SCRS/2018/148 Collect. Vol. Sci. Pap. ICCAT, 75(7): 2247-2258 (2019)

EXTERNAL REVIEW OF ICCAT BIGEYE TUNA STOCK ASSESSMENT

Carmen Fernández[1]

SUMMARY

This document provides a review of the bigeye stock assessment conducted by ICCAT in 2018. The work done for this assessment is of high quality, both in terms of data aspects as well as methods for assessment and provision of management advice. Very considerable efforts appear to have been made over the years to improve the quality of the data available for the stock assessment; this is very important given the high complexity of the fisheries. Surplus production models and the age-structured SS3 model (fitted to length but no age data) were implemented. Of these, the SS3 implementation is considered to be the most suitable for the assessment of this stock and to provide management advice, mostly because of the very different selectivities of some of the main fleets which has led to changes in the overall selectivity for the total fishery on the stock over time. All methods applied reached the conclusion that there is presently a high probability of B<Bmsy and F>Fmsy, corresponding to the red quadrant of the Kobe plot. Efforts to continue to develop and improve the longline CPUE index used in the stock assessment are strongly encouraged, as is the development of an abundance index for juvenile bigeye tuna. Future work considering ways to more fully incorporate all relevant sources of uncertainty in the projections and evaluating harvest strategies taking due account of uncertainty via MSE is of high relevance and seems to be planned already.

RÉSUMÉ

Le présent document passe en revue l'évaluation du stock de thon obèse réalisée par l'ICCAT en 2018. Le travail effectué pour cette évaluation est de grande qualité, tant du point de vue des données que des méthodes d'évaluation et de formulation d'avis de gestion. Des efforts considérables semblent avoir été déployés au fil des ans pour améliorer la qualité des données disponibles pour l'évaluation des stocks, ce qui est très important compte tenu de la grande complexité des pêcheries. Des modèles de production excédentaire et le modèle SS3 structuré par âge (ajusté aux données de taille mais pas aux données d'âge) ont été employés. Parmi ceux-ci, l'exécution du SS3 est considérée comme la plus appropriée pour l'évaluation de ce stock et pour la formulation d'avis de gestion, principalement en raison des sélectivités très différentes de certaines des principales flottilles, ce qui a entraîné des modifications de la sélectivité globale pour la pêcherie totale sur le stock au fil du temps. Toutes les méthodes appliquées ont abouti à la conclusion qu'il existe actuellement une probabilité élevée que B<BPME et que F> FPME, ce qui s'inscrit dans le quadrant rouge du diagramme de Kobe. Les efforts visant à poursuivre le développement et l'amélioration de l'indice de CPUE palangrière utilisé dans l'évaluation du stock sont fortement encouragés, de même que l'élaboration d'un indice d'abondance pour le thon obèse juvénile. Les futurs travaux sur les moyens d'intégrer plus complètement toutes les sources d'incertitude pertinentes dans les projections et d'évaluer les stratégies de capture en tenant dûment compte de l'incertitude liée à la MSE revêtent une grande importance et semblent déjà être prévus.

[1] Instituto Español de Oceanografía, Avda. Príncipe de Asturias 70 bis, 33212 Gijón, Spain.

RESUMEN

Este documento presenta una revisión de la evaluación del stock de patudo realizada por ICCAT en 2018. El trabajo realizado para esta evaluación es de gran calidad, tanto en términos de los aspectos relacionados con los datos como de los relacionados con los métodos de evaluación y la formulación del asesoramiento de ordenación. Parecen haberse realizado considerables esfuerzos a lo largo de los años para mejorar la calidad de los datos disponibles para la evaluación de stock, algo muy importante dada la gran complejidad de las pesquerías. Se implementaron modelos de producción excedente y el modelo estructurado por edad SS3 (ajustado a los datos de talla pero no de edad). De ellos, la implementación del SS3 se considera la más adecuada para la evaluación de este stock y para proporcionar asesoramiento de ordenación, debido principalmente a las muy diferentes selectividades de algunas de las principales flotas que han conducido a cambios en la selectividad global para la pesquería total de este stock a lo largo del tiempo. Todos los métodos aplicados llegaron a la conclusión de que actualmente existe una elevada probabilidad de $B<B_{RMS}$ *y* $F>F_{RMS}$*, correspondiente al cuadrante rojo del diagrama de Kobe. Se insta encarecidamente a que continúen los esfuerzos para desarrollar y mejorar el índice de CPUE de palangre utilizado en la evaluación del stock, al igual que a continúe el desarrollo de un índice de abundancia para el patudo juvenil. Es de gran importancia que el trabajo futuro considere formas de incorporar de forma más completa todas las fuentes de incertidumbre pertinentes en las proyecciones y de evaluar las estrategias de captura teniendo debidamente en cuenta la incertidumbre mediante la MSE, algo que parece ya planificado.*

KEYWORDS

Bigeye tuna (BET), Thunnus obesus*, Atlantic, Stock assessment, Review*

1. Terms of reference

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock and if appropriate recommend alternative approaches to be accomplished in the future.
3. Evaluate the methods used to estimate population benchmarks and stock status (e.g., MSY, FMSY, BMSY or their proxies).
4. Evaluate the adequacy, appropriateness, and application of the methods used to evaluate future population status, given the commissions objectives.
5. Evaluate the adequacy, appropriateness, and application of methods used to characterize the uncertainty in estimated parameters. Comment on whether the implications of uncertainty in technical conclusions are clearly stated.
6. Comment on whether the stock assessment results are clearly and accurately presented in the detailed report of the Stock Assessment.
7. Comment on potential improvements on the stock assessment SCRS process (CPC participation, transparency, objectivity, documentation, uncertainty characterization, etc.). as applied to the reviewed assessments.
8. Comment on the adequacy of the workplan for the assessment and whether it was adequately addressed by the Data or Assessment Working groups.
9. Consider the research recommendations provided by the working group and suggest any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment considering control rules or management strategy in effect.

2. General findings

I find the work done for this bigeye tuna (BET) assessment to be of high quality, both in terms of data aspects as well as methods for assessment and provision of management advice. I support the work done and I agree with the conclusions reached and my comments should be interpreted as aiming to provide some additional thoughts that could help when developing further aspects of this assessment in the future.

It is clear from the Data Preparatory (hereafter DP) and Stock Assessment (hereafter SA) reports that very considerable efforts have been made over the years by national scientists and the ICCAT Secretariat to improve the quality of the data available for the stock assessment. This is very important given the high complexity of these fisheries (wide-ranging stock with fisheries from different parties often catching a mix of tuna species) and I encourage the continuation of these efforts in coming years. In addition, any improvements that could occur in the knowledge of biological parameters (growth, M, length-weight relationships) would be very useful for the assessment.

Three types of models were used for the stock assessment this year: a surplus production model “mpb” without process error, a surplus production model “JABBA” including process error, and an age-structured “SS3” model fitted to length data (no age data used in the fit). All three alternatives were implemented in a thoughtful and competent manner, considering an appropriate suite of diagnostics.

In line with the conclusions from the SA report, I consider the SS3 implementation to be the most suitable for the assessment of this stock and to provide management advice. This is mostly because it can take into account the very different selectivity of longline (LL), baitboat (BB) and purse seine (PS) fleets, and the increasing proportion of catches from FADs in recent years, which results in changes in the selectivity of the total combined fishery on the stock.

A key development this year is the joint combined index from CPUE of LL fisheries, derived by combining operational data from several main LL fisheries. The index has a wider spatial and temporal coverage of the stock than any of the separate indices previously used and represents, in principle, a very useful advancement for the stock assessment. At the same time, it is crucial to ensure that the index is reliable and I strongly encourage efforts to continue to develop and improve this index in coming years, focusing among other things in demonstrating the ability of the clustering method to differentiate between changes in targeting strategy and changes in relative abundance of the different species and further considering how best to represent the selectivity of the index. I also strongly recommend the development of an abundance index for juvenile BET.

All methods applied reached the conclusion that there is presently a high probability of $B<B_{MSY}$ and $F>F_{MSY}$, corresponding to the red quadrant of the Kobe plot.

Projections of future population development under different future levels of constant catch were performed, considering a range of model configurations. I support the further development of the SS3 projections by considering ways to more fully incorporate all relevant sources of uncertainty. I also recommend evaluating harvest strategies taking due account of uncertainty via MSE (as I understand will be done for this stock in the near future).

3. Tor 1: Data

3.1 Fisheries data

It is clear from the reports that very considerable efforts have been made over the years by national scientists and the ICCAT Secretariat to improve the quality of the data available for the stock assessment. This is particularly the case for fisheries-dependent data (catch, effort, spatio-temporal distribution of catch, length frequencies and catch-at-size matrices). These efforts are both commendable and highly necessary given the high complexity of these fisheries (wide-ranging stock with fisheries from many different parties and often catching a mix of tuna species) and I encourage their continuation in coming years.

The overall fishery selectivity on the stock appears to have changed substantially over time, with the strong rise of PS catches on FADs in the last two decades or so (see e.g. Figure 5 of the SA report). The difference in the sizes of BET caught by the different fisheries is striking (e.g. Figure 7 of the SA report), with the BB and PS fisheries catching much smaller sizes of BET (mostly juveniles) than the LL fisheries. This makes it very important to have good-quality data on the PS fisheries and I think efforts in that respect are particularly relevant.

In respect to the EU PS fleet, the DP report highlights that the length-weight relationships used in the T3 software to apportion the combined catch of yellowfin, skipjack and BET to species are old (from 1982 in the case of BET) and correspond to a time when the fisheries were quite different from nowadays. Therefore, the length-weight relationships should be reviewed and modified, if needed, in order to improve the accuracy of the catches for the 3 tropical tuna species. The DP report also notes that using smaller spatio-temporal strata in the T3 software for the said fleet would improve the estimates of species composition.

The DP report also identifies the need to more generally collect length and weight measures (including different metrics) in order to review and modify, if needed, the length-length, weight-weight and length-weight relationships currently used by ICCAT.

The SS3 assessment splits the catches into 15 different fleets, which in broad terms are defined based on gear type (PS, separating between PS on free schools and FADs, BB and LL), geographical area (tropical area, which is where most catches take place, north area and south area; i.e. 3 areas in total), sometimes country (e.g. LL Japan and LL Others) and sometimes time period (e.g. differencing between earlier and later years in PS or BB fisheries). I expect the general idea was to end up with a set of "fleets" each of which could be assumed to have a roughly constant selectivity over time (though I note that some block changes in selectivity are included in the SS3 assessment for some fleets). This partitioning into fleets seems appropriate and has focussed the work of the group on obtaining total catches and length-frequency distributions, by quarter, for each of them. A number of decisions had to be made as to the fleet to which certain catches or length samples should most sensibly be allocated; this is unavoidable given the complexity of the data and, even though I do not know the full details of how it has been implemented, I have no reason to question the decisions taken in this regard.

3.2 Stock abundance indices from standardised fishery CPUE:

A key development in the stock assessment this year is the joint combined index from LL fisheries CPUE, derived by combining operational (set-level) data from main LL fisheries (Japan, US, Korea, Chinese Taipei). The combined index is used as the sole abundance index in this year's stock assessment; this is in contrast to previous assessments, which used separate LL indices corresponding to the different national fleets. Using the joint combined index is, in principle, a useful advancement for the stock assessment, as the index has a wider spatial and temporal coverage of the stock than any of the separate indices; moreover, it provides a common agreed methodological approach for the standardisation of CPUEs from different fleets, which I expect will increase the transparency and replicability of the work. This said, my understanding is that this index is a main driver of the stock status results and, therefore, it is crucial to ensure that the index is reliable and represents a step in the correct direction for the stock assessment.

The LL fisheries catch multiple tuna species, adding substantial difficulty to the CPUE standardisation process. In the derivation of the joint index, a clustering method based on species composition (by vessel-month) was applied to better characterize effort directed to BET. The main question for me is whether the clustering method has been able to differentiate between changes in targeting strategy and changes in relative abundance of the different species. For example, if during a series of years the proportion of BET in the catches is higher than observed in previous years, is this because in the said series of years there has been more targeting of BET or because the abundance of BET has increased relative to the abundance of other species? Have there been simulation studies to validate the performance of the clustering approach in this respect?

I think the results from the clustering analysis should be complemented with knowledge of the fisheries in each country. I think one should try to understand whether the clusters obtained can be interpreted meaningfully given what is known about the fisheries, e.g. in terms of combination of species in each cluster, distribution of variables such as gear characteristics, geographical location or quarter of the year in the different clusters; examine also how the proportion of vessel-months allocated to the different clusters has evolved over the years and, if trends emerge over the years, whether it seems sensible to interpret those trends as changes in targeting strategy based on what is known about the fisheries. I also think the results from the CPUE standardisation including cluster as an explanatory variable should be contrasted with results from CPUE standardisation not using cluster as an explanatory variable, and the gains obtained from the use of clusters should be shown. It is possible that many of the aspects I indicate in this paragraph have already been examined; however, they have not been reported in the SCRS/2018/058 document that describes this work, so it is difficult for me, as an outside reader, to get a feel for how the approach has worked.

As noted in the DP report, other issues to consider regarding the standardised LL index are how to weigh the different spatial strata (i.e. the 5º cells) and what is the selectivity (i.e. length composition) of the resulting index. I agree that weighting all 5º cells equally, while removing (year, quarter, 5º cell) combinations with a very low number of sets, seems a reasonable approach at this stage and is in line with previous recommendations, as noted in SCRS/2018/058. Regarding the selectivity of the resulting index, the approach currently applied in the SS3 assessment (which is to assume it is the same as that of the Japanese fishery), is likely the best the group could do at this stage, but this should be further investigated in the future, as noted in the SA report.

As I said above, my understanding is that this index is a main driver of the stock status results. Therefore, I strongly encourage efforts to continue to develop and improve this index in coming years.

Given the focus of some of the main fisheries on young fish, an abundance index for this fraction of the population seems like a key gap in the stock assessment data, and I strongly encourage efforts to develop such an index, e.g. by appropriate standardisation of catch rates from a FADs fishery.

3.3 Biological parameters:

Growth and natural mortality are key biological parameters when trying to make any kind of inference about stock status from length frequency data, and any improvements that can be made in the knowledge of these parameters would be very useful to the stock assessment. The ongoing ICCAT tagging program AOTTP should be able to contribute useful data in this direction. This program also provides information about movement / migration, which is also relevant to understand the stock dynamics, even though I agree that a single-area SA model is the most appropriate at present.

4. TOR 2: Stock assessment methods

Three types of models have been used for the stock assessment: a surplus production model "mpb" without process error, a surplus production model "JABBA" including process error, and an age-structured "SS3" model fitted to length data (no age data used in the fit). All three alternatives were implemented in a thoughtful and competent manner, and considered an appropriate suite of diagnostics.

In line with the conclusions from the SA report, I consider the SS3 implementation to be the most suitable for the assessment of this stock and to provide management advice. This preference is mostly because of the very different selectivity of the LL, BB and PS fleets, and the increasing proportion of catches from FADs in recent years (see Figures 5 and 7 of the SA report). As a consequence, the overall selectivity for the total fishery on the stock has changed through time and evolved towards increased selectivity of young fish in recent years; I do not think this situation is well-suited for the application of surplus production models, in which the fish population is represented by the aggregate exploitable biomass and F is defined as the ratio of catch (tonnes) to exploitable biomass without taking into account the age or length composition of the population or catch. Moreover, the LL CPUE used as abundance index in the assessments corresponds only to the adult fraction of the population, and therefore excludes the juveniles which constitute a substantial proportion of the catch in recent years. So I think considerable caution has to be applied when interpreting the results of the surplus production models for this stock. Whereas the stock assessment developed with SS3 is not exempt from problems, I feel it constitutes an appropriate basis to make inference on the BET stock status and to provide management advice.

4.1 Specific comments on the SS3 assessment

A number of reasonable initial configurations and sensitivities were explored before arriving at a reference case (the so-called Run 19), around which an uncertainty grid (18 different model configurations, based on combinations of M, steepness h, and the magnitude of the annual recruitment deviations from the stock-recruitment curve sigmaR) was subsequently built in order to characterize stock status and provide management advice. I do not feel there is much I can add in the way of improvement to the work done, so I simply include a few comments and considerations below.

The 15 fishing fleets used in the assessment (essentially defined by combination of gear, geographical area and, in some cases, country and range of years of operation) seems an appropriate choice.

The change from the 3 geographical areas, with possible movement between them, included in the 2015 assessment to a single combined area in this year's assessment also seems appropriate at this stage, given the scarcity of information about movement; this, however, should be reconsidered in the future, if more movement information becomes available (e.g. via the AOTTP tagging programme).

The choice of biological parameters (weight-at-length, growth, M, maturity) is in line with current knowledge of the stock. However, any possible improvements in the knowledge of these parameters would be very useful for the stock assessment, as I noted earlier, so I encourage studies pursuing such improvements.

The joint LL CPUE index developed this year was used as an abundance index, which I find appropriate, although the comments I made earlier about this index and its selectivity remain.

The description in the SA report (Sections 3.2.8 and 3.2.10) of the selectivities of the 15 fishing fleets, is quite short and does not seem to capture the full complexity of what has been done. Further explanation would have been generally useful, for example:

- Explain why some fleets were modelled with cubic splines and others with double-normal selectivity.
- Several selectivity parameters were assigned symmetric beta prior distributions, with different values of the prior standard deviation. Some explanation about this would have been useful.
- The selectivities of some fleets used blocks for some of the parameters, but not much explanation about this is provided in the SA report in most cases. I agree with the statement at the end of Section 3.2.10 that it would be useful to develop guidelines for setting time blocks for fishery selectivity.
- The selectivity of Fleet 3 is estimated to have undergone a very big change in 2008 (see Figure 25 of SA report). The selectivity of this fleet before 2008 appears to have been entirely fixed in the SS3 input file but I could not find any comment or explanation about this in the SA report. The overall fit to this fleet around the secondary mode of 150 cm is not good (see Figure 24 of SA report) and I wonder if the fit might have been improved by allowing the selectivity to be estimated before 2008.
- The selectivity of Fleet 14 was forced to be asymptotic from 2005 onwards and the stock assessment model estimates a very big selectivity jump in 2005 (see Figure 25 of SA report). This may not be unreasonable, given the observed change in the length composition data in the early 2000s. However, such a big sudden change in selectivity seems surprising and I think it would merit some explanation in the SA report.
- The selectivity of Fleet 7 has been estimated, even though no length composition data are available for this fleet. As far as I can see, there is no significant information in the available stock assessment data from which this selectivity might be estimated. A double-normal selectivity form was assumed for this fleet, with a symmetric beta prior for the first 4 parameters and fixed values for the last 2 parameters. The point estimates for the 4 estimated parameters were close to the mid-point of the set parameter range, as expected given the absence of information about this selectivity in the data. Given this absence of information, I wonder why a prior distribution was used instead of entirely fixing the selectivity of this fleet.
- There is a minor error in Section 3.2.8 of SA report, where it indicates double-normal selectivity was used for Fleet 9; according to the SS3 input file, a cubic spline was used.

The length composition data were assigned very low weights in the overall likelihood (final Multinomial sample sizes <3 for most fleets and quarters). This was done so as to allow an improved fit to the abundance index and better retrospective behaviour. This is in line with the usual understanding in integrated stock assessment models that primacy should be given to achieving a good fit to abundance indices over fitting the composition data (e.g. as recommended by Francis, 2011).

As part of the model diagnostics, likelihood profiles were conducted for R0, sigmaR and steepness, but the SA report (Section 3.2.11) does not include any comment to help interpret the results.

Given that the stock data are not sufficiently informative to estimate steepness, including several fixed values of steepness in the uncertainty grid seems appropriate to me. Steepness is an influential parameter in the determination of MSY reference points and stock status relative to them, so it important that the selected range can be considered realistic. I do not know enough about steepness for tuna populations, but I assume experts at the meeting will have considered this to be the case for the selected range (0.7-0.9).

My understanding is that estimating sigmaR via maximisation of the (penalized) likelihood is generally not possible, so I am not convinced that examining the profile likelihood on sigmaR is informative. Including several fixed values of sigmaR in the uncertainty grid, as done by the group, seems appropriate. I wonder if there is a strong basis for including sigmaR=0.2 in the grid, is such a low sigmaR considered realistic?

In short, the only stock-recruit parameter estimated within the stock assessment is R0 (and the individual annual recruitment deviations). This is a fairly common situation and seems like a reasonable approach in this case.

Recruitment deviations were estimated starting only in 1974 (and fixed to 0 for earlier years). This may not be unreasonable given the very little data information available to estimate such deviations in the early years. I thought it might be more usual to include these deviations but using a near-zero bias correction in those early years, but I could be wrong and it may not have any significant impact on the assessment results in any case. The maximum bias correction applied (0.2) is, I would say, unusually low, but again, I could be wrong and it may not affect the results of the assessment in any significant way.

I find hindcasting analyses very useful, because they help evaluate the predictive ability of the system, which is very relevant for the provision of catch advice. I assume the hindcasting has been implemented as follows:

(1) run the assessment until year X, removing *all* data after year X (i.e. as one does in retrospective runs);
(2) project the results of the assessment ending in year X forward to years X+1, ..., 2017, conditioning on the catches observed to have occurred in each of the years X+1,… , 2017 [when doing projections with SS3, the overall fishery selectivity on the stock needs to be specified for the projection years];
(3) at this point, one could compare the projection results with those from the stock assessment ending in 2017 (as seems to have been done for mpb, see Figure 12 of SA report) or predict the abundance indices for years X+1,..., 2017, and compare the predicted with the observed index values (as seems to have been done for SS3, see Figure 30, and for JABBA, see Figure 19, of SA report).

Something seems a bit weird in the length bins defined in the SS3 input data file (this may not create any problem, but I mention it just in case):
- Population length bins have widths of 4 cm, then switch to 6 cm, then switch to 5 cm, then again switch to 6 cm…
- Data length bins do not fully align with the population length bins (a couple of interval bounds are different between data and population length bins)

Table 12 of the SA report: the settings of runs 17-19 may be incorrectly noted in the table.

Table 1 and Section 3.2.3 of SA report: the values of K and t0 stated for the Richards model may be incorrect.

I would have found it useful if the SA report displayed the formula for the Richards growth model; there seems to be multiple ways of parameterizing this growth curve and e.g. I found it difficult to figure out how to transform the parameter values in Table 1 into those needed as inputs to SS3.

Comment on code: I was a bit surprised that SS3 version 3.24 was used instead of the newer 3.30 version. In terms of developing this assessment further in coming years, I thought it would be more practical to have it based on the newer SS3 version.

4.2 Specific comments on the surplus production assessment models (mpb and JABBA)

As I already said, even though I feel the assessments were performed competently, I do not think the changing selectivity of the total combined fishery on the stock, with increased selectivity on young fish in recent years, is well-suited for the application of surplus production models.

For both mpb and JABBA, the reference assessments put forward use the newly developed joint abundance index from LL CPUE. This seems logical given the efforts spent this year on developing an improved index (however, see my earlier comment about this index representing only the adult fraction of the population). I think the SA report should clarify if the index used in these assessments is an index of biomass (as would be logical in a surplus production model) or numbers (as I had understood was the case for the LL index developed this year).

The mpb analysis used the Fox model (corresponding to Bmsy/K = 0.37), as in the 2015 assessment. I think the SA report should have included some discussion and interpretation of the found diagnostics (Section 3.1.1), for example, concerning the shape of the likelihood profiles for K and r (Figure 10 of SA report) and the difference in hindcast results obtained from removing different numbers of years (Figure 12 of SA report).

The JABBA analysis is a Bayesian implementation of a surplus production model including process error. Structural uncertainty was handled by running 3 alternative model configurations, respectively corresponding to the values of the production function shape parameter that result in Bmsy/K = 0.332, 0.306 and 0.278, where K is the virgin biomass. The SA report indicates that these Bmsy/K values correspond to SSBmsy/SSB0 with steepness h=0.7, 0.8 and 0.9, respectively, in the SS3 assessment. It should, however, be remembered that the SSBmsy/SSB0 ratio in SS3 depends not only on h, but also on the fishery selectivity, which has varied over time (so SSBmsy/SSB0 has changed through time, as can be inferred from the left panel of Figure 42 in the SA report); nevertheless I find this a useful way to capture a range of productivity scenarios roughly comparable to the SS3 assessment.

Table 10 of SA report: It would have been useful to explain the parametrisations used for the lognormal (applied for K, r and psi) and inverse-gamma (applied for sigma_process^2) prior distributions, bearing in mind that these distributions are not always parameterized in the same way. For K and r, I could not work out the correspondence between the parameters quoted for the lognormal distributions and the values stated for the 2.5 and 97.5 percentiles, hence the reason for my question.

Figure 17 of SA report (priors and posteriors): Lines for the prior distribution are missing in some of the panels. This is possibly because the priors are very wide and have very low densities in any given range of values, but a way to illustrate this in the plots should be found, or clarification provided in the figure caption.

I think the SA report should have included some discussion and interpretation of the found diagnostics for JABBA (Section 3.1.2), for example, concerning the difference in hindcast results obtained when removing different numbers of years (Figure 19 of SA report).

5. TOR 3: Population benchmarks and stock status

For all three assessment methods employed, stock status was measured in terms of B/Bmsy (where B denotes exploitable biomass for the surplus production models and SSB for SS3) and F/Fmsy (where F denotes harvest rate for the surplus production models and the average F over ages 1-7 for SS3). Reference points are key in determining stock status and, therefore, can strongly influence the perception of the stock status and the catch advice provided.

Under all three assessment models, the reference points were calculated deterministically, as the B and F values that correspond to maximum equilibrium yield, where the equilibrium yield was calculated deterministically from the parameters defining the production function in the case of surplus production models and from biological and fishery selectivity parameters in the case of SS3. The approach followed here for the calculation of Bmsy and Fmsy is, I think, common practice in many parts of the world and I do not consider it problematic.
However, I think one should also keep in mind that the long-term sustainability of a population subject to any level of fishing mortality also depends on the stochastic properties of the process error (such as the process error included in JABBA, see e.g. Bordet and Rivest (2014), or the annual recruitment deviations included in SS3, see e.g. Clark (1993)). In line with this thinking, ICES takes into account the stochastic properties of the annual recruitment deviations when calculating Fmsy (ICES, 2017). I think this generally points to the need to be cautious when interpreting what Bmsy and Fmsy obtained from a deterministic calculation actually mean and highlights the relevance of properly evaluating harvest strategies taking due account of stochastic variability in the population and uncertainty via MSE (as I understand will be done for this stock in the near future).

The MSY reference points depend on biological parameters such as steepness or M, about which there is little information currently available. For this reason, one might alternatively have considered using as reference points proxies based on e.g. 40%SPR (corresponding to an equilibrium spawner per recruit level equal to 40% of what would be achieved in the absence of a fishery) or 40%B0 (corresponding to an equilibrium SSB equal to 40% of what would be achieved in the absence of a fishery), which have generally been considered to provide sustainable and high yields under a variety of circumstances. It is, however, often difficult to justify the choice of a particular proxy over another, so the approach used by the group, which is to work with Bmsy and Fmsy (no proxies) but considering a realistic range of biological parameters and displaying stock status relative to the range of resulting Bmsy and Fmsy values, seems appropriate to me.

The overall selectivity for the total fishery on the stock has been changing over the years and different selectivities correspond to different values of the MSY reference points (see e.g. Figure 42 of SA report); therefore, I consider it appropriate to display historical stock development and current stock status based on reference points calculated with year-specific selectivities, as done in the SS3 assessment (Figures 40 and 41 of SA report).

6. TOR 4: Future population status (projections)

Projections 15 years into the future (2018-2032) were conducted under mpb, JABBA and SS3. In all cases, the catch in 2018 was fixed at a value considered as realistic (~78 kt) whereas a range of annual catch levels were applied in each year after 2018 (assuming the same catch in all projection years).

6.1 Projections with mpb

500 future trajectories were calculated for each future catch level, where the different trajectories arose from assessment uncertainty calculated from bootstrap. Each of the 500 future trajectories was calculated deterministically, from the K and r values estimated in that particular bootstrap assessment. For each projection year, results were summarised by the following statistics: median(B/Bmsy), median(F/Fmsy), P(B>Bmsy), P(F<Fmsy) and P(being in the green quadrant of the Kobe plot, i.e. F<Fmsy and B>Bmsy).

6.2 Projections with JABBA

30,000 future trajectories were calculated for each future catch level, where the different trajectories arose from 3 assessment models (corresponding to the 3 values of Bmsy/K selected to reflect structural uncertainty) and 10,000 MCMC posterior samples from each assessment. I think each of the 10,000 future trajectories was calculated using the (K, r, sigma_proc) value of the corresponding MCMC draw, simulating the process error stochastically and independently from year to year; this, however, was not explained in the SA report, it is just my understanding from a (fairly quick) look at the R code. Projection results were summarised in the same way as for mpb, with each of the 3 assessment models receiving the same weight.

6.3 Projections with SS3:

18 future trajectories were calculated for each future catch level, with each trajectory corresponding to an assessment model in the agreed uncertainty grid. The overall selectivity (i.e. selectivity of individual fleets and fleet allocation) in the projection years was assumed to be the average of 2016-2017. I think each of the 18 future trajectories was calculated deterministically assuming no annual recruitment deviations from the SR curve during the projection years. The 18 future trajectories of SSB/SSBmsy and F/Fmsy were all shown in graphs (Figures 47-50 of SA report); no summary statistics were presented in the SA report.

The SA report states that a full characterization of SS3 projections will be conducted intersessionally, and the results will be presented in a separate SCRS document during the September Species Group meeting, including Kobe strategy matrices with bootstrap estimates of uncertainty across the 18 SS3-uncertainty grid. My understanding of this statement is that the intention is to include within-model uncertainty around the 18 deterministic trajectories, through bootstrapping the corresponding assessment. Thus, the projections would include uncertainty both within and between models, similarly to what was done in the JABBA projections, which seems sensible. I think one should also aim to incorporate annual stochastic recruitment deviations from the SR curve in the projection years, in order to capture uncertainty in future recruitment, otherwise uncertainty would be underestimated in the projections; however, how to do this in an appropriate manner would require some thinking, as simply randomly drawing annual recruitment deviations with the sigmaR value fixed in the assessment may not provide sufficient realism. As a projection diagnostic, I think it would be useful to compute, for each projection year, the proportion of the catch and SSB made up of year-classes born in 2018 and later years, as these are year-classes about which no information is available at this stage.

At the time I prepared this review (beginning of September), the additional SCRS document on the SS3 projections was not available so I can not comment on it, but I support the aim of further developing the SS3 projections by giving more consideration to the appropriate inclusion of uncertainty.

7. TOR 5: Characterization of uncertainty

The work conducted by the group pays due account to uncertainty, considering both within-model uncertainty (i.e. uncertainty in parameter estimates under a particular stock assessment model) and between-model uncertainty. I consider the explicit consideration of between-model uncertainty (such as was done with the 18 SS3-model uncertainty grid, recognising uncertainty in key biological parameters) to be a very useful step towards more realistically characterizing uncertainty in the results, so I very much support this approach.

Quantifying the true degree of uncertainty is always extremely difficult, but I feel a reasonable job has been done here, bearing in mind the usual time/practical limitations. I consider the implications of uncertainty in technical conclusions to be clearly stated.

8. TOR 6: Presentation of results in the SA report

I found both the DP and SA reports well-structured and, on the whole, clearly written. However, as an external reader not present at the meetings, I generally felt that explanations and/or discussion were often a bit too brief or entirely lacking for important issues such as e.g. the reasoning followed to arrive at a particular assessment model configuration and at a reference case, or the interpretation of diagnostics and action taken to account for the issues detected by the diagnostics.

9. TOR 7: Stock assessment SCRS process

This was the first time I conducted a stock assessment review for a tuna RFMO, and I was totally new to most aspects of the data, fisheries and stock assessment. I was offered the opportunity to take part in the SA meeting, but this was not possible because of other commitments. I therefore reviewed the work by correspondence and my understanding of the SCRS process is probably very incomplete.

My impression, though, is that the process was well-organised. Having a DP meeting followed by the SA meeting a few months later seems like a sensible way to organise the process. I also got the impression that these meetings were open to a broad range of participants and that the scientific work was conducted and discussed in an open manner, giving participants the opportunity to engage in teams to conduct various aspects of the work.

As a reviewer by correspondence, I was given full access to all documentation available to the scientists participating in the meetings. Unavoidably, there was a large amount of documents (several working documents on the fisheries, on various kinds of data and on stock assessment; background documents; presentations; files with code or input-output files for the different stock assessment methods and ensuing projections...), so, even though the online site containing the material was well organised, it was still quite time-consuming to get an understanding of what was there and to work my way through the main materials. I think conducting these reviews should be easier and faster if the reviewer can attend the SA meeting, as many issues can be discussed and clarified at that point. Nevertheless, I can also see merits in a review by correspondence, for example, it gives the reviewer more time to think through some issues before commenting.

In summary, I think the process was appropriate and have no particular suggestions for improvement at this stage.

10. TOR 8: Workplan

As an external reviewer, I do not know the workplan leading up to the present assessment. For what I can see, the scientists involved in the various aspects concerning the data and assessment made good efforts to develop the assessment and advice in a sound manner, addressing issues identified in previous work to the extent possible and making recommendations for future work when considered needed.

11. TOR 9: Recommendations

My recommendations for future priority work were identified throughout this report, and I list them here. I note that there is a high degree of overlap between my recommendations and those provided by the group (Section 9 of the DP report and Section 6.1 of the SA report), which I support.

11.1 Recommendations

- Continue efforts to improve fisheries-dependent data (catch, effort, spatio-temporal distribution of catch, length frequencies and catch-at-size matrices).
- Given the strong rise of PS catches over time, it is very important to have good-quality data on the PS fisheries and consider that efforts in that respect are particularly relevant.
- It is crucial to ensure that the newly developed abundance index from LL CPUE is reliable and represents a step in the correct direction for the stock assessment. I strongly encourage efforts to continue to develop and improve this index in coming years, focusing among other things in demonstrating the ability of the clustering method to differentiate between changes in targeting strategy and changes in relative abundance of the different species (simulation studies could aid in this regard) and further considering how best to represent the selectivity of the index.
- Given the focus of some of the main fisheries on young fish, an abundance index for this fraction of the population seems like a key gap in the stock assessment data, and I strongly encourage efforts to develop such an index, e.g. by appropriate standardisation of catch rates from a FAD fishery.
- Growth (Linf and K) and natural mortality (M) are key biological parameters to make any kind of inference about stock status from length frequency data, and any improvements that can be made in the knowledge of these parameters would be very useful to the stock assessment. The ongoing AOTTP tagging program should be able to contribute useful data in this direction.
- The AOTTP tagging program also provides information about movement / migration, which is also relevant to understand the stock dynamics, even though I agree that a single-area SA model is the most appropriate at present. I recommend that efforts be made to make the best possible use of the scientific data generated by this program.
- I do not have any particular recommendation concerning the stock assessment method. I consider the assessment built with SS3 to be appropriate for the characteristics of this assessment and I support its further development in the future.
- I note the need to be cautious when interpreting what Bmsy and Fmsy obtained from a deterministic calculation actually mean and recommend properly evaluating harvest strategies taking due account of stochasticity in population dynamics and uncertainty via MSE (as I understand will be done for this stock in the near future).

References

Bordet, C., and Rivest, L.-P. (2014). A stochastic Pella Tomlinson model and its maximum sustainable yield. Journal of Theoretical Biology 360, 46–53.

Clark, W. G. (1993). The effect of recruitment variability on the choice of a target level of spawning biomass per recruit. Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, Alaska Sea Grant College Program, AK-SG-93-02.

Francis, R.I.C.C. (2011). Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124-1138.

ICES (2017). ICES fisheries management reference points for category 1 and 2 stocks. ICES Advice Technical Guidelines.