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

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 **t**he 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.