNPO) and one in the Eastern Pacific Ocean (EPO). The latest stock assessment for the WCNPO stock was produced in 2012. The review of that stock assessment is presented here.

The model chosen to simulate the dynamics of the population was a seasonal, length- and age-structured, forward-simulation model. This was part of the Stock Synthesis software that was used for the stock assessment. Catch and CPUE data as well as catch-at-length information were used to estimate the status of the stock.

The results of the calculations showed a long-term decline of the population. Its current (2010) size is estimated to be 15% of its virgin size while the current size of the spawning stock (i.e. spawning stock biomass) is less than 6% of its virgin size. The latter figure is equal to 35% of the spawning stock biomass that can produce maximum sustainable yield (MSY). So, the population is below its size at MSY. Fishing mortality is 24% more than the fishing mortality that corresponds to MSY. These results suggest that, relative to MSY-based quantities, the population is overfished and overfishing is currently occurring. The results of model projections showed that a reduction in fishing mortality is likely to lead to an increase in the size of the population.

Dr Panayiota Apostolaki was commissioned to provide an independent review of the stock assessment report entitled "Stock assessment of striped marlin in the Western and Central North Pacific Ocean" authored by Hui-Hua Lee, Kevin R. Piner, Robert Humphreys and Jon Brodziak, in accordance with the SoW. The review was desk-based and this document provides the outcome of this review.

The software used for both the stock assessment and projection is appropriate for the stock considered and the parameterisation of the model is valid. Overall, the stock assessment is a good first attempt to capture the dynamics of this newly defined stock and make use of a wide range of information to feed into the calculations. However, although the general approach is sound, parts of the analysis could be strengthened or need to be revisited for both the stock assessment and projections.

Specific comments and recommendations under each Term of Reference are shown below:

TOR 1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data

Overall the choice of the assessment method (Stock Synthesis) is appropriate and the methods are used properly. The model used to simulate the dynamics of the population simulates key biological processes such as recruitment, growth, natural mortality and provides flexibility in how biological processes are simulated. A flexible model is also used in SS to simulate the dynamics of the fishing fleets and their interaction with the fish stock. Nevertheless, the current stock assessment does not take advantage of some additional features that Stock Synthesis has and would be useful for this assessment.

Recommendation 1.1. Stock Synthesis allows for Bayesian estimation; it is recommended that this feature of the software is used to provide more flexibility in the parameters that can be estimated.

Recommendation 1.2. Work to incorporate information about movement and spatial heterogeneity in the distribution of the population into stock assessment discussions is recommended. That would provide a better understanding of the dynamics of the stock and help scientists decide whether a more detailed model is needed to simulate the dynamics of the stock.

TOR 2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner-recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for

The model configuration is a good representation of the dynamics of the stock and fisheries and associated uncertainty. Assumptions and input data are generally good reflection of current knowledge of the dynamics of striped marlin (but there are some exceptions). The values of input parameters are valid and have been chosen after consideration of wide range of sources and based on appropriate analysis. The model estimates a number of parameters and further sensitivity analysis is used to provide information about how varying values of the fixed parameters might affect model results. This approach covers the different areas of uncertainty well. However, there are some aspects of the analysis that could benefit from further sensitivity analysis or from treating more biological parameters as uncertain. The methods used to standardize CPUEs are appropriate but it is noted that the adopted models could not fully explain the variability in the data. Recommendations for additional work to address those issues are provided below.

Recommendation 2.1 The authors need to explain why the values of the CVs for length at age estimates used in the stock assessment differ from those calculated and presented in the Billfish working group workshop in December 2011 and what impact that change in the CV values had in the model results.

Recommendation 2.2. The authors need to confirm/check whether the value for *k* in the growth equation used in the stock assessment model is 0.24 or 0.34.

Recommendation 2.3 Given the sensitivity of the model to values of natural mortality, I recommend that another sensitivity run is done assuming that the minimum age for adults is lower than 4 years (e.g. 3 years) to test the effect of that assumption on model fit and results.

Recommendation 2.4 I would recommend two additional runs in which *h* is estimated by the model; one using the distribution calculated by Brodziak and Mangel as a prior for *h* and another run in which a less informative prior would be used. To do this, the authors need to take advantage of the features of the Stock Synthesis that allow Bayesian statistics to be used for the calculations.

Recommendation 2.5 The authors need to check the effect of the assumption used about the sum of deviates in annual recruitment on the pattern of deviates found (and the recruitment values calculated).

Recommendation 2.6. An explanation is needed as to whether the wide deviation in the recruitment values seen in the predicted recruitment is biologically plausible and should be permitted in the model.

Recommendation 2.7. The values of the CVs for the estimated recruitment differ considerably from values found in earlier calculations. The reason behind such noticeable difference should be explained.

Recommendation 2.8 Further work is needed to assess the influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing strategies, etc.) on CPUE values and improve the explanatory power of the models used to standardise CPUEs.

Recommendation 2.9 More information should be provided about model convergence including information about whether any of the estimated parameters converged at the minimum or maximum allowed value (hit the boundaries) and diagnostics/plots used to check MCMC algorithm convergence.

Recommendation 2.10 A table to present the values of life history parameters used in the model and clarify whether they are fixed or estimated should be included in the report. A table with the CPUEs used and their original CVs is also needed.

TOR 3 Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status

The stock assessment calculated reference points based on MSY that included yield, fishing mortality (calculated as the average of fishing mortality on ages 3 and older), population size, spawning biomass and the spawning potential ratio. Those values are then used in the presentation of the assessment results. Those management parameters are appropriate and are commonly used to describe current and future status of stocks. However, it is not clear why fishing mortality at MSY was calculated using the fishing mortality for fish of age 3 or older only.

Recommendation 3.1 The authors need to explain why they chose to use values of fishing mortality for fish of age 3 or older to calculate the average fishing mortality at MSY and also present annual fishing mortality. This will not capture changes in fishing pressure on young fish that might increase as the size of the population decreases.

Recommendation 3.2 Some of the reference points would need to be recalculated if the additional runs for steepness suggested under TOR2 estimate a different value for steepness from that used in the base case run.

TOR 4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status

The model used to do the projections is appropriate and describe the dynamics of the stock and fleet well. Some of its features are less flexible than those used in the Stock Synthesis, which meant that the output of the stock assessment had to be processes/simplified before being used for the projections. The adjustments made are unlikely to have any significant impact on the calculations and results of the model. It would be preferable though if the same model structure was used to do the stock assessment and projections to avoid introducing any errors as part of the transition; future work needs to address that. The application of the method is valid and captures an adequate number of scenarios to explore different levels of fishing pressure and management approaches. However, there are some assumptions and input values that are not consistent and need to be revisited. I have provided more information below.

Recommendation 4.1 I would recommend that the same software/model is used to do the stock assessment and projections.

Recommendation 4.2 I recommend that the grouping of selectivities is done based on estimated selectivities not the starting assumptions.

Recommendation 4.3 The values for selectivity shown in Table 9 of the stock assessment report need to be checked and explain why the selectivity chosen does not reflect the catch-at-age behaviour shown in Figure 16 of the report.

Recommendation 4.4 Some of the projections need to be repeated to use only a single assumption about exploitation in years 2010 and 2011.

TOR 5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices

Recommendation 5.1 I recommend further work to better understand and describe the recruitment process as well as the level of variation in recruitment that is biologically plausible and inform population dynamics model formulation and future stock assessments.

Recommendation 5.2 If there is fishery independent information that has not become available for the 2012 stock assessment then priority should be given to making it available for future stock assessments.

Recommendation 5.3 The influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing strategies, etc.) on CPUE values needs to be explored further. Also, further work is recommended to explore the factors that might influence or contribute to the area-dependent changes in catchability found in the stock assessment.

Recommendation 5.4 Work to incorporate information about movement and spatial heterogeneity in the distribution of the population into stock assessment discussions is recommended.

Background

Striped marlin (*Tetrapturus audax*) is commonly harvested multi-nationally from commercial and recreational fisheries in Pacific Ocean regions. Previous striped marlin stock assessments provided estimates of stock status for striped marlin in the whole north Pacific. The last stock assessment for striped marlin in the whole North Pacific was conducted in 2007. Since then, considerable work on the biology of the species has been completed and the definition of the stock was revised to reflect results of genetic studies. Based on those studies, two stocks were identified:

a) Western and Central North Pacific Ocean (WCNPO) stock West of 140 W and north of the equator
b) Separate Eastern Pacific Ocean (EPO) stock. East of 140 W and north of the equators

The latest stock assessment for the WCNPO stock was produced in 2012. The review of that stock assessment is presented here. The assessment was conducted in two phases; the stock assessment was conducted in December 2011 and the stock projections were developed over April 2-9, 2012. The analysis undertaken for both phases and relevant results are presented in the document entitled "Stock assessment of striped marlin in the Western and Central North Pacific Ocean" which is the subject of this review.

The objectives of the assessment were to (1) understand the dynamics of WCNPO striped marlin by estimating population parameters such as time series of recruitment, biomass and fishing mortality, (2) determine stock status by summarizing results relative to MSY-based limit reference points, and (3) formulate scientific advice on conservation needs for fisheries managers by constructing a decision table based on projections using both constant catch and constant fishing mortality scenarios.

The Stock Synthesis software was used for the stock assessment. The model chosen to simulate the dynamics of the population was a seasonal, length- and age-structured, forward-simulation model. Catch and CPUE data as well as catch-at-length information were used to estimate the status of the stock.

The results of the calculations showed a long-term decline of the population. Its current (2010) size is estimated to be 15% of its virgin size while the current size of the spawning stock (i.e. spawning stock biomass) is less than 6% of its virgin size. The latter figure is equal to 35% of the spawning stock biomass that can produce maximum sustainable yield (MSY). So, the population is below its size at MSY. Fishing mortality is 24% more than the fishing mortality that corresponds to MSY. These results suggest that, relative to MSY-based quantities, the population is overfished and overfishing is currently occurring. The results of model projections

showed that reduction in fishing mortality is likely to lead to an increase in the size of the population.

Three Center for Independent Experts (CIE) reviewers have been commissioned to provide an impartial and independent peer review (desk review) of this stock assessment in accordance with the SoW and ToRs listed in Appendix 2. This document presents my comments on the stock assessment report. Further details on the reviewer's role and the review request of the CIE are presented below and in Appendix 2.

Description of the Reviewer's Role in the Review Activities

I was contracted by the CIE to provide an independent review of the document entitled "Stock assessment of striped marlin in the Western and Central North Pacific Ocean" authored by Hui-Hua Lee, Kevin R. Piner, Robert Humphreys and Jon Brodziak in accordance with the SoW (listed in Appendix 2). As part of this review, I also read background documents and reports that are relevant to the reviewed stock assessment report (listed in Appendix 1). The review was desk based and took place in November and December 2012. This document provides the outcome of that review.

Summary of Findings

TOR 1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data

The stock assessment used the Stock Synthesis (SS) software (http://nft.nefsc.noaa.gov/Stock_Synthesis_3.htm) which includes a seasonal, length- and age-structured, forward-simulation population dynamics model. The model used to simulate the dynamics of the population is appropriate and provides flexibility in how biological processes are simulated. The model simulates key biological processes such as recruitment, growth, natural mortality, etc. A flexible model is also used in SS to simulate the dynamics of the fishing fleets and their interaction with the fish stock. The latter provides flexibility in the description of fleet selectivity and simulates multiple fleets. Seasonality has also been captured in the calculations. A number of observed quantities can be used by the statistical sub-component to adjust the simulated dynamics using likelihood; for the striped marlin stock assessment, the model was fit to CPUE data as well as to data on length composition of the catches.

Overall the choice of the assessment method is appropriate and the methods are used properly. Nevertheless, the current approach does not take advantage of some additional features that Stock Synthesis has and would be useful for this stock assessment. In the stock assessment report the authors note that Stock Synthesis allows for Bayesian estimation and use of the MCMC algorithm. However, priors are not used in the analysis for the estimated parameters (or if they do, it is not said so in the report) which would be useful for this stock assessment (I have suggested a potential use in TOR 2).

Simulating the dynamics of a stock using more than one area is also another feature that could be useful but not used in the current stock assessment. Analysis of the catch data suggests that, for the same fleet, catchability or catch composition differs among different parts of the area that the stock occupies. A spatially disaggregated model can simulate biological or other processes that might need to be defined at a finer spatial scale to describe those patterns. However, it is not clear whether there is enough information about the dynamics of the stock in different parts of the area it occupies although distinct spawning grounds have been identified and tagging studies have also been undertaken. Further work to incorporate such information into stock assessment discussions is recommended. That would provide a better understanding of the dynamics of the stock and whether a more detailed model is needed to simulate the dynamics of the stock.

It is not clear why the authors chose to use a separate model to do projections despite the fact that Stock Synthesis has a module for forecasting. Notwithstanding that, the model chosen for projections is appropriate for the data available and simulates all key processes in adequate detail (more comments on this are provided under TOR 4).

TOR 2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for

Assumptions and input data are a good reflection of current knowledge of the dynamics of striped marlin. There are a couple of exceptions to that which I have listed below. The values of input parameters have been chosen after consideration of wide range of sources and their calculation has been based on appropriate analysis. The model configuration puts emphasis on allowing the model to adjust parameter values that determine the dynamics and characteristics of the fleet. From the biological processes, recruitment is estimated; for all other processes the variability is given as input value. The authors did suggest that it would be difficult to estimate

certain biological parameters hence the chosen fixed values. Sensitivity analysis is used to provide information about how varying values of the fixed biological parameters might affect model results. This approach covers the different areas of uncertainty well. However, I have indicated below some aspects of the analysis that could benefit from further sensitivity analysis or from treating more biological parameters as uncertain. The methods used to standardize CPUEs are appropriate but it is noted that the adopted models could not fully explain the variability in the data. So, further work on CPUE standardization is needed.

Growth

a) The report of the Billfish working group workshop in December 2011 (2012a) describes further work that was done to calculate the CVs on length at age for youngest and oldest fish based on the aging study of Sun et al. (2011). The report reads "The empirical size at age CVs were consistent with the values used in the assessment (i.e., WP-6); the CVs were larger for younger fish (8%) and the CVs were smaller for older fish (4%)." However, the values for CV used in the stock assessment are 14% and 8% respectively. The authors need to explain why different values were used in the stock assessment and what impact they had in the model predictions.

b) The value of *k* for the length at age relation is not specified in the stock assessment report. The value of *k* calculated in Sun et al. (2011), which is the study that the stock assessment used to get those parameters, was 0.34. However, the value for *k* that is specified in Table 3.2 in the report of the BILLWG in December 2011 (2012a) is 0.24. Is this a typo? Please, confirm/check that the right value for this parameter has been used in the stock assessment model.

Natural Mortality

As part of the sensitivity analysis conducted, the model was run assuming that natural mortality for the adult population was higher or lower than the value used for the base case run (0.38). However, the minimum age of the adult population has remained the same (4 years) as did the mortality of younger fish for all the sensitivity runs. Given that the $L_{50\%mature}$ is 2.3 years a lower age than 4 years for adult fish is a plausible assumption. Given the sensitivity of the model to values of natural mortality the effect of that change (which would also change the values of natural mortality at age for juvenile fish) on model prediction and model fit should be evaluated. So, I recommend that another sensitivity run is done with the alternative parameterisation.

Steepness of the stock-recruitment function

The model uses a fixed value for *h* which is equal to 0.87. That value is the mean value estimated by Brodziak and Mangel (2011). There are two points to highlight relating to the assumptions used for the steepness *h* of the Beverton-Holt curve:

a) The value of *k* that Brodziak and Mangel (2011) used in the growth model for their calculations was 0.34. If the value for *k* used in the stock assessment is different (see comment above about growth parameters) then the assumptions about growth used in the two studies are different and the value used for steepness should be revised.
b) More importantly, it is not clear why the choice was made to fix the value of steepness. From sensitivity runs, it was clear that the value of steepness is one of the factors to which the model was sensitive. The other factor to which the model was sensitive was natural mortality and the values of it were also fixed. The explanation given for fixing the value of the steepness was that it is unlikely that the model could estimate it. However, it is not clear whether such estimation was attempted. Furthermore, Brodziak and Mangel (2011) have provided a distribution for the value of that parameter that is very informative. It is important to clarify whether the value of steepness can be estimated by the stock assessment model and what changes in the predictions of the model that would cause. Therefore, I would recommend two additional runs in which *h* is estimated by the model, one using the distribution calculated by Brodziak and Mangel as a prior for *h* and another run in which a less informative prior would be used. To do this, the authors need to take advantage of the features of the Stock Synthesis that allow Bayesian statistics to be used for the calculations. This is a good example why this extra feature could benefit the stock assessment for striped marlin (I referred to that in the previous section). These calculations are also of relevance to calculations of MSY (see comment below).

Deviation in annual recruitment

The assumption used about the sum of deviates in annual recruitment over the estimated period was that it would be zero. To identify the years for which the recruitment could be estimated a model run with all recruitment deviations estimated was done and the CV of the recruitment estimates was plotted. The years for which information was available to estimate recruitment were those years for which the CV had been stabilised. However, it is not clear to what extent the stability observed in the CV was the effect of the assumption used about the sum of deviates over the estimated period. The authors need to check if that assumption influenced the pattern of deviates found (and the recruitment values calculated). In other words, would have that stability in the deviates been found if the model had been run without the assumption about the sum of deviates?

As shown in Figure 30 of the stock assessment report, there are a couple of values of recruitment that are much higher than the other values. Those values are between two and three times greater than the average recruitment values estimated in the same period. However, as Table 8 in the same report shows, those values are up to 10 times

higher than the recruitment values calculated for similar spawning stock biomass in other years. For example, recruitment in 1984 was equal to 1,620,000 fish for a spawning biomass of around 3,000 t. The recruitment for similar spawning stock biomass in 1993 was 116,000 fish; that is less than one tenth of the former value. These values for recruitment might be the values that produce the better fit for the model but do they make sense from a biological point of view? An explanation is needed as to whether such wide deviation in the recruitment values is biologically plausible and should be permitted in the model.

CV for recruitment values

The CVs for recruitment shown in Figure 11 of the stock assessment report are very different to the value of the CV calculated from the figures shown in Table 8. As described in the report, the values shown in Figure 11 were calculated assuming that the recruitment for all years could be estimated while those in Table 8 were found when 2009 and 2010 were excluded from the estimation process. Did the exclusion of two points created such noticeable difference in the CVs? Please, check/clarify.

CPUE series

The results from the standardisation of CPUEs show that the explanatory power of the adopted models is moderate, so some of the variability in the data cannot be explained. More work is needed to assess the influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing strategies, etc.) on CPUE values.

Model convergence

The report provides very little information about convergence of the model. The information provided shows that the model converges at almost the same value of the likelihood function if the starting values of the estimated parameters are changed by 10%. That is not enough information to assess convergence; additional information is needed such as: a) whether any of the estimated parameters converged at the minimum or maximum allowed value (hit the boundaries) b) diagnostics used to check MCMC algorithm convergence.

Presentation of information

A table to present the values of life history parameters used in the model and clarify whether they are fixed or estimated should be included in the report (similar to Table 3.2 which is included in the report from the BILLWG meeting in December 2011). A table with the CPUEs used and their CVs (before adjusted so their average value was above 0.2) is also needed since this information is not available in a single document and therefore, one needs to refer to a number of documents to get this information.

TOR 3. Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status

The stock assessment calculated reference points based on MSY that included yield, fishing mortality (calculated as the average of fishing mortality on ages 3 and older), population size, spawning biomass and the spawning potential ratio. Those values are then used in the presentation of the assessment results. Those management parameters are appropriate and are commonly used to describe current and future status of stocks. However, it is not clear why fishing mortality at MSY was calculated using the fishing mortality for fish of age 3 or older only and the same approach was used to present fishing mortality for each year in the calculations. Fish of all ages are exploited in the fisheries described in the model. If certain age classes are excluded then any increase or decrease in the fishing pressure on those ages classes would not be captured in the value of annual fishing mortality. The authors need to explain why they chose to use values of fishing mortality for fish of age 3 or older to calculate the average fishing mortality at MSY and also present annual fishing mortality. This will not capture changes in fishing pressure on young fish that might increase as the size of the population decreases.

The model uses the steepness, *h*, of the stock recruitment function to calculate some of those parameters. Therefore, estimating the most plausible value of steepness will directly affect the results of calculations of the management parameters. This provides one more reason why it is important to make use of the statistical capabilities of the assessment framework to estimate the value of *h* (as suggested in TOR 2). Therefore, the reference points would need to be recalculated if the additional runs for steepness suggested under TOR2 estimate a different value for steepness from that used in the base case run.

Comments provided in TOR4 about the use of results in 2012 as a reference point are also relevant here.

TOR 4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status

The Default Rebuilding Analysis software (Punt 2010) was used to do model projections. The population and fleet dynamics model used in the software includes many of the features that were also available in the Stock Synthesis and its use is valid. However, it does not offer the option to do seasonal calculations (i.e. it uses an annual time step) which was something used for the stock assessment. Although this

is not so unusual it does mean that adjustments had to be made to make the transition from the Stock Synthesis configuration to the Default Rebuilding Analysis. For example, the spawning stock biomass in the Stock Synthesis was calculated at the beginning of April (second season in the model) while for the projections, that needs to be calculated at the beginning of July due to the annual time step used for the projections. I do not expect that such adjustments have any serious effects on the results. It would be preferable though if the same model structure was used to do the stock assessment and projections to avoid introducing any errors as part of the transition; future work needs to address that.

I have also listed below some other issues with the way the projections were done that would influence the results and need to be addressed. Those issues are related to how the model was applied and input values were calculated rather than the model itself.

Selectivity

The authors used 3 fisheries for the projections which represented the 18 fisheries simulated in the stock assessment calculations. The 18 fisheries were grouped into the 3 groups based on similarity of the selectivity pattern. Then the total catches at age for each cluster of fisheries and the population size at age in the last three years of the stock assessment (2007-2009) were used to calculate selectivity for each cluster. There are two issues with this approach:

a) The clustering of fisheries was done based on the original assumptions about the shape of the selectivity curve (logistic, domed, etc.). However, the estimated selectivity curves for each fishery diverted from starting assumptions about their shape (see Figure 16). Therefore, the combined fisheries do not necessarily have those properties that the authors attribute to them. For example, the results in Figure 16 in the stock assessment report show that the 2nd cluster, which describes longline fisheries that do not catch age 0 fish, does catch age 0 fish although the authors suggest they will not. So, if the fisheries were to be reduced from 18 to a smaller number then the clustering should have been based on the estimated selectivities not the starting assumptions (unless there was other information to suggest that the predicted selectivity for some of the fisheries was not correct/realistic). That is important since it could introduce an error in the interpretation of the projection results and what they mean for the management of fisheries with different selectivities.

b) More importantly, the results for the combined selectivities do not seem to allow for catches of age 0 fish for the 2nd group of fleets (Table 9) which is not what Figure 16 suggests. The fact that the 2nd group of fleets catches age 0 fish is also confirmed by the results in Table 9 which show that the model has calculated a value for weight-at-age of age 0 fish caught by this group of

fleets. So, the values for selectivity shown in Table 9 need to be checked and explain if there is a typo or an error in the way the combined selectivity was calculated.

Assumption about exploitation in the first years of projection runs

The assumption made for the first two years of the projections (2010, 2011) was that the exploitation level was the same as the current (2009) level of exploitation. The alternative assumption was that catches were constant and equal to 80% of current catches. Those two assumptions are not consistent to each other. What is the logic behind the latter assumption? A single assumption for the first two years should have been used for all projections. As it is now, the population size in 2012 differs depending on whether the constant initial exploitation level or constant initial catches assumption was used. This means that the results of the projections for each of those assumptions are not comparable. It also means that the recovery of the stock under the constant catch assumption takes less time because it starts with a higher population size in 2012 than that found under the constant initial exploitation level assumption (i.e. 1640 mt of spawning biomass in comparison to 1333 mt). Furthermore, it is confusing because some of the results are presented relative to the values of the parameters in 2012 but the values in 2012 differ depending on the assumption about initial exploitation used.

Some of the projections need to be repeated to use only a single assumption about exploitation in years 2010 and 2011. Also, it would be more informative if MSY-based values (or virgin population values) are used as reference when the results are presented as relative values.

TOR 5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices

Work to better understand and describe the recruitment process as well as the level of variation in recruitment that is biologically plausible is recommended to inform population dynamics model formulation and future stock assessments

The stock assessment does not use any fishery independent abundance indices because none were available at the time when the analysis was taking place. If there is fishery independent information that has not become available then priority should be given in putting it in a format that would allow its use for stock assessment.

The results from the standardisation of CPUEs show that the explanatory power of the adopted models is moderate. The influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing

strategies, etc.) on CPUE values needs to be explored further. Similarly, the catchability of a fishery seems to change considerably among different areas. Further work is recommended to explore the factors that might influence or contribute to those changes.

As mentioned in previous sections, spatial heterogeneity is a feature of the dynamics of the assessed population. Work to incorporate information about movement and spatial heterogeneity in the distribution of the population into stock assessment discussions is recommended.

Conclusions and Recommendations

The software used for both the stock assessment and projection is appropriate for the stock considered and the parameterisation of the model is valid. Overall, the stock assessment is a good first attempt to capture the dynamics of this newly defined stock and make use of a wide range of information to feed into the calculations. Although management targets have not been formally adopted for this stock, the benchmarks used to express the status of the stock are appropriate.

As explained in previous sections, there are aspects of the stock assessment that need refinement and to either be explained better or extended to address some additional considerations. Similarly, the projections need to be revisited to rectify the issues identified above. I have described that work as well as other future work in my recommendations below.

TOR 1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data

Recommendation 1.1. Stock Synthesis allows for Bayesian estimation; it is recommended that this feature of the software is used to provide more flexibility in the parameters that can be estimated.

Recommendation 1.2. Work to incorporate information about movement and spatially heterogeneity in the distribution of the population into stock assessment discussions is recommended. That would provide a better understanding of the dynamics of the stock and help scientists decide whether a more detailed model is needed to simulate the dynamics of the stock.

TOR 2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models

are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for

Recommendation 2.1 The authors need to explain why the values of the CVs for length at age estimates used in the stock assessment differ from those calculated and presented in the Billfish working group workshop in December 2011 and what impact that change in the CV values had in the model results.

Recommendation 2.2. The authors need to confirm/check whether the value for *k* in the growth equation used in the stock assessment model is 0.24 or 0.34.

Recommendation 2.3 Given the sensitivity of the model to values of natural mortality, I recommend that another sensitivity run is done assuming that the minimum age for adults is lower than 4 years (e.g. 3 years) to test the effect of that assumption on model fit and results.

Recommendation 2.4 I would recommend two additional runs in which *h* is estimated by the model; one using the distribution calculated by Brodziak and Mangel as a prior for *h* and another run in which a less informative prior would be used. To do this, the authors need to take advantage of the features of the Stock Synthesis that allow Bayesian statistics to be used for the calculations.

Recommendation 2.5 The authors need to check the effect of the assumption used about the sum of deviates in annual recruitment on the pattern of deviates found (and the recruitment values calculated).

Recommendation 2.6. An explanation is needed as to whether the wide deviation in the recruitment values seen in the predicted recruitment is biologically plausible and should be permitted in the model.

Recommendation 2.7. The values of the CVs for the estimated recruitment differ considerably from values found in earlier calculations. The reason behind such noticeable difference should be explained.

Recommendation 2.8 Further work is needed to assess the influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing strategies, etc.) on CPUE values and improve the explanatory power of the models used to standardise CPUEs.

Recommendation 2.9 More information should be provided about model convergence including information about whether any of the estimated parameters

converged at the minimum or maximum allowed value (hit the boundaries) and diagnostics/plots used to check MCMC algorithm convergence.

Recommendation 2.10 A table to present the values of life history parameters used in the model and clarify whether they are fixed or estimated should be included in the report. A table with the CPUEs used and their original CVs is also needed.

TOR 3 Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status

Recommendation 3.1 The authors need to explain why they chose to use values of fishing mortality for fish of age 3 or older to calculate the average fishing mortality at MSY and also present annual fishing mortality. This will not capture changes in fishing pressure on young fish that might increase as the size of the population decreases.

Recommendation 3.2 Some of the reference points would need to be recalculated if the additional runs for steepness suggested under TOR2 estimate a different value for steepness from that used in the base case run.

TOR 4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status

Recommendation 4.1 I would recommend that the same software/model is used to do the stock assessment and projections.

Recommendation 4.2 I recommend that the grouping of selectivities is done based on estimated selectivities not the starting assumptions.

Recommendation 4.3 The values for selectivity shown in Table 9 of the stock assessment report need to be checked and explain why the selectivity chosen does not reflect the catch-at-age behaviour shown in Figure 16 of the report.

Recommendation 4.4 Some of the projections need to be repeated to use only a single assumption about exploitation in years 2010 and 2011.

TOR 5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices

Recommendation 5.1 I recommend further work to better understand and describe the recruitment process as well as the level of variation in recruitment that is biologically plausible and inform population dynamics model formulation and future stock assessments.

Recommendation 5.2 If there is fishery independent information that has not become available for the 2012 stock assessment then priority should be given in putting it in a format that would allow its use for future stock assessments.

Recommendation 5.3 The influence of factors such as area-dependence and changes in the fleet behaviour (target species, use of a combination of gears/fishing strategies, etc) on CPUE values needs to be explored further. Also, further work is recommended to explore the factors that might influence or contribute to the area-dependent changes in catchability found in the stock assessment.

Recommendation 5.4 Work to incorporate information about movement and spatially heterogeneity in the distribution of the population into stock assessment discussions is recommended.

Appendix 1: Bibliography of materials provided for review

Report reviewed

Stock assessment of striped marlin in the Western and Central North Pacific Ocean (2012) *authored by* Hui-Hua Lee, Kevin R. Piner, Robert Humphreys and Jon Brodziak. Billfish Working Group, International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean.

Additional material

Billfish Working Group (BILLWG). 2011a. Report of the Billfish Working Group Meeting, 19-27 January, 2011, Honolulu, HI, USA. Annex 7. Report of the Eleventh Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 20-25 July, 2011, San Francisco, CA, USA.

Billfish Working Group (BILLWG). 2011b. Report of the Billfish Working Group Meeting, 24 May-1 June, 2011, Taipei, Taiwan. Annex 8. Report of the Eleventh Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 20-25 July, 2011, San Francisco, CA, USA.

Billfish Working Group (BILLWG). 2012a. Report of the Billfish Working Group Meeting, 6-16 December, 2011, Honolulu, HI, USA. Annex 5. Report of the Twelve Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 18-23 July, 2012, Sapporo, Japan.

Billfish Working Group (BILLWG). 2012b. Report of the Billfish Working Group Meeting, 2-9 April, 2012, Shanghai, China. Annex 7. Report of the Twelve Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 18-23 July, 2012, Sapporo, Japan.

Brodziak, J. and Mangel, M. 2011. Probable values of stock-recruitment steepness for north Pacific striped marlin. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/11: 13p

Kanaiwa, M., Kimoto, A., and Yokawa, K. 2011. Standardized CPUE of striped marlin in north western central Pacific Ocean by using GLM. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/06: 16p.

Piner, K.R. and Lee, H.-H. 2011a. Meta-analysis of striped marlin natural mortality. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/10: 09p.

Piner, K.R. and Lee, H.-H. 2011b. Correction to Meta-analysis of striped marlin natural mortality. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/08: 01p.

Piner, K.R., Lee, H.-H., Taylor, I.G., Katahira, L., Tagami, D., and DiNardo, G. 2011. Preliminary Striped marlin stock assessment. Working paper submitted to the ISC Billfish Working Group Meeting, 6-16 December 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-3/01: 34p.

Punt, A.E. 2010. SSC Default Rebuilding Analysis: Technical specifications and User Manual. Jan. 2010.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011a. Reproductive biology of female striped marlin (*Kajikia audax*) in the waters off Taiwan (preliminary). Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/11: 14p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011c. Age and growth of striped marlin (*Kajikia audax*) in waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/09: 15p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011d. Age and growth of striped marlin (*Kajikia audax*) in waters off Taiwan: A revision. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/07: 12p.

Walsh, W. and Lee, H.-H. 2011. Standardization of striped marlin, *Kajikia audax*, CPUE with generalized linear models fitted to pelagic longline observer fata from the Hawaii-based fishery: 1995-2009. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/08: 45p.

Yokawa, K. and Kimoto, A. 2011. Standardized CPUE of striped marlin caught by Japanese coastal large-mesh drift fishery in the northwest Pacific in the periods between 2001 and 2009. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/05: 5p.

Appendix 2. Statement of Work for Dr Yiota Apostolaki

External Independent Peer Review by the Center for Independent Experts

Stock Assessment of Striped Marlin

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide an impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Striped marlin (*Tetrapturus audax*) is one of six species of billfishes commonly harvested multi-nationally from commercial and recreational fisheries in the western and central Pacific Ocean regions. Fishery management requires high quality science to effectively manage and conserve our living marine resources, and the scientific peer review of stock assessments by external CIE expertise is an important process in the determination of best scientific information available. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise, working knowledge and recent experience in the application of fish stock assessment, mathematical modeling, and statistical computing. Scientists who are employed by or have significant interactions with the Inter-American Tropical Tuna Commission (IATTC) and the Western and Central Pacific Fisheries Commission (WCPFC), and the Secretariat of the Pacific Community (SPC), should not be considered as reviewers. Scientists associated with the ISC also should be excluded as reviewers. Each CIE reviewer's duties shall not

exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, and other pertinent information. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than **18 November 2012**, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, and to Dr. David Die, CIE Regional Coordinator, via email to ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

18 October 2012	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
25 October2012	NMFS Project Contact sends the CIE Reviewers the report and background documents
1-16 November 2012	Each reviewer conducts an independent peer review as a desk review
18 November 2012	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
1 December 2012	CIE submits the CIE independent peer review reports to the COR
7 December 2012	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long

as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Gerald DiNardo, Stock Assessment Program Leader (NMFS Project Contact)
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Gerard.DiNardo@noaa.gov Phone: 808-983-5397

Kevin Piner
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Kevin.Piner@noaa.gov Phone: 858-546-7003

Jon Brodziak
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Jon.Brodziak@noaa.gov Phone: 808-983-2964

Hui-Hua Lee
University of Hawaii, Joint Institute for Marine and Atmospheric Research
2570 Dole St., Honolulu, HI 96822
Huihua.Lee@noaa.gov Phone: 808-983-5352

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2: Terms of Reference for the Peer Review

Stock Assessment of Striped Marlin

1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.
2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.
3. Comment on the proposed population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT*); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.
4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.
5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.

Center for Independent Experts (CIE) Independent Peer Review of the 2012 Stock Assessment of Striped Marlin in the Western and Central North Pacific Ocean[1]

Tom Carruthers

t.carruthers@fisheries.ubc.ca

335 AERL 2202 Main Mall

University of British Columbia

Vancouver

Canada

December 2012

[1] Lee, H-H, Piner, K.R., Humphreys, R., Brodziak, J. 2012. Stock assessment of striped marlin in the western and central North Pacific Ocean. Billfish working group of the International Committee for tuna and tuna-like species in the North Pacific Ocean.

1. Executive Summary

The assessment document provides a clear description of a methodology with a number of central strengths. The modelling was undertaken using established software (SS3) that has been subject to extensive review in stock assessments elsewhere. Where available, the assessment makes use of new information on the biology and ecology of the striped marlin stock. The report provides an excellent introduction to the fishery and presents a complete account of the different stock structure and mixing hypotheses. The assessment modelling itself is very detailed, particularly in the attempt to capture the historical fishing mortality at size of multiple fleets. In general, the authors do an excellent job of highlighting and discussing many of the implicit assumptions of the model. Such an assessment can produce a large quantity of output. However, the results of the assessment are clearly presented in a range of informative figures and tables including several that allow the reader to easily make cross-stock comparisons. A comprehensive range of sensitivity analyses and projections is another key strength of the assessment.

There are however important areas in which greater detail is required in order to have confidence over the assessment results. In particular, much greater transparency is needed in the description of CPUE standardization approaches used to construct relative abundance indices in order to understand whether these were calculated correctly. The account of the size-frequency data is also too cursory for a reader to understand whether these data are sufficiently informative to support a large and intricate component of the assessment model.

There is not a clear justification for assessment model structure such as seasonality and the inclusion of regional fisheries. In some cases the base-case assessment model appears to include complexity that is demonstrated to be unnecessary by sensitivity analysis. In general the model seems very complex in its approximation of fishery dynamics relative to the simplicity of the spatial population assumptions. It is problematic therefore that the document does not provide sufficient assurance that the model is not overparameterized; that it converges reliably and can robustly estimate management reference points. Let us assume that the purpose of stock assessment is the provision of reliable management advice. It is difficult to know whether the assessment model adopted here can be expected to perform better than more simple approaches.

Various parts of the methodology contain disparities in assumptions. For example, some important standardization models assume a viscous stock with regional abundance trends and seasonally variable biomass that are not fully accounted for by the assessment model.

There are also some important overarching problems. The lack of clear management objectives for the striped marlin stock prevents assessment results from being presented in a meaningful framework and may hinder the development of quantitative tools to support decision making.

2. Background to the review

Striped marlin (*Tetrapturus audax*) is one of six species of billfishes commonly harvested multi-nationally from commercial and recreational fisheries in the western and central Pacific Ocean regions. Fishery management requires high quality science to effectively manage and conserve our living marine resources, and the scientific peer-review of stock assessments by external Center for Independent Experts (CIE) expertise is an important process in the determination of best scientific information available (from the Statement of Work Appendix 2).

Important fisheries for striped marlin have operated in Western and Central North Pacific Ocean (WCNPO) since the early 1950s. Catches in the period before 1975 are considered to be in the range of 4000-8000mt (the 'equilibrium catch' prior to 1975 is assumed by this assessment model to be 5000mt). Since 1975, catches in the longline fishery fluctuated but the trend remained broadly flat until the early 1990s. The substantial contribution of the driftnet fishery after 1970 (approximately 1/3 of catch) lead to peak catches around 10,000mt in the mid 1980's. After 1992, catches in all fleets can be observed to decline to current (2010) total levels of around 2600mt. While simplistic, the consistent decline in both catches and catch rates of key fleets (*e.g*. Japanese longline fleet, Kanaiwa *et al*. 2011) provides some reason for concern about the status of the WCNPO striped marlin stock.

Striped marlin is a highly migratory species occupying a large spatial range over which large discrepancies in population density and size structure are likely to occur. While spatial structure is a central challenge in the approximation of population dynamics, size-dependent exploitation by multiple fleets poses difficulties when attempting to represent fishing dynamics. An additional challenge for stock assessment is the paucity of reliable data prior to 1975, a period over which a relatively large degree of stock depletion may have occurred.

The previous assessment in 2007 (MAR&SWO 2007; Piner *et al*. 2006, 2007) was the first detailed attempt to characterize fishing /population dynamics and estimate current stock status. The central conclusion of the 2007 assessment was that spawning potential ratio (a measure of 'health' of a fish stock) was less than 10% that of unfished levels. It proved difficult to interpret these results with confidence due to the generally sparse biological and ecological information available to support the assessment. Central uncertainties included "stock structure, spawner-recruit resilience (h), natural mortality (M) and the growth rate of the species" (from the document under review here, Lee *et al*. 2012, hereafter referred to as 'the document' or 'the report'). Subsequently a number of studies have provided more credible estimates of these inputs (*e.g*. Sun *et al.* 2011a,b,c; Piner and Lee 2011a,b) potentially improving the basis for a revised stock assessment.

The current assessment considerably updates and expands on the 2007 analysis taking into account new information on life-history, recent total catch, catch composition and relative abundance indices. The current assessment prescribes a higher level of stock productivity with higher base-case assumptions regarding natural mortality rate and steepness (age specific natural mortality ranging from 0.54-0.38 as opposed to 0.3 across all ages in the 2007 assessment; steepness of 0.85 as opposed to 0.7 in 2007).

3. Description of the Individuals Reviewers Role in the Review Activities

A detailed description of the reviewer's role can be found in the Statement of Work (Appendix 2). The supporting documents, many of which are listed in the bibliography (Appendix 1) were received on the 26th October 2012. The main assessment report was received on the 10th of November 2012. Most of the remaining documents listed in the Bibliography can be found on the ISC website.

4. Summary of findings in regard to TORs (weaknesses and strengths)

4.1. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

A specific discussion of these issues is included in the detailed break-down of Section 4.2 below. It is hard to judge whether the data available are of sufficient quality to support an assessment of this complexity. Given the lack of information regarding model convergence it is difficult to know whether the assessment methods are reliable, properly applied and appropriate for the species (Section 4.2.1), fisheries and the available data (4.2.2).

4.2. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.

4.2.1 Model configuration and assumptions

Description of the model

The authors provide a comprehensive description of the model in the main text of the document but the transition equations (page 31) do not include the seasonal structure. This raises a number of questions about model assumptions. These are probably ignorable but it would be instructive to know when recruitment was assumed to have occurred during a modelled year.

A seasonal population dynamics model

From the assessment document it is hard to completely understand the rationale for a seasonally disaggregated assessment. It is argued that a seasonal model is necessary to (1) reliably estimate selectivities and (2) account for difference in the magnitude of catch between seasons (page 21, paragraphs 3 and 4). That size composition of catches differs between seasons is not necessarily a problem for an annual assessment. The definition of selectivity simply changes to the fraction of modal fishing mortality rate exhibited on an age class over the course of a year. That the magnitude of catch differs between seasons is also not necessarily a problem for an annual assessment; perhaps fishing effort is greater in certain seasons. The justification for a seasonal model is usually based on the seasonal interaction between the stock and the fleet among years. If the seasonality in the stock is constant over time (roughly the same seasonal pattern in abundance and size composition among years) and the distribution of fishing is also constant seasonally over time (roughly the same fraction of effort in each season among years) then an annual assessment may not be strongly biased. If however there are large

fluctuations in the seasonal pattern of fishing (between years a very different sub population structure is being fished) then it may be important to model seasonal population dynamics. However this rationale is not referred to by the authors and there is insufficient information in the assessment document to understand whether the seasonality in fishing and/or the population are changing dramatically over time. It would be interesting to see the difference in estimated reference points from an annual model with the other assumptions held constant.

Disparity in the magnitude of spatial assumptions and mortality-at-size assumptions

I agree with the authors' decision to fix a number of inputs that are generally not well estimated such as steepness, mortality at age and maturity-at-age. It is interesting that projections (and therefore the basis for management advice) depend most heavily on a very small number of estimated parameters, in particular stock size (*R0*) and recent recruitment deviations (that may not be that well estimated). The remaining model complexity (napkin arithmetic points to over 120 parameters) is spent approximating size-specific fishing dynamics. It is possible that this demands too much of the size composition data and it seems disproportionately concerned with approximating size-specific mortality, particularly considering that the model relies on relatively large spatial assumptions.

A general comment on model complexity

From my perspective, the tendency for *ad-hoc* adjustment of model structure to remove fine-scale residual patterns may be problematic and inconsistent with the broader objectives of stock assessment: "Seventeen fisheries were initially defined but further analysis indicated that a residual pattern and quarterly size observations from the Japan other fishery showed a substantial seasonal pattern of larger fish caught in the first two seasons (see Section 3.5 below on length frequency data and Figure 5). Seasonality in selectivity was modeled by splitting the Japan "other fishery" into two seasonal fisheries corresponding to seasons 1-2 and 3-4 of the calendar year in order to reduce the influence of the misfit" (Page 20, last paragraph).

It may be possible to tweak assessment models to remove patterns in residual errors. But the objective of stock assessment is not usually to provide a complete account of all of the historical complexities of the system. Generally we wish to robustly capture stock size and productivity in order to make reliable decisions about how to manage the stock. Adding parameters to chase down areas of model misfit will certainly increase the challenge for the numerical optimization, may lead to over-fitting and spurious predictions without necessarily improving the reliability of decision making. While I acknowledge the comprehensive attempt made by the stock assessment scientists to account for important changes in fishery dynamics in terms of residual errors, I am concerned that there may be little benefit in terms of the provision of reliable management advice. Other areas where complexity has been added to remove residual patterns is the blocking of historical selectivities. My experience reviewing similar models applied to stocks elsewhere (with much fewer fleets) has been that the model may be highly overparameterized. There is an onus in such instances on providing conclusive evidence of convergence that in my view is not reflected in the stock assessment document.

In Section 5.1 of the results (page 35) it is stated that: "There is no evidence of substantial differences in the scaling parameter (R0) and total likelihood showing a better fit (Figure 12). Based on these results, the BILLWG concluded that the base-case assessment model is relatively stable with no evidence of lack of convergence to the global minimum". In my view this does not constitute a sufficiently detailed assurance

of convergence. In the most overparameterized models I have reviewed previously, the scale of the population represented by *R0* remains consistently estimated under jittered runs. The fact that the objective function does not change may also mask a serious problem (*e.g*. a very flat and poorly-defined objective surface). It would be much more informative to see the variance in the estimates of quantities of management reference points (*e.g*. *MSY*, $B_{current}/B_{MSY}$ and $F_{current}/F_{MSY}$) from different starting values. Other important diagnostics of model overparameterization that could be included here are parameters estimated at their bounds, strong posterior cross-correlation among parameters, poor MCMC mixing (high auto-correlation requiring heavy thinning to satisfy convergence diagnostics) unrealistically high precision in model predictions, high sensitivity in model predictions to credible changes in input parameters and an inability to recover known parameter values from simulated data (a built-in feature of SS3). To have confidence over the model predictions it would be highly informative to have a more complete account of whether the model suffers from such phenomena.

In several instances, reducing the complexity of the model or the way in which data are included, does not significantly alter results (*e.g*. insensitivity to spatially disaggregated index for the Japanese longline fleet; section 5.6.1.1, page 39). It is not clear why the base-case model should be the more complicated of the two configurations that produce similar results.

The objective function

The authors make a sensible decision to reweight the effective sample size of the length composition data to reflect non-independence in observations. This helps to combat a well-established problem in which the analysis becomes dominated by the composition data.

The authors state that there are three components to the likelihood function: the total catch data, CPUE indices and the length-frequency data. I suspect that this is not strictly correct and that there are a number of other likelihood components that go unmentioned. For example, penalties for parameters as they reach their bounds or as predicted stock size approaches zero. The difficulty for most assessments undertaken in SS3 to completely describe the objective function remains a central criticism I have of the software.

4.2.2 Input data

Length-frequency data

Due to the relatively complex fishery dynamics (multiple fleets, multiple seasons, blocked selectivities *etc*.) the model may be strongly dependent on the quality of the length-frequency data. Section 3.5 (page 22) provides a brief description but it is not easy to gain an intuition about the amount of data gathered and the process used to derive the lengths. Additional information would be useful here to allow a reader to better understand to what extent these data can be expected to inform the more intricate aspects of the assessment model.

The division of relative abundance indices into temporal blocks

It is argued that the Japanese longline indices should be divided into three periods (1975-1986, 1987-1999 and 2000-2009) to "account for changes in operation, hook-per-basket (HPB) distribution, targeted fish and length distribution of catch". It is usually the role of standardization to remove these confounding influences. Indeed Kanaiwa *et al*. (2011) account for HPB in their standardization. By severing the abundance time series into three parts, long-term information regarding depletion is largely removed and

three nuisance parameters (*q*'s by time period) must be estimated instead of one. The likely result is that depletion is now informed to a much greater extent by other data that may not be as reliable. For example, the inference regarding total mortality rate *Z,* from in the catch composition data.

The inclusion of relative abundance indices for multiple fleets

There is only one trend in real abundance and several standardized indices that each provide a different inference of stock trend. Let us assume that the objective of the standardizations is to produce an index of population-wide abundance. Assuming that the stock is fully mixed and vulnerability schedules are comparable among fleets there are two possible conclusions: that (1) all but one of the standardization methods are not operating correctly or (2) all of the standardization methods are not operating correctly. In such cases it is not defensible to fit the model to multiple sets of derived data of which the majority are known to be incorrect. The assessment report includes a note to this effect (page 28, last paragraph) but then continues with a multiple index approach.

On a practical level, the objective surface on which the optimization algorithm operates usually becomes less well-defined with multiple local minima at different parameter vectors that suit particular abundance indices.

On a theoretical level, the inclusion of multiple abundance indices may increase the strength of the spatial assumptions of the assessment. Clearly a spatially aggregated model is to some extent inconsistent with known spatial characteristics such as size structuring, regional abundance trends and spatial heterogeneity in abundance. In particular spatial heterogeneity in abundance is evident from spatial plots of CPUE among fleets. The implicit assumption of the spatially aggregated model is that while population density may vary in space, the distribution of the fleet is constant in relation to this regional abundance (effort is distributed consistently on the population). Fleets with the most complete spatial coverage sample a greater range of the stock and indices are weighted in proportion to regional abundance (in a GLM with marginal time and area effects for example). Including an index for a regional fishery increases the spatial assumptions of the assessment: we now have to assume that regional trends reflect overall population trends. Additionally the inclusion of a regional index may provide extra weight to a particular area of relatively low abundance and may bias the assessment by the location of observations. For this reason those indices that are applicable to small sub-areas such as the Japanese coastal large-mesh drift fishery (Yokawa and Kimoto 2011) should not be used to infer-population wide abundance trends in a base-case assessment. This is particularly the case if the density of the population in these areas varies among years due to temporal changes in rates of migration (as implied by the year x season interaction effects modelled by Yokawa and Kimoto 2011).

Similarly, the theoretical basis for including three regional abundances indices for the Japanese longline offshore fishery (JPN_DWLL1, JPN_DWLL2, JPN_DWLL3) is not clear. It is theoretically inconsistent to incorporate such data in an assessment model that assumes that there are no regional abundance trends. However, it may be argued that while regional trends in abundance exist we may approximate stock dynamics by a single area model. Since these regional indices are essentially time x area interactions it would be better to undertake a single CPUE standardization including these interactions. This would derive a single index weighted by the predicted level of abundance in each region. Currently, the assessment assigns equal prior weight to these indices when in fact one may be tracking the abundance of a substantially smaller fraction of the population.

It is indicated in the assessment report (page 28, last paragraph) that the reason for including multiple series was due to a lack of consensus in expert judgement regarding the representativeness of the different indices. An attempt was made to evaluate the different indices objectively based on correlation analysis and sensitivity analysis. The rationale for why these approaches should be helpful was not provided. It might be better to undertake a basic simulation exercise. In their simulation analysis of CPUE for tuna and billfish fisheries in the Atlantic (Carruthers *et al*. 2010), there were clear grounds to favour some fleets over others due to their relatively complete spatial coverage and recording of catch rate covariates over time. For example, if it can be assumed that the population is spatially well mixed, CPUE standardization methods with marginal area effects were best informed by the Japanese longline fleet that has the most complete spatial coverage over time and has relatively complete records of hooks-per-basket.

Japanese longline CPUE 1975-2010, Kanaiwa *et al*. 2011.

Kanaiwa *et al*. (2011) provide a cursory account of their standardization method and provide insufficient information to reproduce the method. Fundamental equations are missing. For example, it is not clear how annual relative abundance and standard error were calculated. Future standardization approaches should include the equation $I_t = \ldots$ where I is the derived index and t the time subscript. Kanaiwa *et al*. (2011) provide no information about how they dealt with potentially critical problems regarding the interpretation of catch rate data. Did they account for differences in the size of areas? How were the confounded year x area and quarter x area interactions dealt with in the calculation of annual index? Did the authors account for different sample sizes among strata (*e.g*. Campbell 2006)? It is not clear why the model should use numbers/effort as opposed to weight/effort as a unit of CPUE since (as far as I can tell) the indices are compared with biomass in the assessment model.

It is not clear why the standardization was undertaken on a spatially aggregated dataset as opposed to the more detailed trip-level data that contains 1x1 degree detail and hooks-per-basket covariate information. I suspect this was due to the computational constraints of modelling time x area interaction effects. If aggregating data to 5x5 degree cells was necessary to model interactions it would be desirable to understand whether this additional model complexity was necessary. Unfortunately Kanaiwa *et al*. (2011) provide no plots of the time x area interactions (or in fact any plots of indices) to examine whether there are differences among the predicted regional abundance trends. It should be noted that modelling CPUE with time x area interactions is based on spatial assumptions that are inconsistent with that of the assessment model; namely that there the stock is not well mixed and there are regional abundance trends in contrast to the assessment model that assumes a perfectly mixed stock.

The absence of basic figures is inexplicable. In a paper generating relative abundance indices it is strange not to include a graph of the derived abundance trends over time. Such plots should include a representation of uncertainty and the sensitivity of the derived index to credible alternative assumptions.

It is my experience that accounting for the marked changes in set depth of fleets such as Japanese longliners has a profound impact on the standardization of billfish CPUE. As noted in this document (Page 18, paragraph 1) longlining in the 1970s shifted towards deeper sets to target bigeye, a phenomenon also observed in longline fishing in the Indian and Atlantic Oceans (Miyake 2004). The expected effect is a marked increase in the inferred relative abundance for species inhabiting shallower waters. It is therefore surprising that the standardization of Kanaiwa *et al*. (2011) that accounts for both spatial effects

and depth effects leads to an inferred relative abundance that is so similar to the unprocessed nominal CPUE (see Figure 1 below). Without providing CPUE data and a reproducible method it is not possible to investigate this atypical result and I remain concerned about this important input to the stock assessment.

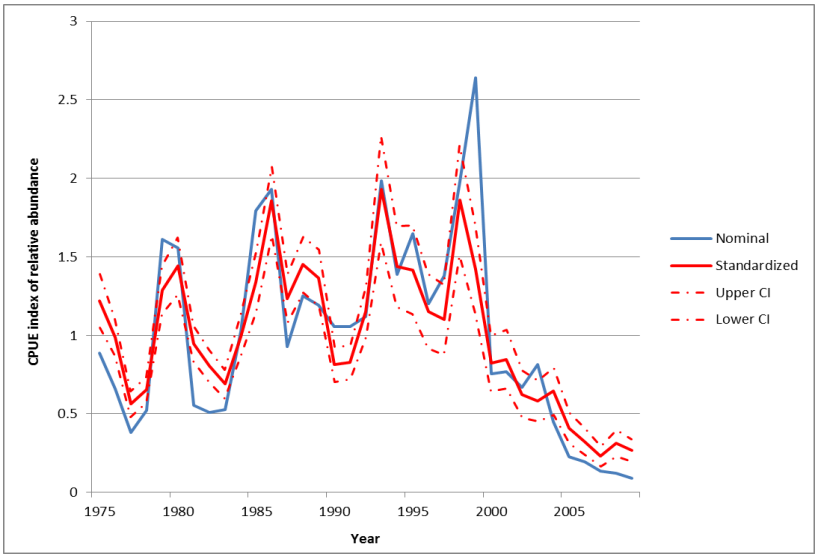

Figure 1. Base-case standardization of CPUE versus nominal CPUE for striped marlin in the WCNPO presented by Kanaiwa *et al*. (2011). The dotted lines represent the 95% confidence interval assuming that the standard deviation reported in Kanaiwa *et al*. (2011) refers to the mean and can be interpreted as a standard error.

The reliance on AIC to select the structure of the GLM standardization model is problematic. In general AIC is known to lead to the selection of GLM models that are overparameterized (Kadane and Lazar 2004). This tendency is magnified in the specific case of CPUE standardization because the objective is the extraction of a reliable relative abundance index not the prediction of the next CPUE observation (Carruthers *et al*. 2010); AIC is suited to the latter. Instead of model selection criteria it is desirable to use current ecological and fishery knowledge to select a defensible model (or small set of competing models) *a priori* rather than use AIC to undertake an ad-hoc search for GLM model that selects the best predictor of CPUE observations. For example: (1) we may have reason to believe that striped marlin inhabit different depths in different areas due to spatial changes in the depth of the mixed layer (supporting the inclusion of depth x area interactions); (2) tagging studies may indicate that the stock is viscous and spatial differences in the intensity of fishing would imply that regional abundance trends could be different (supporting the inclusion of time x area interactions).

Japanese coastal large-mesh drift fishery CPUE, Yokawa and Kimoto 2011

Similarly to Kanaiwa *et al*. (2011) above, Yokawa and Kimoto (2011) do not present basic calculations that are necessary to evaluate their standardization approach. There is no clear explanation of how they calculated their index. In particular it would be useful to know how they derived an annual index given a year effect that is confounded with the year x quarter interaction effects. This is not an ignorable matter and entirely determines the credibility of the index. In many standardization papers that I have previously reviewed, the marginal year effects are interpreted directly without consideration of the confounded interactions leading to spurious abundance trends.

It is not clear why the categorical effects were estimated in log-space. This is unusual and it would be useful to see the SAS code for this modelling.

The authors provide no explanation for the modelling of a year x quarter interaction. It implies that the stock is seasonally transitory in this fishing area over years. It follows that Yokawa and Kimoto (2011) acknowledge that the stock moves in and out of this area seasonally and that a low/high annual abundance could be driven by migration. This is problematic because it requires a level of spatial detail absent in the spatially aggregated assessment model. In my view these assumptions are sufficiently disparate to preclude the use of this index in the assessment.

Japanese high sea large-mesh driftnet CPUE, BILLWG 2011b

The 2011 Report of the Billfish Working Group Workshop (BILLWG 2011b) is provided in the stock assessment as a reference for the Japanese high sea large-mesh driftnet fishery. This is an update of the standardization of Yokawa (2005) that derived an index from 1977-1993. Yokawa (2005) provides more information regarding the underlying data. For example it is noted that a moratorium of large scale drift net fishing in the open ocean in 1993 led to all but 4% of catches being taken in a single near shore area. Unfortunately no information is provided about how this issue was dealt with in either the analysis of Yokawa (2005) or the update (BILLWG 2011b). Similarly to the standardization papers above it is not clear exactly how the index was calculated or the sensitivity of the index to other credible assumptions about stock and fishery dynamics. Again, it is not clear why the categorical effects were estimated in log space.

Taiwanese distant-water longline fishery 1967-2009, Sun *et al.* 2011d

The lack of transparency in the description of standardization methods described above is also applicable here. Similarly, no sensitivity analyses are conducted. There are additional curiosities however. The equation describing the GLM standardization model does not include a transformation of the nominal catch rate data which implies that the authors were applying a normal error model to a variable that cannot be negative (later they refer to a log-normal error distribution in contradiction to their GLM equation). The GLM model applied includes latitudinal and longitudinal marginal effects and a latitude x longitude interaction effect. They do not state whether these are continuous variables or discrete categorical spatial blocks. If they are categorical there is no reason to model the marginal effects as they are already accounting for areas defined in two-dimensional geographic space. We do not know how the authors accounted for potential sources of bias such as uneven spatial distribution of observations and differences in sizes of modelled areas.

The authors do not mention whether covariate set depth data are also available for these fisheries. Given the spatial range of this fishery, if depth and other covariate information regarding gear (*e.g.* bait) were available, this CPUE series could offer an interesting and credible alternative to the Japanese longline index as a primary source of information for stock-wide abundance trends.

Hawaiian Pelagic longline observer CPUE 1995-2009, Walsh and Lee 2011.

Of the indices used in the assessment Walsh and Lee (2011) provide the most comprehensive account of their standardization approach. Given that they did not model year interaction effects it is easier to deduce how they may have calculated their index (however the reader still does not know how they dealt with

unbalanced sampling and areas of different size). The analysis provides a number of valuable findings in particular the relative impact of sea surface temperature and set depth.

In regional index standardization of this type it would be useful to have illustrations of the spatial extent of the observations in order to assess how general indices are likely to be in regard to population-wide abundance. Additionally several of their residual plots show distinct structuring (clumping of plotted points) that suggest population or fishery characteristics not accounted for by the GLM modelling (*e.g.* the plot of sea surface temperature observed vs. standardized residuals).

Walsh and Lee (2011) conclude that "…striped marlin catches and catch rates have decreased considerably in the last 15 years. The relatively similar estimates of change in catch per set and CPUE and the finding that the standardized trends were highly correlated with differences that could be ascribed to operation changes in this fishery reinforces this conclusion". This is a strange statement in the context of CPUE standardization. If the indices are correlated with operational trends then the apparent changes may be an aberration of observation processes (that we wish to standardize for) rather than population changes. For example, consider a fishery in which a species is increasingly targeted at a particular depth. This operational shift may lead to apparent declines in species inhabiting other depth ranges. In the case of either species the correlation does not provide corroboration of the apparent trend in CPUE.

4.2.3 Input parameters

The fundamental biological and ecological relationships assumed by the model appear to be sound and the supporting documentation is generally detailed and carefully presented.

Growth, maturity and the stock-recruitment relationship

The assessment is well supported by the growth studies of Sun *et al.* (2011b; 2011c), the spawning analysis of Sun *et al*. (2011a) and the stock recruitment work of Brodziak *et al*. (2011). The estimate of steepness 0.87 may be a good base-case assumption but the '+/-' 0.05 appears to be quite a precise range for a quantity that is generally poorly informed. This is recognised in the sensitivity analysis and the range of values considered is much wider.

Natural mortality rate

Piner and Lee (2011a;b) provide a comprehensive account of the derivation of their natural mortality rate-at-age estimates. Their method is interesting and defensible relative to other papers on the subject and they acknowledge potential shortcomings of the approach. As they note, the same approach has been adopted elsewhere and has the core benefit that it provides a measure of uncertainty with which to bracket sensitivity analyses.

Sensitivity analyses

In general, a very comprehensive range of sensitivity analyses are included in the assessment document and these help the reader to understand more about the behaviour of the model.

In some cases changes in input parameters are used as the basis for sensitivity analysis. For example natural mortality rate: "Results indicate that models for both natural mortality rates fit worse by a moderate amount for length compositions (9 and 4 likelihood units worse than base case for high M and

low M, respectively). However, fit to CPUE series appeared similar based on the small changes in likelihood (< 2 likelihood units). In summary, total likelihood favors the base case model". Various structural aspects of the base-case model have been changed *ad-hoc* in order to reduce residual error given base-case inputs (such as natural mortality): "The authors note that many additional sensitivity runs were conducted in the development of the base case (*e.g*. bin definitions, initial conditions, alternative data sets *etc*.) that are beyond the scope of this paper to describe" (page 30, paragraph 3). It is therefore not surprising that alternative inputs lead to poorer fit and it may not be possible to interpret this as evidence with which to favour one natural mortality rate assumption over another.

4.2.4 Accounting for uncertainty

In general the assessment does a good job of expressing uncertainty by providing confidence intervals for predicted quantities and undertaking a range of sensitivity analyses. As noted in the report, structural uncertainty is not well accounted for and it would have been instructive to see the difference in predictions of key reference points generated from a more simple assessment model (*e.g.* with/without seasonal structure).

The assessment document refers to Bayesian methods of characterising uncertainty: "The structure of the model allows for Bayesian estimation processes and full integration across parameter space using the Monte Carlo Markov Chain (MCMC) algorithm" (page 23, paragraph 1). However this is not mentioned elsewhere in the document with the exception of the 'MCeval' lines of the SS3 starter file (page 104). There is no formal reference to chain thinning, convergence diagnostics or other matters that are necessary when interpreting MCMC outputs. A number of other comments imply that the MLE estimates and a hessian approximation to the standard error were used to construct the estimates of uncertainty expressed in tables and figures. Additionally the use of the term 'confidence interval' (a frequentist concept) implies that the MCMC run was not used to provide estimates of uncertainty.

It would be desirable to have the Bayesian posterior estimates since they are much more straightforward to understand and interpret. It may also be the case that the MLE estimate (the posterior mode) is not a suitable estimate of an expected value if the posterior is strongly skewed (in which case a posterior median is preferable). Several Bayesian outputs would also be useful in diagnosing model overparameterization, in particular convergence diagnostics and the joint posterior parameter cross-correlation plots. However it is likely that given the complexity of the model the MCMC evaluation would take several days (perhaps weeks) to run in order to satisfy convergence diagnostics and perhaps this is reason why these results were not included.

4.3. Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.

The stock assessment uses standard metrics such as spawners per recruit and spawning stock biomass and fishing mortality rate relative to *MSY* levels. These metrics are widely applied elsewhere and are relatively easy to interpret. Standard outputs such as Kobe plots (*e.g*. Figures D and E, page 7) are included here that provide a transparent account of predicted historical stock status and exploitation level.

It appears that there are no agreed management reference points for this stock: "No target or limit reference points have been established for the WCNPO striped marlin stock under the auspices of the

WCPFC" (page 45). However the authors make sure to use standard definitions of over/underfishing and over/underfished to describe the outputs of the assessment. In future assessments it would be highly desirable to have not only target and limit reference points but also acceptable probabilities for exceeding these levels. These definitions would allow assessments to phrase results in a clear way in direct reference to management objectives (for example probability of an MCMC projection not exceeding the limit reference point in a projected year). Other related benefits include the foundation for undertaking value-of-information analysis and management strategy evaluation.

4.4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.

The assessment report includes a detailed series of projections including both input and output control scenarios. The decision to limit projections to a relatively short time period of 5 years is defensible given the high level of recruitment variability and lack of apparent autocorrelation in recruitment. The suite of SS3 tools for conducting projections has been subject to review and revision in many other stock assessment settings and can certainly be considered to be adequate and appropriate.

I am however concerned that projections may be optimistic and do not express a degree of uncertainty that is intuitive given the historical data and historical predictions of the assessment model. Over the years 2005-2010, the base-case model has predicted consistent declines in spawning stock biomass from catches as low as 2560t (Table 1 below, Table 4 of the report). Recent recruitment is estimated to be high which presumably drives the model projections that lead to the conclusion "When catch is reduced 20% from current level (average 2007-2009) which is about 2,500mt, the stock is projected to have zero chance to fall below 2012 level for both states of nature" (page 43). To have zero probability of falling below 2012 levels by 2017 implies a degree of certainty in projections at 2,500mt y^{-1} that does not appear to be consistent with the past data and projections (Table 1 below, Table 4 of the report). In my view this indicates that the projections may be somewhat optimistic and that recovery of the stock at 2,500mt may not be as pronounced as predicted here.

Table 1 (Table A of the report, page 4). Observed catch and base-case model predictions of stock status.

Year	2004	2005	2006	2007	2008	2009	2010	Mean[1]	Min[1]	Max[1]
Reported Catch	4047	3703	3706	3195	3691	2560	2560[2]	6011	2560	10528
Population Biomass	11679	9545	10371	8430	7414	5335	6625	14141	5335	24886
Spawning Biomass	1731	2010	1992	1824	1625	1106	938	2439	909	5104
Relative Spawning Biomass	0.64	0.74	0.73	0.67	0.60	0.41	0.35	0.90	0.33	1.88
Recruitment (age 0)	116	434	125	204	133	349	326	453	116	1620
Fishing Mortality	0.58	0.56	0.62	0.58	0.86	0.84	0.75	0.79	0.53	1.46
Relative Fishing Mortality	1.22	0.95	0.92	1.01	0.95	1.41	1.37	1.30	0.86	2.38
Exploitation Rate	35%	39%	36%	38%	50%	48%	38%	44%	29%	69%
Spawning Potential Ratio	19%	19%	17%	19%	12%	13%	14%	14%	7%	21%

4.5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.

Increasing the transparency of the CPUE standardization process may help to arrive at methods that best characterize declines in relative abundance. This could involve making CPUE data available or allowing standardization models to be investigated interactively during a stock assessment meeting.

A research priority is undertaking simulation evaluation to identify management strategies and assessment models that are robust to uncertainty in stock and fishery dynamics. This type of simulation may also provide a formal basis for selecting abundance indices and understanding the value of additional information in terms of making reliable management decisions. For example, how cost effective is a conventional tagging programme (given the 1% return rates) in improving management performance?

It may be useful to investigate the cost-efficacy of newer tagging technologies that do not rely on reporting of tags (*e.g.* an integrated PIT tag – reader system) that may offer an alternative source of information regarding current stock size and fishing mortality rate.

5. Conclusions

Strengths

The assessment was conducted using established software that has been subject to extensive review in other stock assessment settings.

The assessment makes good use of updated information regarding the biology and ecology of the stock.

The assessment is very detailed particularly in its attempt to capture the historical fishing mortality at size of multiple fleets.

The results of the assessment are clearly presented in a range of informative figures and tables.

In general, the authors do an excellent job of describing the model and discussing the implicit assumptions.

A comprehensive range of sensitivity analyses and projections is another key strength of the report.

Weaknesses

In many cases an assessment may be evaluated in terms of its ability to predict the trends inferred by the relative abundance indices that are provided by stakeholders. These derived data may therefore directly determine one of the most important reference points of the assessment, stock depletion. Small changes in the standardization approach can have dramatic impacts on the inferred abundance trends. In most cases the standardization research used by this report is not presented in a manner that is consistent with basic standards of scientific publication elsewhere. Methods are not described in sufficient detail to be reproducible. The inherent assumptions of the standardization methods are not made clear or discussed. It is not possible to understand whether the authors of these indices have satisfactorily accounted for a number of potentially serious biases that affect CPUE standardisation that could preclude the use of the index in stock assessment.

The gathering and processing of critical inputs such as length composition data are not described in sufficient detail for a reader to gain an intuition of the quality of these data.

The complexity of the modelled fishing dynamics appears disproportionately high relative to the probable quality of the data and the size of the assumptions regarding spatial population dynamics. Given their experience and expertise there is good reason to trust the authors' judgement. However in its own right the document does not include sufficient detail regarding convergence diagnostics and evidence against model overparameterization. These are a precondition for interpreting assessment outputs and based on the report alone it is difficult to be confident about the robustness of the assessment results.

The assessment document does not include enough detail about model structure such as seasonality (*e.g.* seasonal transition equations that include recruitment).

In several cases the base-case model appears to include unnecessary complexity which is undesirable (lack of sensitivity to spatial disaggregation of the Japanese longline index for example).

The use of regional data to infer population-wide stock dynamics is questionable and may lead to results that are biased by regional characteristics.

The CPUE standardization methods of different fleets rely on assumptions about population dynamics that conflict with one another and the assessment.

The index derived from the fleet with probably the best spatio-temporal coverage (the Japanese longline fleet) is severed into discrete times and areas strongly reducing the extent to which these data inform stock depletion.

6. Recommendations

Conduct simple tests of model overparameterization.

Having developed a range of spatial multi-fleet operating models, it is my experience that models of the complexity presented here may not robustly estimate variables of management interest such as $F_{current}/F_{MSY}$ and $B_{current}/B_{MSY}$. This can be tested relatively easily by using the data-generation facility of Stock Synthesis to produce simulated fishery data in order to determine how well the model can retrieve known parameter values. It should be noted that such a test is likely to offer an optimistic evaluation of assessment performance since the same model structure is used to simulate the data. This is a 'first test' for overparameterization and will not help to diagnose problems due to differences between the real observation processes and dynamics and those assumed by the assessment model. Other tests include variability in the estimation of management reference points from different 'jittered' starting locations.

Include a more comprehensive evaluation of structural uncertainty: examine the marginal effect of adding different levels of complexity (seasonality, multiple fleets).

It would be instructive to see future assessments carried out in parallel with simple approaches that are quick and easy to apply. For example, delay-difference models or age-structured production models. For example a model assuming annual population and fishery dynamics, similar selectivities (1) among all longline fleets and (2) all other fleets and fitted to a single relative abundance index such as the Japanese longline fleet.

Identification of clear management objectives including target and limit reference points

The authors have done remarkably well at assessing a stock without established yardsticks with which to evaluate stock status and exploitation rate. It is paramount that managers establish target and limit reference points for the WCNPO striped marlin stock. Without a clear statement of what managers wish to achieve and avoid, it is not clear how to draw conclusions regarding the stock, make recommendations or develop management strategies. Currently, progress in the development of the quantitative tools with which to inform management may be hamstrung by a lack of clarity in objectives.

Greater clarity and rigor in the presentation of CPUE standardization methods

Given their importance for stock assessment, future CPUE standardization papers should endeavour to meet fundamental standards of scientific publication such as the clear description of a reproducible method and a transparent account of implicit assumptions (so that it is clear whether these are aligned with the assessment). The CPUE data may be confidential but at the very least the methods should be described in such a way that they may be applied to other CPUE datasets. Future standardization approaches should include maps of spatial coverage, graphs of the index itself (with an expression of uncertainty) and a suite of sensitivity analyses. Importantly the equation $I_t = \ldots$ should be included where I is the derived index and t the time subscript. While not preferable to annotated equations, a simple solution that could help meet this requirement would be for each standardization paper to include an Appendix of the computer code used to conduct the standardization.

A greater degree of theoretical consistency between CPUE standardization models and the assessment

Several CPUE GLM models assume time x area interactions or year x season interactions that imply population dynamics that are not accounted for by the spatially aggregated model. The model should avoid the use multiple regional abundance indices of equal prior weight (*e.g.* JP_DWLL1-3). If time x area interactions must be accounted for, this should be limited to a single standardization that models these interactions simultaneously and produces a single index weighted by the magnitude of regional abundance.

A greater degree of theoretical consistency among CPUE standardization models

We can see that year x season interactions, time x area interactions, sea surface temperature and set depth are all important factors in standardization of striped marlin CPUE. Some of these characteristics such as set depth are particular to specific gears. However where possible the methods should be consistent to avoid contradictory assumptions regarding the dynamics of the same stock.

Provision of CPUE data to increase the transparency of standardization and allow for cross-evaluation of methods

I acknowledge the difficulties in making publicly available, the fine-scale commercial catch and effort data with sufficient covariate data to investigate defensible standardization methods. For example, Japanese longline trip-level data at the 1x1 degree resolution with covariate hook-per-basket information, species composition, sea conditions, *etc*. However, since the derived relative abundance indices are critical to the assessment it is highly desirable that stock assessment scientists are provided with the basis

to determine standardization procedures that may provide more consistent information about relative abundance across fleets.

In many stock assessment settings there are two central reference points, stock level (depletion, underfished / overfished) and exploitation level (underfishing / overfishing). I find the current *status quo* worrisome; that stakeholders arrive at the assessment with input data (the indices) that may largely predetermine one of these two key reference points (depletion). In my view, CPUE standardization is a critical part of stock assessment (interpreting fishery data in terms of abundance trends) and should be reviewed and investigated in the assessment meeting (perhaps this already occurs). It would, for example, be highly desirable for scientists involved in index calculation to attend the meeting with working code and allow the assessment group to interactively investigate the sensitivity of derived indices to different assumptions.

Simulation evaluation

It is important to understand where the trade-off lies in terms of assessment complexity versus reliability of management decision making. Consider the following hypothesis that may be tested by a multi-fleet operating model with some spatial population dynamics: quota recommendations derived from a simple one-stock, two fleet (longline and other) delay-difference model provides comparable management performance (assuming we have agreement on management objectives) to those of a more complex multiple-fleet, fully age-structured SS3 model (such as that applied here). This may well be rejected, but in other simulation settings simple assessment methods have provided a comparable and sometimes better basis for decision making whilst being more straightforward to apply and review. More complex assessments such as that applied here may offer detailed insights into the past characteristics fishing and the population but is that the objective of stock assessment? From my perspective it is important to know whether the additional complexity of such an assessment can be expected to offer practical benefits in terms of managing the stock.

Simulation evaluation also provides a basis for establishing management strategies that are robust to uncertainties of the population and fleet dynamics. For example, the regional abundance indices described may be valuable in defining spatial operating models to evaluate the magnitude of the single mixed WCNPO region and devise management procedures that perform well subject to a range of credible spatial hypotheses (seasonality in migration, incomplete stock mixing, *etc.*).

Spatial simulation models may provide a more objective basis for eliminating candidate abundance indices and determining the potential value of collecting other sources of data.

Consider a simpler base-case assessment model

It may be the case that a similar model with annual structure inferred by a two fleets (longliners and other fleets) fitted to a single relative abundance index (Japanese longline offshore) could provide credible management advice. The scientists involved in the design of the assessment are experts with a very good reputation for high- quality work. However there is simply not enough detail presented in the document to be sure whether the model is overly complex and capable of producing spurious predictions. My suspicion is that the assessment is simply asking too much from the data that are available.

Note that if the current assessment were to include many of the recommendations made here the document would be much larger and take substantially more resources to produce. This is also a problem associated with more complex assessment models; they pose difficulties for both the presentation and review process due to the manifold permutations of assumptions that may determine model behaviour in a complex way. If a comprehensive review is a key requirement of a stock assessment, a model of greater simplicity may be desirable from a practical standpoint.

Avoid using regional abundance indices to infer population-wide stock dynamics.

Instead, fit the base-case model to a single index of abundance derived from a fleet with good spatial coverage and important covariate information (in this case this may be the Japanese offshore longline fleet) and use other indices for sensitivity analysis.

Where possible, avoid the division of indices into temporal blocks to account for changes in fishing that may be incorporated in CPUE standardization.

Appendix 1. Bibliography

BILLWG. 2011b. Report of the Billfish Working Group Workshop. International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean. 24 May – 1 June 2011. Chinese Taipei, ROC.

Campbell, R.A. 2004. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fish. Res. 70(2–3): 209–227.

Carruthers, T.R., McAllister, M.K. and Ahrens, R.N.M. 2010. Simulating spatial dynamics to evaluate methods of deriving abundance indices for tropical tunas. Can. J. Fish. Aqua. Sci. 47: 1409-1427.

Kadane, J.B., Lazar, N.A. 2004. Methods and criteria for model selection. J. Amer. Statistical. Assoc. 99: 279-290.

Kanaiwa, M., Kimoto, A., Yokawa, K. 2011. Standardized CPUE of striped marlin in north western Central Pacific Ocean by using GLM. ISC/11/BILLWG-2/06.

MAR&SWOWG. 2007. Report of the marline and swordfish working group joint workshop. International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean. Annex 8. http://isc.ac.affrc.go.jp/pdf/ISC7pdf/Annex_8_ISCMARSWOWG_Mar_07.pdf

Miyake, P.M. 2004. A brief history of the tuna fisheries of the world. *In* Second Meeting of the Technical Advisory Committee of the FAO Project Management of Tuna Fishing Capacity: Conservation and Socio-economics, Madrid, Spain, 15–18 March 2004 *Edited by* W.H. Bayliff, J.I. De Leiva Moreno, and J. Majkowski. Food and Agriculture Organization of the United Nations, Rome, Italy. pp. 23–50.

Piner, K. Lee, H.H. 2011a. Meta-analysis of striped marlin natural mortality. ISC/11/BILLWG-1/10.

Piner, K. Lee, H.H. 2011b. Correction to Meta-analysis of striped marlin natural mortality. ISC/11/BILLWG-2/08

Piner, K.P., Conser, R., and DiNardo, G. 2006. Stock status of striped marlin in the north Pacific Ocean in 2005. Marlin and Swordfish Working Group Joint Meeting. 20-22 March, 2006, La Jolla, California, U.S.A. ISC/06/MARWG&SWOWG/07.

Piner, K.P., Conser, R., DiNardo, G., and Brodziak, J. 2007. SS2 sensitivity runs for striped marlin assessment WG 2007. Marlin and Swordfish Working Group Meeting. 19-26 March, 2007, Taipei, Taiwan. ISC/07/MARWG&SWOWG-1/02.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011a. Reproductive biology of female striped marlin (Kajikia audax) in the waters off Taiwan (preliminary). Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/11: 14p. Available at: http:// isc.ac.affrc.go.jp/ pdf/BILL/ ISC11_BILL_3/ ISC11BILLWG-3_WP01.pdf or http:// isc.ac.affrc.go.jp/ reports/ bill/ bill_2011_1.html1

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011b. Reproductive biology of male striped marlin, Kajikia audax, in the waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/09: 17p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011c. Age and growth of striped marlin (Kajikia audax) in waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/09: 15p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP09.pdf or http://isc.ac.affrc.go.jp/reports/bill/bill_2011_1.html1

Sun, C.-L., Su, N.-J., Yeh, S.-Z., Chang, Y.-J. 2011d. Standardized catch-rates for striped marlin (Kajikia audax) for Taiwanese distant-water longline fishery in the North Pacific Ocean for 1967-2009. ISC/11/BILLWG-1/07.

Yokawa, K. 2005. Standardized CPUE of striped marlin caught by Japanese (High Seas) large-mesh drift fishery in the north Pacific for the periods between 1977 and 1993. ISC\05\MAR&SWOWG\18.

Walsh, W.A., Lee, H.H. 2011. Standardization of Striped Marlin Kajikia audax CPUE with Generalized Linear Models fitted to Pelagic Longline Observer Data from the Hawaii-based Fishery in 1995–2009. ISC/11/BILLWG-1/08.

Appendix 2. CIE statement of work

Attachment A: Statement of Work for Dr. Tom Carruthers

External Independent Peer Review by the Center for Independent Experts

Stock Assessment of Striped Marlin

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide an impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Striped marlin (*Tetrapturus audax*) is one of six species of billfishes commonly harvested multi-nationally from commercial and recreational fisheries in the western and central Pacific Ocean regions. Fishery management requires high quality science to effectively manage and conserve our living marine resources, and the scientific peer review of stock assessments by external CIE expertise is an important process in the determination of best scientific information available. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise, working knowledge and recent experience in the application of fish stock assessment, mathematical modeling, and statistical computing. Scientists who are employed by or have significant interactions with the Inter-American Tropical Tuna Commission (IATTC) and the Western and Central Pacific Fisheries Commission (WCPFC), and the Secretariat of the Pacific Community (SPC), should not be considered as reviewers. Scientists associated with the ISC also should be excluded as reviewers. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country,

address, email) to the COR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, and other pertinent information. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than **18 November 2012**, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, and to Dr. David Die, CIE Regional Coordinator, via email to ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

18 October 2012	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
25 October2012	NMFS Project Contact sends the CIE Reviewers the report and background documents
1-16 November 2012	Each reviewer conducts an independent peer review as a desk review

18 November 2012	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
1 December 2012	CIE submits the CIE independent peer review reports to the COR
7 December 2012	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Gerald DiNardo, Stock Assessment Program Leader (NMFS Project Contact)
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Gerard.DiNardo@noaa.gov Phone: 808-983-5397

Kevin Piner
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Kevin.Piner@noaa.gov Phone: 858-546-7003

Jon Brodziak
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Jon.Brodziak@noaa.gov Phone: 808-983-2964

Hui-Hua Lee
University of Hawaii, Joint Institute for Marine and Atmospheric Research
2570 Dole St., Honolulu, HI 96822
Huihua.Lee@noaa.gov Phone: 808-983-5352

Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Appendix 3. Typographical, grammatical errors *etc*.

Page 17, third paragraph, fourth sentence: "Japan longline fleets were targeting predominantly albacore for canning and occasionally caught striped marlin at the surface waters, whereas harpoon fisheries operating in coastal waters of Japan directly targeted striped marlin were."

Page 20, third paragraph, second sentence: "Although some fisheries have catch data time series extending back to at least 1952 and model were developed in parallel that included this early data…"

Page 24, first paragraph, second sentence. Incomplete reference.

Page 29, last paragraph, second sentence. "More uncertain dataset due to…"

Page 39, paragraph 4, first sentence. "The BILLWG identified important sebsutuvuty runs…"

Page 39, paragraph 5, second sentence. "Two indices (2 time periods) were used to replace night indices (3 area times 3 time periods)..."

CIE Independent Peer Review Report

on

Stock Assessment of Striped Marlin, *Tetrapturus audax,*

in the Western and Central North Pacific Ocean

Prepared by

Yong Chen

Professor of Fisheries Population Dynamics
School of Marine Sciences
University of Maine
Orono, ME 04469

Contents

SECTION **Page**

I. Executive Summary…….3

II. Background……..5

III. Description of the Individual Reviewer's Role in the Review Activities……...7

IV. Summary of Findings…….7

V. Conclusions and Recommendations……..15

VI. References……19

VII. Appendices……..21

VII-2. List of Documents ……....21

VII-1. Statement of Work for Dr. Yong Chen…….…23

I. Executive Summary

Striped marlin, *Tetrapturus audax*, supports multinational and multi-gear fisheries in the Pacific Ocean. Two stocks are defined in the North Pacific Ocean: the Western and Central North Pacific (WCNPO) stock and Eastern North Pacific stock. The WCNPO stock is the focus of this stock assessment covering the time duration from 1975 through 2010 and the waters to the west of 140^0W in the North Pacific Ocean. The Billfish Working Group (BILLWG) of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) is responsible for conducting regular assessment for the WCNPO stock.

Stock Synthesis (SS, version 3.11b; Methot 2011) was used as the stock assessment modeling platform to provide estimates of stock parameters and an age-structured projection model was used to project the stock dynamics from the time period from 2010 through 2017 under six different levels of constant fishing mortality and two levels of constant catch. Based on the stock assessment and projection, the WCNPO striped marlin stock status was determined and scientific advices on fisheries management were provided. Both the assessment and projection models implicitly assume that there is a single well mixed stock of striped marlin in the WCNPO. The assessment model uses seasons (quarters) as its time step and includes eighteen fisheries and the projection model uses year as its time step.

Based on temporal and spatial variability in fishing operation, hook-per-basket distribution, targeted species and length distribution of fish, fifteen sets of standardized annual relative abundance indices were developed for eight fisheries, including ten Japanese longliner indices, two driftnet indices, two Taiwanese longliner indices, and one Hawaii-based longliner index. The timing (i.e., season) of each index was determined by the timing when the most landings were recorded for the fishery. No fishery-independent data were available. The BILLWG developed a base-case model for the assessment, and conducted a well-planned and structured sensitivity analysis to evaluate impacts of different input data and model configurations and parameterization on stock assessment. The BILLWG concluded that the WCNPO striped marlin stock was overfished and overfishing occurred and recommended that the current fishing mortality be reduced to allow for increased stock biomass. The BILLWG further projected how the stock biomass might change with different levels of fishing mortality/catch.

I independently evaluated the WCNPO striped marlin stock assessment report with respect to a set of pre-defined Terms of Reference. *I conclude that overall this stock assessment is based on the best science available. I conclude that this assessment is scientifically sound and adequately addresses needs for management advice. I agree with the conclusion regarding the WCNPO striped marlin stock status and management advice made in the stock assessment report.* In particular, I would like to commend the efforts of the BILLWG for compiling updated fisheries and biological data for the multinational and multi-gear fisheries and conducting a well-planned and structured stock assessment and sensitivity analysis to evaluate and address uncertainty regarding data quality and quantity and model configuration and parameterization. However, I believe some important questions did not receive enough attentions or were not addressed in the assessment. These issues include lack of retrospective

analysis; lack of considering all uncertainty in the projection; and failure to explicitly define target and limit reference points for stock biomass and fishing mortality and relevant harvest control rules. I also believe more studies are needed to further improve the quality of fisheries data and biological information on the stock spatial structure and spatial variability in key life history processes such as growth and maturation.

Accordingly, I recommend that future research be done in the following areas: (1) develop a management strategy evaluation (MSE) framework to evaluate the performance of the striped marlin stock assessment model and identify key assumptions that may significantly influence the model performance; (2) conduct more studies to evaluate the quality of the input data and the consistency of the data from different fisheries and reduce the uncertainty in the data before they are used in modeling; (3) evaluate possible spatial and temporal variability in fish life history parameters (e.g., growth and maturation) and fisheries data (e.g., catch, catch size compositions, and CPUEs) and coordinate research efforts to collect samples over a large spatial scale; (4) conduct retrospective analysis to evaluate possible retrospective errors associated with stock biomass, recruitment, and fishing mortality estimated for the recent years and calculate Mohn's rho to explicitly describe and quantify the nature and magnitude of retrospective errors; (5) explore the use of the dynamic binning option in the SS or robust multinomial likelihood functions to reduce impacts of non-informative zero-observation or outliers in size composition data on the model fitting; (6) plot each set of CPUE against the estimated stock biomass to evaluate if gear saturation exists; (7) evaluate likelihood profiles for a range of values for steepness h and natural mortality M to determine if they can be assumed independently in input data; (8) evaluate the roles of recruitment deviation penalty functions in estimating annual recruitment deviation; (9) evaluate the performance of MSY-based biological reference points in the management of the WCNPO striped marlin stock and identify alternative biological reference points (e.g., some historical fishing mortality and stock biomass); (10) examine the uncertainty associated with biological reference points and its impacts on the determination of stock status; (11) evaluate potential impacts of discrepancies between the stock assessment and projection model outlined in Table 5 (BILLWG 2012) on the evaluation of stock status; (12) consider more measures for comparing performance of different management options; (13) consider an alternative management time period, other than eight years, in the projection in order to identify possible differences in long-term and short-term projections; and (14) develop priors for key fishery and population parameters and apply the Bayesian estimator to better quantify uncertainty in stock assessment and projection. More detailed recommendations and their justifications can be found in the sections of Summary of Findings and Conclusions and Recommendations.

II. Background

Striped marlin, *Tetrapturus audax*, supports multinational and multi-gear fisheries in the Pacific Ocean. One unit stock was assumed in the North Pacific Ocean in the 2007 stock assessment. More recent studies suggest that there are two distinct genetic populations in the North Pacific Ocean (McDowell and Graves 2008, Purcell and Edmands 2011). As a result, two stocks are defined in the North Pacific Ocean for the assessment: the Western and Central North Pacific (WCNPO) stock and Eastern North Pacific stock. The WCNPO stock is the focus of this stock assessment covering the time duration from 1975 through 2010 and the waters to the west of 140^0W in the North Pacific Ocean. The Billfish Working Group (BILLWG) of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) is responsible for conducting regular stock assessment.

Estimating growth of billfish species is difficult because of minute sizes of their otoliths and the challenge to obtain samples of good size coverage. The first growth study based on calcified parts in the North Pacific was done in 2003 (Melo-Barrera et al. 2003). The result was used in the last striped marlin stock assessment in 2007. However, the growth estimates for young striped marlin may not be reliable because of lack of young fish in the samples. A more recent study with samples of juvenile striped marlin suggests much faster growth rates for young striped marlin (Sun et al. 2011a,b), which is considered to more realistically quantify the growth of striped marlin. However, the studies are limited in their spatial and temporal coverage, and there is little information available on spatial and temporal variability in the growth.

Sexual dimorphism of striped marlin is related to spawning season and body size (Wang et al. 2006). Male striped marlin tend to mature at a smaller size than females (Kopf et al. 2009, and Sun et al. 2011c,d). Large spatial-temporal variability is evident in the maturation and reproduction of striped marlin in the North Pacific Ocean (Kopf et al. 2009).

Analyses of temporal and spatial variability in the fishery CPUE and catch size composition indicate that the striped marlin population in the North Pacific moves to higher latitudes during summer, but there is no evidence to support trans-ocean movement. Various tagging studies suggest the lack of trans-Pacific and trans-equator movement by striped marlin (Sippel et al. 2011), with ambient oceanographic current being one of main factors influencing individual movement. Vertical movement was mainly limited from the surface to the mixed layer above 90 m depth and regulated by relative changes in water temperature with depth. More studies are still needed to have a better understanding of the extent of movement in the northwest Pacific and into the Hawaiian region.

Most of striped marlin catch is harvested using longline, driftnet and harpoon by Japan, USA, and Taiwan in the WCNPO. Japanese fishing fleets dominated the fishery in the 1950s and 1960s, and striped marlin were caught in longline fisheries targeting albacore and were targeted in harpoon fisheries in coastal waters of Japan. Longline catches of striped marlin reached the highest level in the late 1960s. During the 1970s and 1980s, longline fisheries moved into deeper waters in more tropical waters targeting adult bigeye tunas, where striped marlin were less abundant. This shift of spatial distribution of fishing effort might result in the

reduced striped marlin catch in the 1970s. Catches have continued to decline from approximately 6,000 mt per year in the 1990s to 4,200 mt per year in the early 2000s and 3,500 mt per year during 2005-2008. Reported catch of 2,560 mt in 2009 was the lowest catch reported since 1952. However, this did not necessarily result from the decline in fish abundance; but rather it reflects changes in spatial distribution of the fishery which expanded eastward in WCNPO in the 1950s and 1960s but then reduced in the 1990s and 2000s (MAROWG 2006).

Seventeen fisheries were initially identified, however, a preliminary analysis suggested that a residual pattern and quarterly size composition data from the "Japan other fishery" showed a strong seasonal pattern. Thus the "Japan other fishery" was divided into two separate fisheries: early (seasons 1-2) and late (seasons 3-4) fisheries. Subsequently eighteen fisheries were defined in this assessment based on country/regions, gear, spatial coverage, and season to minimize spatial/temporal variability in selectivity and catchability. These fisheries include: nine longline (USA, Japan coastal, Japan offshore and distant water by area, Japan other seasons 1-2, Japan other seasons 3-4, Chinese Taipei offshore and distant-water, and Korea), two driftnet (Japan high sea and coastal large-mesh and Japan squid), one bait (Japan), one trap (Japan), one set net (Japan), two harpoon (Japan), one coastal fishery (Taiwan offshore and coastal gillnet, coastal harpoon, coastal set net and other) and one miscellaneous longline (WCPO data). These fisheries vary greatly in their spatial and temporal coverage, selectivity and catchability, and differ in nature (striped marlin are targeted or bycatch species).

Catch and size composition data were estimated and compiled by seasons (Jan-March, Apr-Jun, Jul-Sep, and Oct-Dec) for these fisheries from 1975 through 2010. The 2010 catch was assumed to be the same as the 2009 catch because the 2010 catch data were incomplete. Strong seasonal patterns were observed in catch and such seasonality in catch differs among the 18 fisheries defined in this stock assessment. Quarterly length composition data, measured as lower jaw fork length (LJFL) and compiled in 5-cm size bins from 55 to 230 cm, were available for eleven fisheries. The length frequency data represent actual number of striped marlin measured. Because of large spatial and temporal variability in timing of recruitment and rapid growth in early ages, the first size bin was set at 120 cm LJFL, which essentially acts as an accumulation for fish smaller than age 1 size. Based on temporal and spatial variability in fishing operation, hook-per-basket distribution, targeted species, and length distribution of fish, fifteen sets of standardized annual relative abundance indices were developed for eight fisheries, including ten Japanese longliner indices, two driftnet indices, two Taiwan longliner indices, and one Hawaii-based longliner index. The timing (i.e., season) of each index was determined by the timing when most landings were recorded for the fishery. Although different in timing and magnitude of decline, these abundance indices tend to suggest a decreased abundance in the 2000s. Except for the two Taiwan longliner indices for which constant CVs of 0.2 and 04 were used for all the years, the coefficients of variation (CVs) of the standardized indices for other indices were derived from GLM models. The CVs were used to essentially quantify the quality of these data in the assessment.

Stock Synthesis (SS, version 3.20b; Methot 2011) was used as a modeling platform for the development of a seasonal, length-based, age-structured, forward-simulation population model for the current assessment of the WCNPO striped marlin stock (BILLWG 2012). A previous

ISC striped marlin assessment was done in 2007 using Stock Synthesis 2. Seven major differences between the 2007 and current stock assessments were identified including different assumptions on stock structure, selectivity and initial stock condition; different assumed values for steepness parameter (h), natural mortality (M), and growth and maturity parameters; and different time period covered.

The BILLWG conducted the stock assessment to estimate key fishery and population parameters. Various sensitivity analyses were done to evaluate potential impacts of assumptions made implicitly and explicitly in modeling. Various weighting schemes were evaluated and iterative modeling approaches were used in identifying relative weights for CPUE and size composition data of different sources and different quality to improve internal model consistency. Various analyses were done to compare model fitting of different selectivity functions and time blocks to identify optimal selectivity functions and time blocks for different fisheries. Population projection was done using an age-structured model (Punt 2010) for the time period from 2010 through 2017 (but with a year starting on July 1 and ending on June 30) to evaluate the performance of six levels of constant fishing mortality rates and two levels of constant catch in conserving SSB (measured as the ratio of SSB values between 2017 versus 2010). The population projection model is different from the SS in model structure (see Table 5 in BILLWG 2012 for details). Uncertainties in recruitment dynamics were considered in the projection. However, no probability distribution for the ratio of SSB values between 2017 versus 2010 was estimated. No target and limit biological reference points were explicitly specified and no harvest control rules were developed.

III. Description of the Individual Reviewer's Role in the Review Activities

As the SoW states that "*Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs*", my role as a CIE independent reviewer is to conduct an impartial and independent peer review of stock assessment of striped marlin in the Western and Central North Pacific Ocean which are fished by multiple nations with multiple gears, with respect to the pre-defined Terms of Reference.

This is a desk review. Thus, I have no opportunity for face-to-face discussion and questioning. I read the "Stock assessment of striped marlin in the Western and Central North Pacific Ocean" by the Billfish Working Group of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean and all other background documents that were sent to me (see the list in the Appendix II). I also read references relevant to the topics covered in the reports and the SoW. I address each topic covered in the ToRs, evaluate the strengths and weaknesses of what was done in this assessment, and provide recommendations to improve future assessment. Based on these evaluations and analyses, I make research recommendations for future assessment of striped marlin in the Western and Central North Pacific Ocean.

IV: Summary of Findings

1. Review the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.

   Given the complexity of the fishery with multiple fishing fleets and gears from multiple nations and large temporal-spatial variability in the fishing operation and striped marlin life history processes, I conclude that the SS employed is the most appropriate assessment platform which can provide an adequate modeling framework for the assessment of the WCNPO striped marlin. I further conclude that the stock assessment models are properly implemented with caution being paid to the evaluation of data quality and quantity, implicit and explicit assumptions, and alternative modeling options through preliminary analyses, sensitivity analyses, iterative weighting in modeling, and careful evaluation of model fitting and residual patterns. The base-case assessment scenario developed by the BILLWG appears to represent the best knowledge available with respect to the WCNPO striped marlin fisheries.

   However, I could not find that retrospective analyses were done for evaluating possible retrospective errors in the estimation of SSB, recruitment and fishing mortality. The projection model differs from the stock assessment model in time step, model structure, timing for calculating SSB, selectivity and timing for applying M, which may result in inconsistency and add extra uncertainty in the population projection. The assessment only considered the management options of constant fishing mortality rates or constant catch and no target and limit biological reference points were explicitly defined for fishing mortality and stock biomass. Not all uncertainty was considered in the population projection and in the determination of stock status.

   Stock Synthesis used in this assessment is a seasonal age-based, size-structured model. One of the greatest strengths of SS is its flexibility to utilize a wide diversity of age/size-based data and aggregate data of different sources and to account for temporal variability in catchability and selectivity. Using the SS framework, the BILLWG developed a seasonal, length-based, age-structured forward population projection model to predict fishery data (CPUEs and size composition data) which were then compared with corresponding observed data to formulate likelihood functions for the parameter estimation.

   I support the BILLWG's choice of using the SS for the assessment of the WCNPO striped marlin stock with data from such diverse sources. The stock assessment methods developed with the SS allow the modelers to incorporate data from fisheries with different gears and spatial and seasonal coverage and consider temporal changes in catchability and selectivity. The use of quarters as time step is consistent with seasonality of some fisheries described in the stock assessment report (BILLWG 2012). The updated information on the stock area, steepness parameter, natural mortality, growth and maturity parameters, selectivity, assessment duration and setting of the initial stock condition since the last stock assessment in 2007 improved the

assessment. Iterative estimation of effective sample sizes and data variances improved the internal consistency of model fitting. The use of standardized CPUE data removed the impacts of factors other than stock biomass.

Although I am impressed by this well-thought out, planned and structured stock assessment, I was surprised that no retrospective analysis was done for evaluating possible retrospective errors that are often associated with estimates of stock biomass, recruitment, and fishing mortality in the recent years in a stock assessment of this nature. Retrospective analyses are routinely done in stock assessment and I believe should be done in this assessment. I was also surprised at the inconsistency of the model structure and parameterization between the stock assessment model and projection model. Although the projection model is designed for the output from the SS model and widely used on the west coast, I think such an inconsistency may raise some issues in interpreting the results of the stock projection, complicating the interpretations of the ratio of stock biomasses of 2010 and 2017; which is used as the measure to compare the performance of different management options. This would be especially true when uncertainty is considered in the population projection.

2. <u>Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.</u>

   The BILLWG compiled quarterly catch and size composition data of different fisheries, developed standardized CPUEs and assembled updated biological and fisheries parameters for the assessment of this multinational and multiple-gear fishery. This is a major undertaking, and I commend the BILLWG for their efforts to compile such a comprehensive and updated data base for the WCNPO striped marlin stock assessment. I conclude that the stock assessment models are adequately configured, data are properly prepared, screened and used, and assumptions are reasonably satisfied.

   However, I believe uncertainty has not been fully considered. My concern is the failure to conduct retrospective analyses to estimate the magnitude and nature of possible retrospective errors that are often associated with the estimates of stock biomass, fishing mortality and recruitment in the recent years in the stock assessment. A lack of understanding of retrospective errors limits our understanding of the quality (and uncertainty) of the key fisheries parameters in the most recent years, which are used in the projection model for evaluating the performance of various management options. I suggest that retrospective analysis be done and that Mohn's rho (Mohn 1993, 1999) be calculated to explicitly describe and quantify the nature and magnitude of retrospective errors.

The BILLWG has explicitly described some assumptions with respect to assessment models, projection models, and statistical analyses. However, most of the assumptions are embedded in the texts and are sometimes hard to evaluate if such assumptions were violated. The possible consequences if some of the assumptions are violated are also unclear. I suggest that the BILLWG summarize all the assumptions about the model, data, and statistical analyses, explicit and implicit, in a table and describe if a particular assumption is satisfied for a given assessment scenario. Potential consequences of violating these assumptions should also be described in the table. I believe this can greatly help understand potential sources of uncertainty and improve the design of the sensitivity study in the assessment.

Eighteen fisheries were defined in the assessment. These fisheries have different spatial-temporal coverage, tend to target different components of the WCNPO striped marlin, and have different impacts on the stock dynamics. Relative weights of different data sets from different fisheries in the model fitting were mainly determined by the data quality measured by CVs for CPUE data and effective sample sizes for size composition data. I am curious if this is sufficient to reflect the relative importance of different fisheries in driving the dynamics of striped marlin stock. Maybe weighting factors for different likelihood functions should have been considered for the spatial coverage of relevant fisheries, instead all were set at 1.

Size compositions data are available for the eleven fisheries, and size ranges of the data tend to vary among the fisheries because of differences in gear, spatial, and temporal coverage of these fisheries. A size range of 120 to 230 cm LJFL was used to group size composition data from all the fisheries for which the size composition data were available. This might result in a large number of zeros for small and/or large size bins for some fisheries, which might result from limited sampling efforts or spatial coverage, rather than no fish in these size classes. This may affect the model fitting. I suggest that the dynamic binning option in the SS or robust multinomial likelihood functions (Fournier 1996; Chen et al. 2000) be used to reduce impacts of non-informative zero-observation in size composition data on the model fitting.

Choices of selectivity curves and time blocks assumed for different fisheries appear to be reasonable. The selectivity essentially includes both fish availability and gear selectivity. Impacts of changing fishing operations on catchability were considered in the model parameterization. The BILLWG also considered alternative choices of selectivity functions for different fisheries, and carefully evaluated and compared patterns of residuals in fitting size composition data with different selectivity functions to justify the choices of selectivity functions. However, I do not see much discussion about gear saturation. For longline and gillnet, gear saturation might be an issue, which can affect the reliability of CPUE as abundance indices even if it is standardized. A saturation parameter may be needed when CPUE is related to the stock abundance/biomass.

Sensitivity analysis was conducted to evaluate robustness of the modeling results with respect to alternative values for (1) quality of different data sets, which determines

weights of different data sets; (2) biological parameters; and (3) fishery parameters (i.e., selectivity and catchability in this assessment). Most sensitivity runs were conducted with just one parameter being given alternative values while other parameters were held constant as in the base case. Thus, the sensitivity analysis was essentially done for evaluating impacts of a single factor on the assessment. Such a design is important in understanding of roles of each factor. However, limited efforts to change more than one factor at the same time may result in lack of understanding of interactions of these factors. For example, steepness h and natural mortality M are usually negatively correlated, and should not be assumed independently in the assessment. Likelihood profiles should be evaluated for a range of values for these two parameters to identify their relationship and if they can be determined independently.

I did not see the description about the use of recruitment deviation penalty functions, which are usually applied to constrain annual recruitment deviations and prevent the model from yielding biologically unrealistic values for model parameters. I think this issue should be clarified and importance of the penalty functions should be evaluated in the sensitivity analysis (Methot and Taylor 2011).

The CPUE data were weighted in model fitting using CVs estimated from the GLM-based CPUE standardization. The GLM-estimated CVs were small (actually much smaller than what I have seen for other fisheries). Because of limited spatial and temporal coverage of a given fishery, the GLM-estimated CV for CPUE of the fishery is more suitable for describing local variability, but not for the whole stock range. When the CPUE is used to describe the population dynamics of the whole stock range, this level of CV under-estimates the variability. Thus, I believe the GLM-estimated CVs for CPUE data are not appropriate for being used in the likelihoods. However, the BILLWG used an iterative weighting approach to adjust weighting for the CPUEs, which reduces my concerns on this issue.

3. <u>Comment on the proposed population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.</u>

   The MSY-based reference points are estimated in the SS and are implicitly assumed to be limit reference points in the determination of stock status. However, I believe more research is needed in the evaluation of these reference points in their effectiveness of managing the fishery. Uncertainty of the estimated biological reference points and its implications should also be carefully evaluated. Overall I agree with the conclusion regarding the WCNPO striped marlin stock status. However, I believe uncertainty should be considered in the determination of stock status (although the conclusion regarding the stock status would be the same).

   Although the MSY-based biological reference points are estimated in the stock assessment, no limit and target reference points were explicitly defined in the risk analysis of alternative management strategies. However, the B_{MSY} and F_{MSY} were

implicitly defined as limit reference points because they were used in the assessment to conclude that the WCNPO striped marlin stock was overfished and the fishery was in the status of overfishing. Although the use of MSY-based biological reference points for limit reference points is rather common, the performance and implication of using them need to be carefully evaluated.

I recommend that management strategy evaluation (MSE) be developed for the WCNPO striped marlin stock. The MSE can then be used to evaluate and identify target and limit biological reference points. Only the constant fishing mortality (catch) harvest control rules were considered in this assessment. Such control rules lack the ability to adjust fishing mortality based on the updated status of fisheries. The MSE can be used for the evaluation and development of different harvest control rules.

Uncertainty associated with biological reference points was not considered and evaluated in this assessment. Given the uncertainty associated with fisheries and biological parameters, I suspect that the uncertainty may be considerable for biological reference points. The uncertainty associated with biological reference points is important in determining the status of stock.

The biological reference points were estimated in the SS, and used in the projection model which has different model structures and parameterization. I think it is necessary to evaluate the consistency in the performance of fishing mortality and stock biomass biological reference points in the projection model.

4. <u>Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.</u>

Overall I conclude that the methods used to project future population status are adequate. However, discrepancies of the assessment model and projection model need to be addressed. In particular, I recommend evaluating potential impacts of the discrepancies on the evaluation of different management options. I suggest that more measures, in addition to the ratio of stock biomass in 2011 and 2018, are used in evaluating management options. I also suggest that management strategies that can adjust exploitation rates based on stock status should be considered, rather than just considering constant fishing mortality or constant catch strategy.

The projection model used in the assessment was developed to incorporate the output from the SS and was used in the assessment of some groundfish populations on the west coast. However, I believe this is not sufficient to justify the use of this model in this study. The discrepancies of assessment and projection models were outlined in Table 5 in the assessment report (BILLWG 2012), but the potential impacts of such discrepancies on the evaluation of stock status were not evaluated and discussed.

Only constant fishing mortality (6 levels) and catch (2 levels) management strategies were considered in this study. This type of harvest control rule does not adjust fishing

mortality based on the stock status. I suggest developing MSE for the WCNPO striped marlin and evaluate and identify target and limit references for both fishing mortality and stock biomass and to consider and evaluate more forms of harvest control rules which allow the adjustment of fishing mortality based on stock status.

Two types of recruitment dynamics were used in the projection: recruitment randomly drawn from recent recruitment pattern (i.e., 1994-2008) and recruitment determined from the SR curve estimated based on SSB and R from 1975 to 2008 in the assessment model with annual deviation randomly drawn from residuals in SR modeling. The approach is sound and commonly used. However, an implicit assumption associated with the approach is that recruitment variability for the time period covered by the projection would be similar to the recruitment variability in the past. With non-random environmental changes (e.g., climate changes) more frequently observed in marine systems, such an assumption may be violated, which may introduce extra errors in the projection for a given set of management option.

I suggest that the BILLWG evaluate the performance of the projection model for the time period since the last stock assessment (i.e., 2007 to 2010). The BILLWG can use the projection model to project population dynamics from 2007 when the last assessment was done and compare the results derived in the current stock assessment for the same time period to see if the projection model yields feasible results. This can be repeated in a few years when the next stock assessment is conducted to evaluate the discrepancy of the stock biomass projected in the current assessment and stock biomass estimated in the next stock assessment for the time period from 2010 to the years when the next assessment is done.

I only found one measure was used to evaluate the performance of different management options. The measure used is ratio of stock biomass in the beginning and end of the projection period, which may not be sufficient to measure the performance of a management option. I suggest considering more measures (e.g., total catch, lowest stock biomass in the project period, among-year variability in catch). I also suggest quantifying the uncertainty associated with the ratio of stock biomasses at the beginning and end of the projection period to develop a probability distribution of this ratio.

I am a little puzzled by the use of eight years as the projection time period. This seems too short for the long-term projection but too long for the short-term projection. In any case, I suggest evaluating possible differences in long-term and short-term projections.

Maximum likelihood estimator was used in this assessment. I suggest that Bayesian estimators be used in the assessment to better quantify the uncertainty. The posterior distributions for key fisheries parameters derived from the Bayesian estimators can be used in the projection.

5. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.

I suggest the following research priorities to improve the WCNPO striped marlin stock assessment and our understanding of population and fishery dynamics. Some of the research priorities are for the long term, and some are for the short term (e.g., next assessment).

- Evaluate relationships between each pair of CPUE data sets to evaluate their consistency in indexing the temporal trend of stock biomass and to evaluate the coherence of CPUEs derived from different fisheries to identify factors that may influence the quality of the CPUE data and possible discrepancy among different sets of CPUE data, which can help determine relative weighting factors in the objective function;

- Plot each set of CPUE versus stock biomass estimates to identify possible gear saturation effects;

- Continue evaluating spatial and temporal variability in growth and maturation and if the information currently available is not sufficient for such a study, a new research program should be developed to improve our understanding of spatial and temporal variability in these key life history parameters;

- Continue evaluating spatial and temporal variability in fisheries data to identify the dynamics of spatial structure of fish size composition and CPUE to improve understanding of fisheries data quality and quantity and factors influencing them;

- Continue evaluating differences in key life history parameters between females and males to determine if a sex-specific stock assessment is necessary;

- Explore a dynamic binning approach to address potential issues of including too many size bins with non-informative zero to improve model fitting;

- Conduct retrospective analysis to evaluate retrospective errors for stock biomass, recruitment, and fishing mortality estimates;

- Explore the use of robust likelihood functions to identify outliers and then evaluate the identified outliers to determine if they should be removed (because of large measurement errors) or included (because of large process errors) in the assessment;

- Develop priors for key fishery and population parameters and apply the Bayesian estimator to better quantify uncertainty associated with modeling;

- Develop harvest control rules with explicitly defined target and limit reference points;

- Develop MSE for the WCNPO striped marlin stock to evaluate (1) alternative biological reference points, harvest control rules and management strategies; (2) impacts of data quality and quantity on the quality of stock assessment; and (3) impacts of violating some key assumptions on stock structure (e.g., possible meta-populations structure or large spatial variability in key life history process) and fisheries (e.g., selectivity and catchability).

V. Conclusions and Recommendations

The assessment appears to be well-planned and structured, scientifically sound, and adequately addresses needs for management advice for the WCNPO striped marlin stock. Uncertainties in the input data, fisheries and biological processes, and model parameterization were carefully evaluated in the stock assessment and projection. The conclusion on the stock status appears to be robust to uncertainty in the assessment and projection. In particular, I would like to commend the efforts of the BILLWG in addressing data quality issues, designing and conducting a well structured sensitivity analysis, exploring alternative model configurations and parameterization, and evaluating impacts of different fisheries on the assessment of stock dynamics.

However, I do have concerns that I hope the BILLWG could address to further improve the assessment of the WCNPO striped marlin stock. I made the following general comments and specific recommendations.

General comments

Although the new growth data were used in this study, they were derived based on samples collected in limited areas. I believe more study is still needed to improve the estimation of growth for the WCNPO striped marlin stock. Given the difficulty in sampling a large number of striped marlin over a large spatial area, I suggest that research efforts be devoted to develop approaches to back-calculate length-at-age data to derive length at each age for each fish. A nonlinear random effects model explicitly assuming that an individual's growth parameters are samples from a multivariate distribution can then be applied to the back-calculated length at age data (Hart 2001; Pilling et al. 2002) to estimate between-individual and between-region variability.

I am concerned with the failure to conduct retrospective analyses to estimate magnitude and nature of possible retrospective errors that are often associated with the estimates of stock biomass, fishing mortality and recruitment in the most recent years in stock assessment. A lack of understanding retrospective errors limits our understanding of the quality of the key fisheries parameter estimates in the recent years, which are used in the projection model for evaluating the performance of various management options. I suggest that Mohn's rho (Mohn 1993, 1999) be calculated to explicitly describe and quantify the nature and magnitude of retrospective errors for SSB, recruitment and fishing mortality.

Although the BILLWG considered uncertainty in data, model configuration and parameterization, and fisheries processes, some important sources of uncertainty were not considered in the assessment. No uncertainty associated with initial stock structure and biomass was considered in the projection. No uncertainty was considered in the estimation and use of biological reference points. No probability distribution was estimated for the ratio of stock biomass at the beginning and end of the projection, which is used as the measure in the evaluation of the performance of alternative management strategies.

Although the assessment model and projection model used in this study are well known and widely used, I believe the discrepancy of these two models for the WCNPO striped marlin stock may cause some concerns in the consistency and comparability of stock dynamics between the time period covered by the stock assessment model and that covered by the projection model.

No target and limit biological reference points are explicitly defined for fishing mortality and stock biomass. No harvest control rule was explicitly described, which makes the management advice unclear. Although the SS is flexible and has been tested and used in the assessment of many fisheries stocks, the results derived still need to be cross-validated to enhance the confidence in the assessment. I believe that the development of MSE needs to be a future research priority to evaluate the performance of the SS in quantifying the striped marlin stock dynamic, improve the risk analysis of alternative management strategies, and help identify key factors that may influence the quality of stock assessment and projection.

A Bayesian approach was not used in the assessment, and uncertainty in the assessment was not fully quantified in the stock assessment and projection for different harvest strategies. I encourage future assessments to develop priors for key model parameters and utilize the Bayesian estimator in the SS to incorporate uncertainty in the assessment and projection.

I suggest that the assessment model structure be kept relatively stable over time. If a new model needs to be used, it should be run in parallel to the old model to identify changes in stock assessment results resulting from changes in model configurations.

Specific recommendations

Although I have provided comments and recommendations under each TOR, I would like to re-iterate the following recommendations.

- I suggest evaluating spatial and temporal variability in growth and maturation. If the information currently available is not sufficient for such a study, a new research program should be developed to improve our understanding of spatial and temporal variability in these key life history parameters.

- I suggest evaluating spatial and temporal variability in fisheries data to identify the dynamics of spatial structure of fish size composition and CPUE to improve understanding of fisheries data quality and quantity and factors influencing them.

- I suggest evaluating the coherence of CPUEs derived from different fisheries to identify factors that may influence the quality of the CPUE data and possible discrepancy among different sets of CPUE data in quantifying the overall stock biomass, which can help determine relative weighting factors in the objective function.

- I suggest continuing evaluation of possible differences in key life history parameters between females and males to determine if it is necessary for a sex-specific stock assessment.

- I advise conducting a cross validation analysis that leaves some of the growth data out of the SS modeling for testing the growth model estimated within the SS.

- Given the quality of the data from different sources and potential errors in the data, it is likely to have outliers as a result of abnormal observational errors. It is also highly likely that outliers may arise as a result of abnormal process errors because of changes in the ecosystem over such a long time and over such a large area. I believe it is necessary to explore robust likelihood functions (Chen and Fournier 1999) to identify outliers and then evaluate the identified outliers to determine if they should be removed (because of large measurement errors) or included (because of large process errors) in the assessment.

- I suggest that the assessment model structure be kept relatively stable over time. If a new model needs to be used, it should be run in parallel to the old model to identify changes in stock assessment results occurring from changes in model configurations.

- I suggest that retrospective analysis be conducted to evaluate possible retrospective errors associated with stock biomass, recruitment and fishing mortality estimated for the recent years and that Mohn's rho (Mohn 1993, 1999) be calculated to explicitly describe and quantify the nature and magnitude of retrospective errors.

- I suggest that all the assumptions about the model, data, and statistical analyses, explicit and implicit, be summarized in a table and potential consequences of violating these assumptions be described in the table to help understand potential sources of uncertainty and improve the design of sensitivity study in the assessment.

- I suggest that the dynamic binning option in the SS or robust multinomial likelihood functions (Fournier 1996) be used to reduce impacts of non-informative zero-observation in size composition data on the model fitting.

- I suggest that each set of CPUE used be plotted against the stock biomass estimated in the assessment to evaluate possibility of gear saturation. If gear saturation is found to exist I suggest that a saturation parameter be added in the observational models linking the observed CPUE with stock abundance/biomass.

- I suggest that likelihood profiles be evaluated for a range of values for steepness h and natural mortality M to identify impacts of different combinations of these two values on

the assessment results and to determine if they can be assumed independently in input data (which was essentially how their values were determined in the assessment).

- I suggest that a description of the use of recruitment deviation penalty functions, which are usually applied to constrain annual recruitment deviations and prevent the model from yielding biologically unrealistic values for model parameters, be added to clarify the roles of the penalty functions.

- I suggest that the use of MSY-based biological reference points for limit reference points be evaluated for their performance and implication in the management of the WCNPO striped marlin stock and that alternative biological reference points (e.g., some fishing mortality and stock biomass in the past) be considered, evaluated, and compared with the MSY-based reference points.

- I recommend that management strategy evaluation (MSE) be developed for the WCNPO striped marlin stock. The MSE can then be used to evaluate and identify target and limit biological reference points and different types of harvest control rules.

- I suggest that the uncertainty associated with biological reference points be evaluated and its impacts on the determination of stock status be examined.

- I suggest that discrepancies (Table 5 in BILLWG 2012) be examined and discussed between the assessment and projection models with respect to their impacts on the evaluation of alternative management strategies.

- I suggest that MSE be developed for the WCNPO striped marlin to include both target and limit reference points for both fishing mortality and stock biomass and to consider and evaluate more forms of harvest control rules for allowing the adjustment of fishing mortality based on stock status in fisheries management.

- I suggest that the performance of the projection model be evaluated. The BILLWG can use the projection model to project population dynamics from 2007 when the last assessment was done and compare the results derived in the current stock assessment for the same time period to identify if the stock assessment model and project model yield consistent results and if the projection approach yields feasible results.

- I suggest that more measures be considered in evaluating performance of different management options (e.g., total catch, lowest stock biomass in the project period, among-year variability in catch).

- I suggest that alternative management time period, other than eight years, be considered in the projection to identify possible differences in long-term and short-term projections.

- I suggest that Bayesian estimators be used in the assessment to better quantify the uncertainty. The posterior distributions for fisheries parameters derived from the Bayesian estimator can be used in the projection.

VI. References cited

Billfish Working Group. 2012. Stock Assessment of Striped Marlin in the Western and Central North Pacific Ocean. International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean. Document prepared by Hui-Hua Lee, Kevin R. Piner, Robert Humphreys, and Jon Brodziak. NOAA NMFS Pacific Islands Fisheries Science Center, 2570 Dole St., Honolulu, HI 96822, USA.

Chen, Y. and D. Fournier. 1999. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Canadian Journal of Fisheries and Aquatic Sciences 56:1525-1533.

Fournier, D.A. 1996. AUTODIFF: a C++ array language extension with automatic differentiation for use in nonlinear modeling and statistics. Otter Research Ltd., Nanaimo, B.C., Canada.

Hart, D. 2001. Individual-based yield-per-recruit analysis, with an application to the Atlantic sea scallop, Placopecten magellanicus. Can. J. Fish. Aquat. Sci. 58: 2351–2358.

Kopf, R.K., Pepperell, J., and Davie, P.S. 2009. A preliminary report on age, growth, and reproductive dynamics of striped marlin (kajikia audax) in the southwest Pacific Ocean. WCPFC-SC5/BI-WP-01, 5th Meeting of the Scientific Committee of the Western and Central Pacific Fisheries Commission (WCPFC-SC5), 10-21 August 2009, Port Vila, Vanuatu.

Marlin Working Group (MAROWG). 2006. Report of the Marlin Working Group Meeting, 15-21 November, 2005, Honolulu, HI, USA. Annex 5. Report of the Sixth Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 23-27 March, 2006, La Jolla, CA, USA.

McDowell, J.R. and Graves, J.E. 2008. Population structure of striped marlin (*Kajikia audax*) in the Pacific Ocean based on analysis of microsatellite and mitochondrial DNA. Can. J. Fish. Aquat. Sci. 65(7): 1307–1320. doi:10.1139/F08-054.

Melo-Barrera, F.N., Félix-Uraga, R., and Quinonez-Velazquez, C. 2003. Growth and length-weight relationship of the striped marlin, *Tetrapturus audax* (Pisces: Istiophoridae), in Cabo San Lucas, Baja California Sur, Mexico. Ceinc. Mar. 29(3): 305-313.

Methot, Jr., R.D. 2011. User manual for stock synthesis. Model version 3.20. January 2011. NOAA Fisheries, Seattle, WA, USA, 143 p.

Methot, R. D., and Taylor, I. G. 2011. Modeling the Variability of Recruitment in Fishery Assessment Models. Canadian Journal of Fisheries and Aquatic Sciences, in press.

Mohn, R. K. 1993. Bootstrap estimates of ADAPT parameters, their projection in risk analysis and their retrospective patterns. In Smith, S. J., Hunt, J. J., and Rivard, D., (eds.), Risk

evaluation and biological reference points for fisheries management, Canadian Special Publication of Fisheries and Aquatic Sciences 120:173:184.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56:473-488.

Pilling, G. M., G. P. Kirkwood, and S. G. Walker. 2002. An improved method for estimating individual growth variability in fish, and the correlation between von Bertalanffy growth parameters. Can. J. Fish. Aquat. Sci. **59**: 424–432.

Punt, A.E. 2010. SSC Default Rebuilding Analysis: Technical specifications and User Manual. Jan. 2010.

Purcell, C.M. and Edmands, S. 2011. Resolving the genetic structure of striped marlin, *Kajikia audax*, in the Pacific Ocean through spatial and temporal sampling of adult and immature fish. Can. J. Fish. Aquat. Sci. 68: 1861–1875. doi:10.1139/F2011-104.

Sippel, T., Holdsworth, J., Dennis, T. and Montgomery, J. 2011. Investigating behavior and population dynamics of striped marlin (*Kajikia audax*) from the Southwest Pacific Ocean with satellite tags. PLoS ONE 6(6): 1-13.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011a. Age and growth of striped marlin (*Kajikia audax*) in waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/09: 15p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011b. Age and growth of striped marlin (*Kajikia audax*) in waters off Taiwan: A revision. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/07: 12p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011c. Reproductive biology of female striped marlin (*Kajikia audax*) in the waters off Taiwan (preliminary). Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/11: 14p.

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011d. Reproductive biology of male striped marlin, *Kajikia audax*, in the waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/09: 17p.

Wang, S.-H., Sun, C.-L., Yeh, S.-Z., Chiang, W.-C., Su, N.-J., Chang, Y.-J., and Liu, C.-H. 2006. Length distributions, weight-length relationships, and sex ratios at lengths for the billfishes in Taiwan waters. Bull. Mar. Sci., 79(3): 865-869.

Appendix 1: Bibliography of materials provided for review

A. STOCK ASSESSMENT REPORT (MAIN REVIEW DOCUMENT)

Billfish Working Group. 2012. Stock Assessment of Striped Marlin in the Western and Central North Pacific Ocean. International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean. Document prepared by Hui-Hua Lee, Kevin R. Piner, Robert Humphreys, and Jon Brodziak. NOAA NMFS Pacific Islands Fisheries Science Center, 2570 Dole St., Honolulu, HI 96822, USA.

B. ISC BILLFISH REPORTS

Billfish Working Group (BILLWG). 2011a. Report of the Billfish Working Group Meeting, 19-27 January, 2011, Honolulu, HI, USA. Annex 7. Report of the Eleventh Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 20-25 July, 2011, San Francisco, CA, USA. Available at: http://isc.ac.affrc.go.jp/pdf/ISC11pdf/Annex_7_ISC11_BILLWG_FINAL.pdf

Billfish Working Group (BILLWG). 2011b. Report of the Billfish Working Group Meeting, 24 May-1 June, 2011, Taipei, Taiwan. Annex 8. Report of the Eleventh Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 20-25 July, 2011, San Francisco, CA, USA. Available at: http://isc.ac.affrc.go.jp/pdf/ISC11pdf/Annex_8_ISC11_BILLWG_FINAL.pdf

Billfish Working Group (BILLWG). 2012a. Report of the Billfish Working Group Meeting, 6-16 December, 2011, Honolulu, HI, USA. Annex 5. Report of the Twelve Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 18-23 July, 2012, Sapporo, Japan. Available at: http://isc.ac.affrc.go.jp/pdf/ISC12pdf/Annex%205%20-%20Report%20of%20the%20BILLWG%20Workshop%20(Dec%202011).pdf

Billfish Working Group (BILLWG). 2012b. Report of the Billfish Working Group Meeting, 2-9 April, 2012, Shanghai, China. Annex 7. Report of the Twelve Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean, Plenary Session. 18-23 July, 2012, Sapporo, Japan. Available at: http://isc.ac.affrc.go.jp/pdf/ISC12pdf/Annex%207%20-%20Report%20of%20the%20BILLWG%20Workshop%20(Apr%202012)%20FINAL.pdf

C. CPUE STANDARDIZATION

Kanaiwa, M., Kimoto, A., and Yokawa, K. 2011. Standardized CPUE of striped marlin in north western central Pacific Ocean by using GLM. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/06: 16p.

Yokawa, K. and Kimoto, A. 2011. Standardized CPUE of striped marlin caught by Japanese coastal large-mesh drift fishery in the northwest Pacific in the periods between 2001 and 2009. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/05: 5p.

Sun, C.-L., Su, N.-J., Yeh, S.-Z., and Chang, Y.-J. 2011d. Standardized catch-rates for striped marlin (Kajikia audax) for Taiwanese distant-water longline fishery in the North Pacific Ocean for 1967-2009. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/07: 45p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP07.pdf

Walsh, W. and Lee, H.-H. 2011. Standardization of striped marlin, Kajikia audax, CPUE with generalized linear models fitted to pelagic longline observer fata from the Hawaii-based fishery: 1995-2009. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/08: 45p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP08.pdf

D. BIOLOGICAL ASSUMPTION

Brodziak, J. and Mangel, M. 2011. Probable values of stock-recruitment steepness for north Pacific striped marlin. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/11: 13p.

Piner, K.R. and Lee, H.-H. 2011a. Meta-analysis of striped marlin natural mortality. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/10: 09p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP10.pdf

Piner, K.R. and Lee, H.-H. 2011b. Correction to Meta-analysis of striped marlin natural mortality. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/08: 01p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_2/ISC11BILLWG2_WP08.pdf

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011a. Reproductive biology of female striped marlin (Kajikia audax) in the waters off Taiwan (preliminary). Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/11: 14p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP11.pdf

Sun, C.-L., Hsu, W.-S., Chang, Y.-J., Yeh, S.-Z., Chiang, W.-C., and Su, N.-J. 2011c. Age and growth of striped marlin (Kajikia audax) in waters off Taiwan. Working paper submitted to the ISC Billfish Working Group Meeting, 19-27 January 2011, Honolulu, Hawaii, USA. ISC/11/BILLWG-1/09: 15p. Available at: http://isc.ac.affrc.go.jp/pdf/BILL/ISC11_BILL_1/ISC11BILLWG1_WP09.pdf

Appendix 2: A copy of the CIE Statement of Work

Attachment A: Statement of Work for Dr. Yong Chen

External Independent Peer Review by the Center for Independent Experts

Stock Assessment of Striped Marlin

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide an impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Striped marlin (*Tetrapturus audax*) is one of six species of billfishes commonly harvested multi-nationally from commercial and recreational fisheries in the western and central Pacific Ocean regions. Fishery management requires high quality science to effectively manage and conserve our living marine resources, and the scientific peer review of stock assessments by external CIE expertise is an important process in the determination of best scientific information available. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have expertise, working knowledge and recent experience in the application of fish stock assessment, mathematical modeling, and statistical computing. Scientists who are employed by or have significant interactions with the Inter-American Tropical Tuna Commission (IATTC) and the Western and Central Pacific Fisheries Commission (WCPFC), and the Secretariat of the Pacific Community (SPC), should not be considered as reviewers. Scientists associated with the ISC also should be excluded as reviewers. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, and other pertinent information. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than **03 December 2012**, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, and to Dr. David Die, CIE Regional Coordinator, via email to ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 October 2012	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
25 October2012	NMFS Project Contact sends the CIE Reviewers the report and background documents
1-16 November 2012	Each reviewer conducts an independent peer review as a desk review
3 December 2012	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
17 December 2012	CIE submits the CIE independent peer review reports to the COR
21 December 2012	The COR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on changes. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Gerald DiNardo, Stock Assessment Program Leader (NMFS Project Contact)
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Gerard.DiNardo@noaa.gov Phone: 808-983-5397

Kevin Piner
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Kevin.Piner@noaa.gov Phone: 858-546-7003

Jon Brodziak
Pacific Islands Fisheries Science Center
2570 Dole Street, Honolulu, HI 96822-2396
Jon.Brodziak@noaa.gov Phone: 808-983-2964

Hui-Hua Lee
University of Hawaii, Joint Institute for Marine and Atmospheric Research
2570 Dole St., Honolulu, HI 96822
Huihua.Lee@noaa.gov Phone: 808-983-5352

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work

Annex 2: Terms of Reference for the Peer Review

Stock Assessment of Striped Marlin

6. Review of the assessment methods: determine if they are reliable, properly applied, and adequate and appropriate for the species, fisheries, and available data.
7. Evaluate the assessment model configuration, assumptions, and input data and parameters (fishery, life history, and spawner recruit relationships): determine if data are properly used, input parameters seem reasonable, models are appropriately configured, assumptions are reasonably satisfied, and primary sources of uncertainty accounted for.
8. Comment on the proposed population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT*); if necessary, recommended values for alternative management benchmarks (or appropriate proxies) and clear statements of stock status.
9. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status.
10. Suggest research priorities to improve our understanding of essential population and fishery dynamics necessary to formulate best management practices.