SCRS/2021/174 Collect. Vol. Sci. Pap. ICCAT, 78(3): 1114-1124 (2021)

REVIEW OF THE 2021 WEST ATLANTIC BLUEFIN TUNA ASSESSMENT

Mark N. Maunder[1]

SUMMARY

The western Atlantic bluefin stock was assessed using VPA and Stock Synthesis. The main changes from previous years were related to the indices of abundance and the use of dome-shape selectivities. Although these changes may have improved the assessments, further work is needed to ensure that indices of abundance are reliable and to determine if an asymptotic selectivity can be used for the GOM fisheries. There were substantial issues with both the VPA and the SS assessments and I considered neither to be appropriate for management advice. These issues may be a direct consequence of the large proportion of the eastern stock seen in the catch data and indices of abundance for young and intermediate aged bluefin in some years, but not in the catch and indices of the spawners. A more qualitative approach to management based on recent trends in biomass as represented by indices of abundance and indications of recent "recruitment" strength from indices and composition data may be more appropriate. These data indicate that the current catch levels or possible higher levels could be appropriate.

RÉSUMÉ

Le stock de thon rouge de l'Atlantique Ouest a été évalué à l'aide de la VPA et de Stock Synthesis. Les principaux changements par rapport aux années précédentes concernaient les indices d'abondance et l'utilisation de sélectivités en forme de dôme. Bien que ces changements aient pu améliorer les évaluations, des travaux supplémentaires sont nécessaires pour s'assurer que les indices d'abondance sont fiables et pour déterminer si une sélectivité asymptotique peut être utilisée pour les pêcheries du GOM. Les évaluations de la VPA et de Stock Synthesis ont posé des problèmes substantiels et aucun des deux modèles n'a été jugé approprié pour l'avis de gestion. Ces problèmes pourraient être une conséquence directe de la grande proportion du stock de l'Est observée dans les données de captures et les indices d'abondance des jeunes thons rouges et des thons rouges d'âge intermédiaire certaines années, mais pas dans les captures ni les indices des reproducteurs. Une approche plus qualitative de la gestion, basée sur les tendances récentes de la biomasse telles que représentées par les indices d'abondance et les indications de la force du "recrutement" récent à partir des indices et des données de composition, pourrait être plus appropriée. Ces données indiquent que les niveaux de capture actuels ou d'éventuels niveaux plus élevés pourraient être appropriés.

RESUMEN

El stock de atún rojo del Atlántico occidental se evaluó utilizando VPA y Stock Synthesis. Los cambios principales respecto a años anteriores estaban relacionados con los índices de abundancia y con el uso de selectividades con forma de cúpula. Aunque estos cambios podrían haber mejorado las evaluaciones, es necesario más trabajo para garantizar que los índices de abundancia son fiables y para determinar si puede utilizarse una selectividad asintótica para las pesquerías del GOM. Hubo bastantes problemas con las evaluaciones tanto de VP como de SS, y considero que ninguna es adecuada para el asesoramiento en materia de ordenación. Estos problemas podrían ser una consecuencia directa de la gran proporción del stock oriental observada en los datos de captura y en los índices de abundancia para el atún rojo joven y de media edad en algunos años, pero no en la captura ni en los índices de los reproductores. Podría ser más adecuado un enfoque más cualitativo de la ordenación basado en tendencias recientes en la biomasa representada por los índices de abundancia y las indicaciones de la reciente

[1] Inter-American Tropical Tuna Commission, mmaunder@iattc.org

fuerza del «reclutamiento» a partir de índices y datos de composición. Estos datos indican podrían ser adecuados los niveles actuales de captura o posibles niveles más elevados.

KEYWORDS

Review, West Atlantic bluefin tuna, Stock assessment

Summary

The western Atlantic bluefin stock was assessed using two approaches, VPA and Stock Synthesis (SS). These approaches have different inherent assumptions (e.g. VPA has time varying selectivity for the combined fleet and Stock-Synthesis has time invariant fleet selectivity, except as represented by the splitting of catch into fisheries and any explicitly modelled temporal variation), but there were also differences in how the data were used in each model. Results were also presented from the ensemble of operating models used for the Management Strategy Evaluation (MSE) process, but the details of these models were not included as part of this review process and could not be evaluated (although they have been reviewed in another fora). Similarly, I did not attend the data prep meetings (although the relevant reports were made available) and therefore my review of any of the data (e.g. how the CPUE based indices of abundance and the associated composition data were standardized) and biological parameters is limited, but it is likely that these components of the model can also be improved.

The main changes in the stock assessments from previous years were related to the indices of abundance and the use of more dome-shape selectivities for the SS assessment. Although these changes may have improved the indices, further work is needed to ensure that they provide reliable indices of abundance. The use of more dome-shape selectivities was appropriate, but further work is needed on the GOM fisheries composition data to determine if an asymptotic selectivity can be used for those fisheries. This is particularly important given the inter-annual variation in the proportion of the stock vulnerable to the GSL fishery (i.e. spatial variation of the stock), which is the only fishery in the SS assessment that is assumed to have asymptotic selectivity.

There were substantial issues with both the VPA and the SS assessments and I considered neither to be appropriate for management advice. These issues may be a direct consequence of the large proportion of the eastern stock seen in the catch data and indices of abundance for young and intermediate aged bluefin in some years, but not in the catch and indices of the spawners. This inconsistency may have manifested itself as retrospective bias, bias in the bootstrap analysis, sensitivity to the US_RR index, and sensitivity to the new data for the VPA and inconsistencies about information on absolute abundance and inability to find a production function using the ASPM in the SS assessment. The issue is likely to worsen as the eastern stock rebuilds. The MSE operating models explicitly deal with the stock mixing, but, as mentioned above, were not reviewed in this process.

The use of the F0.1 management strategy requires estimates of absolute biomass, which are likely biased from both the VPA and SS assessments. Therefore, it is not recommended that these assessments be used to calculate the catch for 2022. Similarly, the projections of the influence of different future catch levels on spawning biomass will be dependent on the estimated absolute abundance and it is not recommended that these projections be used to calculate the catch for 2022. A more qualitative approach based on recent trends in biomass as represented by indices of abundance and indications of recent "recruitment" strength from recruitment indices and composition data may be more appropriate. (note "recruitment" includes both actual recruitment from the western stock and immigrants from the eastern stock) These data indicate that the current catch levels or possible higher levels could be appropriate. However, there is no reliable quantitative information of how large any catch increases could be under the F0.1 harvest strategy.

It should be noted that this was an update assessment and there was limited time for investigatory analyses and limited scope for any changes to the model. Therefore, expectations for model improvements should not be too high. A few improvements were made (e.g. better specification of fishery selectivity and robustification to outliers for the SS assessment), but these could not address the major issues. It is unlikely that the assessments will be reliable enough for providing management advice until either the western and eastern stocks are explicitly modelled together or some alternative approximation is developed that performs well based on simulation analysis that includes operating models that explicitly model both stocks. It is encouraging that close-kin mark-recapture has been initiated for Atlantic bluefin because it is likely to be the only way that the issues with the assessment can be resolved (stock-structure, estimates of absolute abundance, estimates of natural mortality, and estimates of movement).

Introduction

A review of the assessment of the western Atlantic bluefin stock was conducted through attending virtual meetings, interacting with the assessment scientists, and reading related documents. The review is focused on the application of the VPA and Stock Synthesis (SS) assessment models since the review started after the data prep meetings had already bee conducted. However, I did provide some attention to the methods used to create the indices of relative abundance and related composition data since the indices of abundance represented the major changes in the stock assessment.

Given the short time to conduct the assessment and produce the reports, the methods and results were clearly described, and this was adequate to make judgement on whether the results should be used to prove management advice. The stock assessment analysts conducted additional analyses when requested, which helped understand the assessment and provided improvements in some cases. Despite the issues with the assessments, the working group is commended for their work in producing these assessments within the limited time frame and their initiative to diagnose the issues and investigate possible solutions.

Data

Catch

The high catches off Brazil in the 1960s are inconsistent with the rest of the catch history. It is not clear what these catches represent and whether they come from fish spawned in the Eastern Atlantic. These catches have a large influence on the estimated unexploited biomass, the depletion level, and the steepness of the stock-recruitment relationship from the SS model. They are not used in the VPA analysis. A SS model that starts after this period with an estimated depletion level should be considered.

Composition data

There were some concerns about the large fish seen in the Mexican longline fleet, which appeared larger than expected. There were several hypotheses about the origin of these measurements (e.g. weight to length conversations or round length measurements rather than strait length measurements). The use of tail compression may have resolved this issue, but the source of these large measurements should be further investigated.

Consideration should be given to calculating the composition data for the indices (and possibly the fisheries) based on spatio-temporal models and spatial weighting by CPUE (or catch in the case of removals) [see the section on indices below]

Indices of relative abundance

General

The stock assessments include a large number and variety of indices of abundance. Inclusion of an index of abundance does not necessarily improve the stock assessment since the observation model linking the index to the estimated population dynamics may be misspecified. The misspecification might be in the ages represented by the index (i.e. the selectivity), assuming catchability is constant over time when it is not, not representing the whole population, confounding with the Eastern stock, or other factors. It is therefore important that any index that is included in the assessment is representative of the stock. In general, this means that it represents all fish in the population of the ages caught by the gear, as filtered by the selectivity curve, and therefore generally means that it should be based on data for the whole region. Otherwise, nonrandom variation of the proportion of fish in the area represented by the index will cause a bias in the index of relative abundance. Similarly, the composition data, which determine the ages of fish represented by the index, should be calculated appropriately. They should be weighted by the CPUE, not the catch or simply the sampling effort. A further complication is that the indices of pre-spawning sized individuals include Eastern fish, while those of spawning individuals do not, and without having an estimate of the proportions and how they change over time, given the rebuilding of the Eastern stock, it is difficult to reconcile the two types of indices.

The effect of covariates should be categorized by those impacting density and those impacting catchability. This is particularly important when using area weighting. For example, habitat variables (e.g. SST) can be used to predict the density of fish in unsampled or under-sampled areas, while catchability covariates do not alter the underlying density, but impact the predicted catch rates. Density and catchability covariates can be explicitly modelled in a spatio-temporal analysis (particularly if the density covariates are tied to the spatial location) but are confounded in a general linear model. As a thought experiment, take an example where there are only two covariates: year and another covariate X, which has a positive coefficient. In a particular year X is high for all data points. The GLM might estimate that the year effect for that year to be average and the high catch rates are due to X. If the year effect is used as the index of abundance (or the predicted CPUE for each year under the same fixed value for X), then the index value will be average for that year (X is catchability). However, if the average predicted catch rates are used for the index of abundance (which is not usually the case, but mentioned here for illustrative purposes), then the value of the index for that year will be high (a habitat effect). The latter could be interpreted in two ways. First, the habitat was good for all areas and therefore the abundance was high. Alternatively, the fleet targeted the area with good habitat where all the fish were congregated. This is another argument to use spatio-temporal models, but with the caveat that covariates should be categorized as catchability or density and modelled appropriately when creating the index of relative abundance.

Consideration should be given to using spatio-temporal models to combine information from similar fleets and standardize the indices of relative abundance and associated composition data. The composition data for the indices should be spatially weighted by CPUE and the composition data for the fisheries catch should be spatially weighted by the Catch. Spatio-temporal models can be used to fill in unsampled and under-sampled areas. However, caution is needed when large areas have missing data due to expansion or contraction of the fishery.

Many of the indices show high interannual variability that is large when considering plausible variability based on population dynamics, indicating that they are unreliable, particularly for those indices that represent a range of ages. This variability should be taken into consideration when selecting indices to include in the assessment and/or when weighting the indices in the objective function.

USA_RR_66-144

In previous assessments the data for the US rod and reel fishery was separated into three different indices. This separation was due to management measures being applied differently among the categories. There have been numerous regulations starting in 1992 including minimum legal size, trip bag limits, and seasonal closures, which may bias the index of abundance. In addition, changes in targeting of the fleet towards larger fish may have caused the indices to be biased. It was therefore considered an improvement to combine the school (66-114 cm) and large school (115-144 cm) categories together to produce a single index, which also may be robust to joint-size-class bag limit regulations. There has also been a shift in the spatial distribution of the fish.

The CPUE analysis does not include any spatial covariates. It is argued that the distribution of fish is driven by the spatial distribution of prey abundance and there is no measure of prey abundance to use as a covariate in the model. Therefore, due to the aggregating behavior of prey, it is not appropriate to use spatial covariates in the model. This may not be the most appropriate decision. As a thought experiment, compare an extreme example where in one year all the prey are congregated in a single area and another year where all the prey are evenly distributed. If bluefin congregate where the prey is, then the catch rates are likely to be high in the first year, due to fisheries targeting the aggregation, and low in the second year even if the abundance of bluefin is the same in both years. A spatio-temporal model should be considered to analyses these CPUE data and the associated composition data. The bluefin working group investigated using a spatio-temporal model, which produced an index that was similar to the one used in the assessment. However, further investigation should be conducted.

SST is used as a covariate in the model, and it is not clear whether SST should be a density or a capability covariate and whether it is modelled correctly. The working group did some investigation into this issue, but more work is probably warranted.

Despite the data being restricted to those trips primary targeting small (66 to 144 cm) bluefin, the shift to targeting larger fish, the size-class based regulations, and changes in spatial distribution of the fish is likely to have changed the selectivity of the index. Combing the school (66-114 cm) and large school (115-144 cm) categories has probably exacerbated this issue. There is also evidence that there was cohort targeting of the large cohort that entered the fishery in 2005. In contrast, targeting could change from the strong cohort to other sized fish due to quickly reaching the bag limit of sizes represented by the strong cohort. These changes in selectivity need to be addressed in the standardization method and the associated composition data or time varying selectivity for the index modelled explicitly in the assessment, that latter was investigated by the working group and results are similar to removing the index.

GOM longline index

In the current assessment, the Mexican data was added to the US data to create the GOM longline based index of relative abundance. Combining the data was to overcome the reduction in the catch by the US fishery and was considered appropriate due to similarities in the target species, gear configurations, and the type of information collected by observers. The flag effect was either non-significant or produced similar results even though the CPUE trends for each fleet were quite different. An important component of the analyses was the use of area-month interactions to account for the spatial differences in the two fleets and the seasonal migration pattern of bluefin.

The US fleet has had significant changes over time including time-area closures, changes from J to circle hooks, and the introduction of weak circle hooks. Hook type was included as a covariate in the analysis and the area-month factor address the time-area closures. The use of the weak circle hook reduced catch rates by about 50%.

The analysts should be commended for their efforts to combine the two fleets and account for any differences among them to produce an improved index of abundance. However, there are two major improvements that they should consider for future assessments. The first is to area weight the index of abundance rather than data weight and the second is to also standardize the composition data associated with the index.

It is not clear exactly how the index has been created, but it is likely that each data point in the analysis has the same (or similar) influence on the estimate of the year effect. Therefore, low sampling in areas of high abundance (which may be the case with a non-target species) could cause bias even if an area effect is included in the analysis since the year effect will be more influenced by the area with most samples. A more appropriate approach may be to model the abundance in each area, then sum the abundance over areas to create the index. This can be accomplished using a spatio-temporal model. However, since bluefin exhibit a seasonal migration pattern, simply adding the abundance in each area would not be appropriate and the analysis would have to be modified to take this into account.

SST is used as a covariate in the model, and it is not clear whether SST should be a density or a capability covariate and whether it is modelled correctly.

There is some suggestion from the length composition data that the US and Mexican fleets catch somewhat different sized fish, but variation in sample sizes among different time periods makes the comparison a little complicated. This is particularly important for fleets that catch the largest fish and may be assumed to have asymptotic selectivities, which can have a large influence of stock assessment estimates of biomass and fishing mortality. Including the composition data separately for the US and Mexican fleets as additional fisheries in the Stock Synthesis assessment, but not included in the objective function, would be a convenient way to provide evidence that the fleets have the same or different selectivity. The two fleets fish in different areas and it might be expected that the implied selectivities differ due to availability, particularly if there is differences in seasonal migration by size. In addition, given the weak circle hook reduces catch rates by 50%, it could be possible that the reduction in catch rate may differ with size creating a change in selectivity for the US fleet. In addition, the composition data has a recent large cohort that gets larger than seen in previous years. The composition data should be standardized along with the catch rate data and area weighted. The area weighting should be based on predicted CPUE (density) for the composition data used for the index and by catch for the composition data used for the removals. The change in selectivity of the US fleet needs to be taken into consideration in the spatio-temporal model (i.e. the different fleets and time periods need to have a selectivity curve estimated in the spatio-temporal model) and the US and Mexican fisheries should be modelled separately with the US fleet having time blocks in selectivity. The selectivity for the index and for the fishing fleets should not be shared. The fleet specific selectivity could be modelled in a spatio-temporal model (e.g. VAST) using a fleet*length/age interaction term for catchability.

Canadian and US handline indices

Three separate indices are used for the Canadian and US_RR > 177 cm fisheries. The single Canadian index used in the previous assessment was separated into two indices (SWNS and GSL) for this assessment. This aligns the data from the GSL fishery with the acoustic index (see below) and was considered an improvement. However, there appears to be temporal variation in the distribution of the fish among these fisheries, particularly between the Canadian and US fisheries, and therefore the SS model incorporates an environmental index to model this temporal variation in the catchability parameter. The VPA analysis does not use these indices to tune the terminal fishing mortalities. Indices should, if possible, represent the whole stock (filtered by the selectivity) not only a part of the stock defined on spatial criteria, particularly if there is temporal changes in the spatial distribution. A more appropriate approach, which is consistent with the general statement above, may be to combine the Canadian and US fisheries in a single spatio-temporal model to produce a single index. Differences in selectivity and catchability among the two fisheries may have to be considered, which may be estimable if there is spatial overlap in part of the fishery. The working group attempted a spatio-temporal model, but it was rejected partially due to a lack of spatial overlap. Further investigations are warranted.

There has been a change in the spatial distribution of the Canadian fleet to areas with smaller fish and a ITQ introduced. (see above for USA regulations). A change in selectivity was modelled for the Canadian fishery. The GSL fishery is influential in the SS analysis because it is the only fishery with asymptotic selectivity. Therefore, this fishery should be analyzed in detail. Further work should also be applied to the GOM fisheries (see above) to determine if their selectivities could be made asymptotic in addition or in place of the GSL fishery.

The use of an environmental index for catchability makes sense and reduces the residual pattern seen in the VPA analysis. However, combining the data from all three fisheries using a spatial-temporal analysis would probably be more appropriate.

Acoustic GSL
The acoustic index has similar cyclical patterns as the Canadian handline fisheries and therefore is likely to include the same issues of changes in the spatial distribution of the stock.

Biological assumptions

Natural mortality

It is stated that natural mortality is based on the using Hoenig's method using a maximum age of 35, which infers a value of 0.1 at age 20 and that is scaled for other ages using the Lorenzen curve. Although, from a quick calculation based on the maximum age of 35 using the available Amax formulars, the approach may have been different than described above. Nevertheless, natural mortality is a difficult parameter to estimate, there are multiple Amax based formulas, the choice of the Amax used in the formulars, and how natural mortality changes with age. Since natural mortality is influential on both stock assessment estimates of abundance and on reference points and other management quantities, further work should made to fine tune the estimates of natural mortality (see the recent CAPAM workshop on natural mortality and the associated special issue in Fisheries Research).

Stock-recruitment relationship

The "recruitment" estimated by the stock assessment model is a combination of both recruits from the western stock and immigrants from the eastern stock. Therefore, the inclusion of a stock-recruitment relationship in the assessment is questionable (it might be useful for estimating a trend in recruitment, but should not be interpreted as a stock-recruitment relationship) and use of the estimated stock-recruitment relationship for management is inappropriate.

Age at maturity

There is a large difference between the two maturity schedules. I did not look at how they were derived and what they represent, but one would expect that the age-at-maturity should be better determined. Age-at-maturity influences the stock-assessment through the stock-recruitment relationship and management through the definition of the spawning biomass and how the sock-recruitment relationship is used in biomass-based reference points (e.g B_{MSY}). Given that the management of the western Atlantic bluefin stock is biased on F0.1, the maturity schedule is essentially irrelevant unless it influences the estimates of current biomass. However, if management is to be based on spawning stock biomass (e.g. the objective to increase spawning biomass by X%) or the status of the stock is used by certification organizations (e.g. S/S_{MSY}), then more work needs to be done to ensure the correct age-at-maturity is used.

Growth

The specification and estimation of the growth equation is often more influential that generally considered, particularly when used in contemporary integrated stock assessment models. The asymptotic length can have a large influence on the estimates of absolute abundance when fit to length composition data, particularly from fisheries that assume an asymptotic selectivity. Even the variation of length at age for old individuals can have an influence on estimates of absolute biomass.

The Richards growth curve estimated externally is used for cohort slicing the length composition data used in the VPA and is estimated internally within the SS assessment using age-conditioned on length data. The external and internal estimates were similar. I did not look at details of the data or the estimations, but if not already done, the sensitivity of results to the internal and external estimates, alternative growth curves, parameterization of the variation of length-at-age, data sets used, and differences between the west and eastern stocks should be considered.

Assessments

The bluefin stock was assessed using two approaches, VPA and Stock Synthesis (SS). These approaches have different inherent assumptions (e.g. VPA has time varying selectivity for the combined fleets and Stock-Synthesis has time invariant fleet selectivity, except as represented by the splitting of catch into fisheries and any explicitly modelled temporal variation), but there were also differences in how the data were used in each model.

Model component	VPA	SS
Years	1974-2020	1950-2020
Plus group	16	35
Composition data	Catch-at-age	Catch at length, Age conditioned on length
Index weighting	Fishery independent: CV = 0.3 Fishery: estimate additive constant	CV = 0.2
Indices excluded	Canadian SWNS and GSL, US RR > 177cm	
Stock-recruitment relationship	None	Beverton-Holt with estimated steepnes
Japan_LL time varying selectivity	2011-2015	2010-2014
Asymptotic selectivity	Age 16+ = age 15	Canadian GSL (age 14), but overall F at age decreases after age 14

The assessments were categorized as updates, but there were some significant changes compared to the previous assessments. The USA rod and reel fishery based index was previously separated into 3 categories, but in this assessment, categories 66 to 114cm and 115 to 144cm were combined into a single index. The data from the Mexican longline fishery in the Gulf of Mexico was added to the USA fishery to produce a single index of relative abundance for spawners. There is indication that these changes may not be appropriate (sensitivity of the VPA to the US_RR index and possible differences in selectivity between US and Mexico GOM LL). The Canadian handline fishery was split into the CAN_HL_SWNS and CAN_HL_GSL. The SS assessment changed the selectivities for some fisheries from asymptotic to dome shape.

A priori, the stock assessments of bluefin tuna in the western Atlantic are likely to be biased due to the migration of eastern stock fish into the western area. Indices of abundance for juveniles are indexing both western and eastern fish, while the indices based on spawning fish are indexing only western fish. Similarly, catch-at-age of juveniles comprises both western and eastern fish, while that for spawners only comprises western fish (although some fleets catch both spawners and large fish that are not spawning). When the “stock” being assessed is comprised of a high proportion of eastern fish, which it is in some years (e.g. 2016-2018), indices of juveniles and spawners will be incompatible. In addition, use of a stock-recruitment relationship is inappropriate.

VPA

The VPA assessment is unreliable as indicated by the severe sensitivity to the inclusion of two years of new data, severe sensitivity to the few fish caught at age one, poor fit to the indices of abundance, severe residual pattern in the fit to some indices, possible bias in the estimated spawning biomass indicated by the bootstraps, severe retrospective pattern, and large sensitivity to the inclusion of some indices of abundance.

An analysis looking at a stepwise inclusion of changes from the previous assessment showed that inclusion of the new two years of data and the combining of the USA rod and reel fishery based index categories 66 to 114cm and 115 to 144cm to form a single index, had the largest influence on the results. Cutting the last two years of catch-at-age data in half had a large impact on the estimates of recruitment and spawning biomass verifying the impact of the new catch-at-age data. This appears mainly due to fish aged 3 and older. In 2019, the age 2 catch was low (49 fish) compared to most years (an average of 392 fish for 2010-2018), but the catch of the same cohort in the next year (2020) was very high at age 3. This suggests that either selectivity changed or growth changed (i.e. the cohort slicing is off in some years). Sensitivity to inclusion of the recruitment index was not impacted by the size of the terminal vulnerability penalty. The sensitivity of the new data and the combining of the US-RR index may have to do with the calculation of the selectivity used for this index and the change in selectivity over time. The implied selectivity of the combination of all fisheries over time that removes catch from the population is inherently time varying, but the selectivity's used for the indices of abundance are constant over time. This time invariant assumption for the indices may be violated and cause the bias, and may be more influential when combining the two US-RR indices.

The fit to all abundance indices is poor, which is expected for some indices that have high interannual variability. The fit is notably bad for US_RR_66_144, GOM larval, and Japan LL early, which have severe residual patterns. The GSL acoustic index also has a severe residual pattern. The US > 177 and Canada HL (SWNS and GSL) are not used in the assessment, but they show opposite cyclic trends (likely driven by oceanographic forced changes in spatial distribution) and patterns in residuals when fit in the model. The GSL acoustic index shows similar cyclic patterns to the Canadian HL indices making that index questionable as well. Jack-knife analysis of the indices (leave out one index at a time) shows estimated recruitment since about 2005 is highly sensitive to the inclusion of the US_RR_66_144 index. The estimated recent spawning biomass is also highly sensitive to inclusion of this index, as well as sensitive to the inclusion of the GOM larval index and the GSL acoustic index.

All the diagnostics indicate issues with the stock assessment and many, if not all, were not present in the previous assessment. There are severe retrospective patterns in the estimates of both recruitment and spawning biomass, which were not seen n the previous assessment. There is also possible bias in the spawning biomass estimates indicated by the bootstrap analysis. Changing the catch for age one to one fish for a number of recent years had a large impact on the results even though the maximum was 46 fish in any of those years. Interestingly, this was not a problem using last year's assessment indicating it could be due to combining of the US_RR indices (the analysis should be repeated with RUN0 to check).

There were no obvious ways to address the issues found with the VPA assessment. Further investigation is needed. However, given that the catch has a large proportion of eastern fish in some years, the VPA analysis needs to be modified (e.g. inclusion of movement/mortality estimated before the individuals mature and/or exclusion of the spawning biomass indices from the model fit) to address this issue before it is used for management.

Stock Synthesis

The Stock Synthesis (SS) assessment is unreliable as indicated by the conflict among data sources shown by the R0 likelihood component profile and catch-curve analysis, poor fit to the indices of abundance, and inability to find a production function in the ASPM diagnostic.

The fits to the indices of abundance are generally poor, except for the indices that have time blocks in selectivity, time varying selectivity, or the catchability is related to an environmental index. Of particular concern is the poor fit to the indices that represent spawning individuals (USMEXLL_GOM, GOMlarval, CAN_Acoustic). The assumed CV = 0.2 is not consistent with the fits to the indices. An additive variance component as applied in the VPA assessment, should be estimated for each index's likelihood. For estimation stability, the CV could be set to 0.2 (or 0.3 as done in the VPA) for the index that is assumed to be most reliable.

The R0 likelihood component profile shows conflicting information between the indices of abundance and the length frequency data and between the length composition and the age conditioned on length composition data. Within each of these data types there are conflicts among the different fisheries and indices. These conflicts have been reduced compared to the previous assessment due to the use of more dome shape selectivities, but they are still severe.

The aggregated fits to the length composition data are reasonable. Therefore, fish are removed at about the right size, but there are large residuals for large and small fish in some years for some fisheries, and temporal patterns in residuals. These may influence the parameter estimates, particularly for some years.

The change from logistic to dome shape selectivity for all but one of the fleets is more appropriate than previously assumed as indicated by the empirical selectivity diagnostic. The likelihood component profiles and the jitter analyses were improved. However, further work is needed to evaluate if any other fleets or indices should have asymptotic selectivity. This includes ensuring that the composition data are calculated correctly (spatial weighting by catch for fisheries and by CPUE for indices).

Other diagnostics also show some issues with the assessment. Removing the indices one at a time showed spawning biomass estimates having moderate sensitivity to the GOM_larval, USRR > 177, JPN_LL, and CAN_GSL indices. The retrospective analysis shows little pattern, but recent data has a large influenced on the estimates of high recent recruitment (as would be expected). The bootstrap analysis indicates that the estimates of recent recruitment and the standard deviation of the recruitment deviate penalty may be biased. The catch curve analysis shows that the composition data suggest an increasing biomass trend since the mid 1980s while the ASPM and the full integrated analysis do not. The composition data is weighted using iterative reweighting based on the harmonic mean, which tends to over weigh the composition data and alternative weighting should be considered.

The ASPM diagnostic suggests that there is an inconsistency between the recruitment index (USRR) and the spawning biomass index (USMEXLL_GOM) and that the recruitment index is more consistent with the Japanese longline index. This is consistent with the understanding of the stock dynamics where the recruitment index and the Japanese longline index comprise fish from both the western stock and the migrants from the eastern stock, while the spawning biomass index only comprises fish from the western stock.

The estimation of the stock recruitment relationship steepness in the stock assessment is not appropriate. The large cohorts of fish caught off Brazil in the early part of the time series, which may have been from the eastern stock, are informing the steepness parameter. In addition, estimating recruitment deviates back to the start of the model time period greatly reduces the estimate of the unexploited biomass. The recruits to the fishery are from both the western and eastern stocks invalidating the stock-recruitment relationship concept. However, given the rebuilding of the eastern stock and the larger number of eastern fish migrating to the west, simply modelling recruitment as randomly varying around a constant value may also not be appropriate. A recruitment index related to the abundance of the eastern stock (or the appropriate ages) may be worth investigating. Starting the model from an exploited condition after the large catches off Brazil occurred might also be worth considering.

The US_RR fishery is split into multiple fisheries based on length classes. This causes an artificial break in the selectivity curve that may not be well represented by the double normal selectivity curve. The addition of retention curves should be considered to better model this selectivity.

The model is unable to fit the recent peak in small fish in the length composition data. This means that the model is not interpreting the data correctly and will cause a bias in the estimate of this cohort's size. The estimates of recent recruitment are also sensitive to the value of the standard deviation of the recruitment deviate penalty. Therefore, the estimates of recent recruitment, though high, are unreliable. The estimated uncertainty in the estimated recruitment does not represent the error in the estimate of recent recruitment and caution should be applied when using the estimates of recruitment for projections. The bias may be caused by different growth rates or cohort targeting (temporal changes in selectivity).

It might be useful to include USA and Mex GOM indices and composition data as separate additional fisheries in the assessment without explicitly fitting to them, so that any differences can be observed to evaluate the appropriateness of combining them. The USA GOM fishery has a change in selectivity due to management regulations (e.g. weak hooks to allow large spawners to escape) and this should be taken into consideration.

Butterworth and Rademeyer

The ensemble of operating models (OMs) developed for the MSE were presented as an alternative way to evaluate management advice. These models differ from the VPA and SS models because they explicitly model both the western and eastern stocks. They are conditioned on the data and the results are essentially a modelled averaged stock assessment used to project the stock under a constant catch policy. The models have been previously reviewed by the bluefin working group, but their details were not presented at this meeting and I therefore cannot comment on their validity. However, unless explicit modeling of the western and eastern stocks solves all the issues encountered for the VPA, SS, and data inputs, using the OMs for providing catch limit advice should be done with caution.

Management quantities and projections

Management is based on a F0.1 strategy and therefore requires an estimate of absolute biomass (since catch is F0.1 applied to the current or projected biomass) or, at a minimum, a reliable estimate of current fishing mortality relative to F0.1 to determine if current catches are appropriate. Nether the VPA or SS assessments provide reliable estimates of these quantities. In addition, recent estimates of recruitment are unreliable, which are used for projections under F0.1 to calculate appropriate levels of catch. Nevertheless, there is indication from the indices that the biomass is increasing under current catch levels and recent recruitments (recruitment plus immigration from the eastern stock) are high. Therefore, increased catch is probably sustainable, but the amount of increase is uncertain. Estimates of stock status based on fishing mortality or biomass reference points are also unreliable. The presence of eastern stock fish in the catch and indices of abundance make the estimates of the steepness of the stock-recruitment relationship invalid and any reference points based on the stock-recruitment relationship (e.g. B_{MSY} and F_{MSY}) are not appropriate.

The estimates of current biomass from the SS assessment are higher than the previous assessment due to changing from asymptotic selectivity to dome shape selectivity for many of the fisheries. This translates into a lower estimated fishing mortality and higher projected catch under F0.1. However, due to the unreliability of the SS assessment indicated by the conflict in the data, the estimated increased catch is not reliable.

The estimates of recent recruitments from both the VPA and SS assessments are high, but unreliable due to the sensitivity of the VPA to the new data and the US_RR index and the misfit of the SS model to the length composition data. Therefore, any predictions using these recruitments will also be unreliable (for quantities related to the recruitment level, which will depend on the ages used in the calculation and the duration of the projects), particularly catch levels over the next few years based on F0.1 for fisheries that catch young fish. These unreliable recruitments could be replaced with average recruitment, but this ignores the information indicating recruitment is high and will depend on the years over which the average is taken. The most recent recruitment that is estimated has high confidence intervals and therefore it may be appropriate to replace this with average recruitment. Using average recruitment for these recent years could be viewed as precautionary.

Consideration should be given to the allocation of the catch among fisheries as a strong cohort enters the fishery. This cohort will increase the estimates of catch based on F0.1 projections, but the increased catch is due to young fish and should not be allocated to catch of older fish (e.g. spawners). It may be useful to look at the fishing mortality at age when projecting under different catch levels using the current proportional allocation of catch among fisheries.

Projections based on the MSE OMs predict, on average, an increase in abundance under the current quota, but some scenarios related to the lower recruitment scenario predict biomass will decline. I don’t think these models use the most recent data and may not have information on the recent large recruitments. This is another indication that the current catch levels are sustainable and there is a possibility for increased catch.

The projections presented for the VPA and SS models were conducted using appropriate methods, but due to issues with the assessments, they should not be used for management advice. Any measures of uncertainty including the asymptotic based confidence intervals and the bootstraps are likely to be biased because of model misspecification and/or inappropriate data weighting and should not be used for probability based harvest control rules.

Conclusions

The VPA and Stock Synthesis assessments are not reliable enough for providing management advice. The indices of abundance and composition data suggest that the current catch levels are sustainable and increases in catch may be possible. However, the amount of increase in catch that would be sustainable is uncertain. Future improvements in the stock assessment should focus on integrating the western and eastern assessments and considering spatial-temporal analysis of indices of relative abundance and related composition data. It is encouraging that close-kin mark-recapture has been initiated for Atlantic bluefin because it is likely to be the only way that the issues with the assessment can be resolved (stock-structure, estimates of absolute abundance, estimates of natural mortality, and estimates of movement).

Acknowledgements

The working group is thanked for their cooperation in preparing this review. The views presented in this review are those of the author and do not reflect the views of the IATTC.