REPORT OF THE MANAGEMENT
STRATEGY EVALUATION TECHNICAL TEAM
14-16 FEBRUARY 2022

John Walter[1] and Cassidy Peterson[1]

SUMMARY

This document is an informal report of the February 14-16th, 2022 BFT MSE technical team meeting. The report was not adopted by the Group and the meeting was not a formal meeting of the SCRS Bluefin tuna Working Group. Nevertheless, the details from and discussions at the meeting are pertinent to the Bluefin Working Group. Many of these details will be addressed and included in the material presented to the March Panel 2 meeting.

RÉSUMÉ

Le présent document constitue un rapport informel de la réunion de l'équipe technique sur la MSE tenue du 14 au 16 février 2022. Le rapport n'a pas été adopté par le Groupe et la réunion n'était pas une réunion officielle du Groupe d'espèces sur le thon rouge du SCRS. Néanmoins, les détails et les discussions de la réunion sont pertinents pour le Groupe d'espèces sur le thon rouge. Nombre de ces détails seront abordés et inclus dans le matériel présenté à la réunion de la Sous-commission 2 du mois de mars.

RESUMEN

Este documento es un informe informal de la reunión del Grupo técnico sobre la MSE para el atún rojo, que se celebró del 14 al 16 de febrero de 2022. El informe no fue adoptado por el Grupo y la reunión no fue una reunión formal del Grupo de especies de atún rojo del SCRS. No obstante, los detalles y debates de la reunión son pertinentes para el Grupo de especies de atún rojo. Muchos de estos detalles se abordarán y se incluirán en el material que se presentará en la reunión de la Subcomisión 2 de marzo.

KEYWORDS

Atlantic Bluefin Tuna, Management Strategy Evaluation

---

[1] NOAA-Fisheries, SEFSC, Miami, FL

Introduction

This document represents and informal report of the BFT MSE Technical Team. It is not a formal, adopted report but constitutes a series of notes compiled to record discussion. The document follows the agenda adopted (**Appendix Table 1**).

Results

1.   Round robin from CMP developers

    a. 1_CMPs_Butterworth_Rademeyer_final
    b. Any other papers from developers

Developers were asked to please provide updated mathematical descriptions to include with these meeting materials. Note mathematical descriptions from the previous meeting can be found here (see appendix: https://www.iccat.int/Documents/Meetings/Docs/2021/REPORTS/2021_BFT2_ENG.pdf). It was also proposed that the simplified CMP table (from the ambassador presentation) could be updated (see appended table). Developers were also asked to please ensure "living" CMP excel spreadsheet is updated (https://docs.google.com/spreadsheets/d/1ajxmmG9f8HZ5KB8hy1vXemUtb6Kr-npg_LxJo0NGwC4/edit?usp=sharing).

2.   CMP summary (Tom: **4_Preliminary CMP results_Feb_14_2022_v2.pptx**)
    a. Key plots
        i.   *Tuning objective comparison*
        ii. *Violin plot of Br30 & yield (bimodality)*
        iii. *E/W tradeoff in yield showing no strong negative correlation*

    b. Key trajectories to show
        i.   *Line plot of Br & yield*
        ii. Demo of what responsiveness of an index means and why it will be quite evident if a 'regime' shift occurs in the future.
        iii. Management cycle projection of generic MP (indices + TACs), with 3 different versions for the 3 different recruitment scenarios to show how a generic MP would respond

The MSE contractor presented preliminary results (*4_Preliminary CMP results_Feb_14_2022_v2.pptx*).

Notable highlights include the need for a **new CMP naming convention** as follows:

<u>MP#A</u> – where MP would be replaced by the CMP name (e.g., BR), the number would indicate which tuning target (1-4), and the final letter would indicate which CMP configuration was implemented (A-D,X). Letter appendices are coded following:

    A—no cap in the East (or West), 20% annual allowable TAC change restriction
    B—55K ton cap in East, 20% annual allowable TAC change restriction
    C—no cap in the East (or West), increased flexibility in annual allowable TAC change
    D—45K ton cap in East, 20% annual allowable TAC change restriction
    X—some other configuration

The new naming convention was accepted.

The Group also highlighted the clear bimodality in some resulting performance statistics (e.g., in AvC30). The Group clarified that the bimodality was a reflection of the adaptive nature of CMPs, and of the extent to which they were able to react to the biology in the underlying OM scenarios. This may not seem a natural result to scientists or stakeholders who are unfamiliar with MSE and accustomed to seeing more traditional stock assessment results. This example emphasizes that **the bimodality is actually a very good performance response and that the top performing CMPs should exhibit this, given the bimodal nature of yields from the stock recruitment levels considered in the OMs.** Consideration should be given to presenting violins with associated worm plots, or some explanatory graphical explanation of the appropriate interpretation of results should accompany presentation of results. It was further stressed that performance metrics should be fully explained (e.g., OFT) during presentations.

A notable result from the presentation was **that TAC caps are not as influential on CMP performance** as previously presumed. Counterintuitively, restrictive caps resulted in more aggressive CMPs when tuned to the same Br30 target, since caps served to limit the allowable catch of OMs that performed well and increased their corresponding Br30 such that more aggressive tuning was required to reduce the resulting weighted average Br30 over the reference OM grid to the intended target. However, it was noted that the primary advantage of TAC caps was reduction in annual yield variability AAVY, for which results were not presented due to lack of time. **The impact of TAC caps on AAVY should be more thoroughly explored before the Group might agree to abandon their usage, as they could be needed to achieve greater yield stability**.

The presentation also **demonstrated improved performance when the downward flexibility in the annual allowable % TAC change was increased**, which was to ensure adequate adaptive responsiveness to recruitment scenarios 2 and 3. Though median behavior of CMPs did not change across the reference grid (e.g., Br30 or AvC30), the violin tails (or correspondingly the behavior of the worst performing OMs; 5th %ile Br30) was improved by increased flexibility, such that fewer stocks collapsed when projected. It was pointed out further that annual TAC changes greater than 20-30% were rarely observed across the reference OM grid (only when needed to accommodate for a downward biomass trend), which was reflected by minimally impacted AvC30 values. This should ease some concern regarding realized annual changes when a CMP is ultimately implemented.

3.   Selected performance statistics (colored patchwork or "quilt" plot)

Results of the quilt plot were presented, and it was noted that the CMP ranking would depend on the weight of the performance statistics included in the plot, and further that duplicative metrics could lead to over-weighting of corresponding management objectives. These rankings are also determined by scaling CMP performance from the best to worst performing for each metric, such that relative differences in performance are preserved; as such, changes in any CMP included could impact these rankings. The Group also considered the appropriateness of some metrics and their resulting influence on CMP rank. For instance, though all CMPs tuned to the same Br30, slight differences (e.g., 1.48 vs. 1.52) in median Br30 would be visually emphasized and impact cumulative ranking in the quilt plot despite relatively minor differences. Median Br30 was therefore excluded from the quilt plot. Attention should be paid to the performance statistics values across CMPs when interpreting these plots, so as not to be swayed by patches that look very different in color despite similar results. The Group noted further that these plots may be somewhat overwhelming to examine, and suggested reducing the number of performance statistics to 4-6 per area. Input on the selection and relative importance of each performance statistic will be needed to reduce the dimensionality of these plots.

The Group noted that **different performance statistics that better encapsulate Stakeholder perspectives (e.g., short-term, 10-year statistics) should be considered**.

To reduce the number of metrics, **the Group agreed to drop AvgBr_5%, LD_15%, LD_50%, and C30_50% from this quilt plot**, citing that each was redundant of other metrics. The final column, **"NormPerf" (or to be renamed in the future by the Contractor) will be eliminated from the current plot**. However, it may be retained in some variants of the plot if itsuse would help to clearly demonstrate the value of including/removing TAC caps and increasing downward flexibility in allowable catch. The Group noted that the **weighting of the remaining performance statistics may need to be adjusted to reflect the importance of each management objectives**. The Group also noted that the performance metrics removed would still need to be calculated for subsequent analyses, and could be retained in a separate, master quilt plot. In particular, LD and lower percentile metrics for Br will be informative during the upcoming discussion on the B$_{lim}$ reference point.

The Group discussed the lack of a **Blim** value at present. It was noted that the value of Blim would be considered, if necessary, at the April meeting and that the performance statistics related to the 5th and 15% percentile of lowest depletion reflect different considerations of Blim and the probability of falling below it, both of which need to be considered.

The Group also noted that some measure of overfishing would need to be added to the quilt plot as that metric is developed. Candidate metrics included OFT, median U/UMSY averaged over years 1-30, % of time that B>BMSY, and probability green Kobe (PGK). **The Group determined that further development of these proposed overfishing metrics was needed before they could agree on a single metric for the quilt plot**.

4. Comparing different CMPs
    a. Does the relative rank differ by tuning criteria?
    b. Discussion of process for winnowing CMPs

The Contractor's presentation showed that **CMP ranking was reasonably conserved across tunings**. The rankings were more highly conserved in the East than in the West. Nevertheless, clear correlations between tuning levels 1 & 3 (Eastern Br30 targets of 1.25) and 2 & 4 (Eastern Br30 targets of 1.5) were evident. As such, **it was proposed that at least one of each tuning level (either 1 or 3 and either 2 or 4) might be removed for the purposes of CMP comparison**.

Ways in which the number of CMPs under consideration could be reduced were discussed. Discussion centered around how the selection would vary based on how each individual stakeholder ranked the performance metrics (e.g., as presented in colored quilt plot). One way was to remove CMPs that receive very little support.

Ultimately, the Group clarified that we are **seeking to present the management trade-off space to Panel 2**, and the advice of the Panel will influence the final CMP tuning. It may therefore be beneficial to retain differently-performing CMPs at this stage to ensure that the management trade-off space is fully characterized.

5. Panel 2 requests
    a. F statistic (**Caruthers and Butterworth: 2_Complications FMSY_Carruthers_Butterworth, Butterworth: 3_F performance statistic.docx**)
    b. Mixing table
    c. Operational management objectives

The Group emphasized the desire to understand any anticipated questions or concerns that Panel 2 would have to ensure proper preparation for that meeting. Some stakeholder concerns were accordingly discussed by the Group.

The first was the emphasis of performance statistics 30 years in the future and the perceived lower emphasis placed on near-term fishery dynamics. The Group noted the need for longer time horizons (>1 generation) rather than measuring CMP performance based only on transient, short-term simulation dynamics. Given that CMP rankings change when ranking based on short-term (10-year) or long-term (30-year) dynamics, the **Group agreed to give more emphasis 10-year performance statistics when comparing performance**.

Additional discussion centered around potential negative **stakeholder perspectives on indices** used in several CMPs. Concerns were raised informally about the MexUS GOM PLL index, GOM larval index, US RR 177+ index, and the 66-144 index (note some of these were not recommended for use in CMPs by the Group due to poor diagnostic fit in the OM conditioning. These included the US RR 66-114, US RR 115-144, and US RR 177+; see Table 7.1 from TSD). It was noted that some stakeholders trust JLL indices and the Canadian rod and reel index, though these were not recommended for CMPs due to severe residual pattern. Developers are free to remove these perceived problematic indices from their CMPs to demonstrate whether alternative CMPs consistent with some stakeholder opinions about indices performed better.

Some concern was raised in relation to indices that were representative of west stock-of-origin biomass and not the broader western fishable area, and a stakeholder desire for indices to be representative of areas in which the heaviest fishing activities are occurring was expressed. The Group noted, however, that it could be important to retain stock-of-origin indices to preserve indicators of stock safety, considering that the Commission-defined management objectives include stock-specific biomass reference points. Further, the CMPs are still being scaled to stock-of-origin Br30 which implicitly accounts for some mixing assumptions. At least one class of CMPs (LW) reflected improved performance when including Western area information to west stock-of-origin indices in the absence of complete characterization of mixing data within the western CMP.

The Group further clarified that CMP developers have thus far been more concerned about CMP performance than perceived opinions of indices, stressing the importance placed on robust CMP performance from an analytical perspective. Notably, indices have already undergone a substantial vetting process (including analysis of index fits to the conditioned OMs and scrutinized by a designated bluefin tuna Index Working Group), about which many stakeholders may be unaware. **The Group concluded that they will more explicitly discuss the acceptability of indices with stakeholders**, in addition to analytic evaluation from strictly a performance perspective. The reason that Group has not had that conversation with Panel 2 and stakeholders is that the first step is simply to evaluate

performance, and that indices that do not improve CMP performance do not need further discussion. While stakeholder input should be considered, it should not replace the substantial index investigation already conducted by the Group.

Stakeholders also voiced concern about pressure associated with developing a control rule that would be in place for decades into the future, consequently believing that if CMPs were developed with indices that they did not trust, they would be stuck with those decisions moving forward. The Group emphasized that the **MP review process**, which, though yet to be specified in detail, will involve periodic reconsideration after implementation (likely every ~5-6 years). It was stressed that the Group would be open to revisions during these periodic reviews, including changing which indices were utilized within the MP. Further, there will be determination of **Exceptional Circumstances and the actions that follow**, where some component of the MP becomes "problematic" (e.g., fishing strategy changes, impacting the indices such that they are no longer proportional to abundance) or where future circumstances stray from those modeled in the MSE process. **The Group noted the importance of ensuring that stakeholders were aware of these review processes**.

The Group noted that these are the conversations that will comprise the next step of the MSE process after the next Panel 2 meeting. **The Group will be responsible for working with stakeholders to foster transparency and buy-in to the process**, e.g., receiving suggestions and demonstrating how those decision points result in management trade-offs.

The Contractor presented and the Group subsequently discussed **development of an overfishing metric** (e.g., F/FMSY; *Carruthers and Butterworth: 2_Complications FMSY_Carruthers_Butterworth, Butterworth: 3_F performance statistic.docx*), following fisheries assessment convention and as a measure that is mandated by the Commission. Calculation within the MSE context is both conceptually and computationally challenging, as it is complicated by the large number of OMs, the multi-stock and multi-area nature of the fishery, and stochastic and nonstationary OMs (e.g., changing recruitment and selectivity-at-age). As such, fishing at FMSY within these scenarios would likely not lead to anticipated stock dynamics (e.g., movement to BMSY), and therefore the F ratio does not necessarily reflect the dynamics that it is intended to measure, especially in the context of a simulation wherein the resulting biomass outcomes can be observed. For example, the probability of overfishing statistic does not indicate the magnitude of overfishing, and thereby does not inform on the corresponding biomass outcomes, which is the principal concern when measuring CMP performance. **The Group agreed that given these considerations, an overfishing metric would be redundant in terms of CMP ranking and performance, but nevertheless some overfishing metric will need to be reported to fulfill the Commission's mandate to inform on and reduce overfishing**.

Three potential metrics of overfishing were proposed:

    (1) overfished trend (OFT),
    (2) relative exploitation rate (U=catch/biomass) averaged over many years (Uratio), and
    (3) probability of green Kobe.

Despite challenges associated with stock-measured (biomass) versus area-measured (catch) quantities and the computational inability to measure annual Uratio, the Group supported calculation of relative exploitation rate (Uratio) by averaging over years (either 1-30 or 10-year intervals), as the exploitation rate would be less impacted by changes in selectivity. Ultimately, **the Group convened a small team to discuss options, computational solutions, and the path forward offline. This small group will report back to the Group and updates will be reflected by the Contractor in the TSD**.

The Contractor presented the 'Mixing table', showing the percentage of eastern biomass comprising catch in each area.

6.   Panel 2/Ambassador materials **BFT_Ambassador_Feb2022.pdf**
    a.   4-pager
    b.   Presentation
    c.   1 page

The MSE Communications Group presented and the Group accordingly discussed which plots should be included in the MSE outreach materials (e.g., *BFT_Ambasador_Feb2022.pdf*). Continued presentation demonstrating the limited East/West tradeoff, violin and line plots of Br30 and yield to exhibit bimodality, and the pre-existing figure demonstrating the management cycle were accepted by the Group. **Additional curation and development of these materials will proceed during the MSE Communications Group meetings outside of the MSE technical team meeting**.

The Group also discussed the decision points for which resolutions were desired at the Panel 2 meeting. Main points (and accompanying supporting information) included demonstrating the need for and receiving acceptance for increased downward flexibility in allowable TAC changes, removal of TAC caps unless the desire for stability in yield outweighs reduction in biomass risk, and with input on how to prioritize the management objectives and corresponding performance metrics that will be used to rank CMP performance. The Group highlighted that the **Commission's main decision point would be on deciding final tuning targets, so it was most important to convey the trade-offs among tuning levels**.

The Group noted the need to approach the Panel with care to fully explain and characterize results and concepts as appropriate for the audience, and with very specific descriptions on the guidance that will be sought and why.

The Group also expressed the desire to remind the Panel of the pre-agreed science vs. management roles and responsibilities of the MSE process, with the intent to build confidence in the activities of the MSE technical team.

7.   Other matters
      a.   TSD/HTMP curation
      b.   Workplan

The MSE Contractor agreed to update the TSD as work progressed, particularly with respect to development of an overfishing performance statistic.

Some concern was expressed regarding the Group's progress to date, particularly with receiving input from stakeholders and Panel 2. The Group noted the desire for more formal and thorough timelines to be created for decision points moving forward.

Table 1. Adopted agenda

1. Round robin from CMP developers

   a) 1_CMPs_Butterworth_Rademeyer_final
   b) Any other papers from developers

2. CMP summary (Tom)

   a) Key plots
      i. *Tuning objective comparison*
      ii. *Violin plot of Br30 & yield (bimodality)*
      iii. *E/W tradeoff in yield showing no strong negative correlation*

   b) Key trajectories to show
      i. *Line plot of Br & yield*
      ii. *Demo of what responsiveness of an index means and why it will be quite evident if a 'regime' shift occurs in the future.*
      iii. *Management cycle projection of generic MP (indices + TACs), with 3 different versions for the 3 different recruitment scenarios to show how a generic MP would respond*

3. Selected performance statistics (colored patchwork or "quilt" plot)

4. Comparing different CMPs
   a) Does the relative rank differ by tuning criteria?
   b) Discussion of process for winnowing CMPS

5. Panel 2 requests
   c) F statistic (**Caruthers and Butterworth: 2_Complications FMSY_Carruthers_Butterworth, Butterworth: NOTES REGARDING AN F OR F/Fmsy PERFORMANCE STATISTIC**)
   d) Mixing table
   e) Operational management objectives

6. Panel 2/Ambassador materials **0_BFT_Ambassador_Feb2022_v1.docx**
   f) 4-pager
   g) Presentation
   h) 1 page

7. Other matters
   a. TSD/HTMP curation
   b. Workplan

Table 2. Brief description of CMPs.

CMP	Indices used		Formulae for calculating TACs	References
	EAST	WEST		
FZ	FR AER SUV2 JPN LL NEAtl2 W-MED LAR SUV	US RR 66-144, CAN SWNS RR US-MEX GOM PLL	TACs are product of stock-specific F0.1 estimates and estimate of US-MEX GOM PLL for the West and W-MED LAR SUV for the East.	SCRS/2020/144 SCRS/2021/122
AI	All	All	Artificial intelligence MP that fishes regional biomass at a fixed harvest rate.	SCRS/2021/028
BR	FR AER SUV2 W-MED LAR SUV MOR POR TRAP JPN LL NEAtl2	GOM LAR SUV US RR 66-144 US-MEX GOM PLL JPN LL West2 CAN SWNS RR	TACs set using a relative harvest rate for a reference year (2018) applied to the 2-year moving average of a combined master abundance index. In recent refinement, the weighting range across individual indices on the east area master index has been reduced, given that this resulted in improved resource conservation performance.	SCRS/2021/121 SCRS/2021/152
EA	FR AER SUV2 W-MED LAR SUV MOR POR TRAP JPN LL NEAtl2	GOM LAR SUV JPN LL West2 US RR 66-144 US-MEX GOM PLL	Adjust TAC based on ratio of current and target abundance index.	SCRS/2021/032 SCRS/2021/P/046
LW	W-MED LAR SUV & JPN LL NEAtl	GOM LAR SUV & MEXUS_LL	TAC is adjusted based on comparing current relative harvest rate to reference period (2019) relative harvest rate.	SCRS/2021/155
NC	MOR POR TRAP	US-MEX GOM PLL	TAC is updated using an average of an index in recent years compared to an average in previous years. The scale of TAC increase/decrease is controlled based on the trend in catches and indices	SCRS/2021/122
PW	JPN LL NEAtl2 & W-MED LAR SUV	US-MEX GOM PLL & GOM LAR SUV	TAC is adjusted based on comparing current relative harvest rate to reference period (2019) relative harvest rate.	SCRS/2021/155
TC	MOR POR TRAP JPN LL NEAtl2 W-MED LAR SUV GBYP AER SUV BAR	US RR 66-144	TAC is adjusted based on F/F_{MSY} and B/B_{MSY}.	SCRS/2020/150 SCRS/2020/165
TN	JPN LL NEAtl2	US RR 66-144 JPN LL West2	Both area TACs calculated based on their respective JPN_LL moving averages, unless drastic drop of recruitment is detected by US_RR index.	SCRS/2020/151 SCRS/2021/041