Review of the Indian Ocean Tuna Commission bigeye tuna management procedure and its underlying management strategy evaluation

Philipp Neubauer

June 2025

Abstract

The present report provides an evaluation of the Indian Ocean Tuna Commission's (IOTC) Management Procedure (MP; MP1 Harvest) for bigeye tuna (*Thunnus obesus*), adopted through IOTC Resolution 22/03. MP1 Harvest is a model-based MP designed to set total allowable catches (TACs) for bigeye tuna (BET) in the Indian Ocean, utilizing a Pella-Tomlinson biomass dynamic model fitted to historical catch and standardized longline CPUE data. The MP was underpinned by a management strategy evaluation (MSE) process that tested its performance across a range of plausible operating models and robustness scenarios. The present review focuses on the procedure's design, its underpinning management strategy evaluation (MSE), and its alignment with established best practices for management procedure development, MSE, and process for running MPs.

The review found that MP1 Harvest and the associated MSE process align with many best practices. The MP is explicitly linked to quantitative, probabilistic management objectives related to stock status and overfishing. This allows for explicit tuning of the MP's performance against these pre-agreed benchmarks. MP1 Harvest includes a clearly specified HCR, based on a Pella-Tomlinson model and a hockey-stick rule, which dictates how TACs are adjusted in response to changes in estimated stock status. A strength of the underlying MSE for MP1 Harvest is the identification and incorporation of a number of biological and fishery-related uncertainties into a set of operating models, as opposed to relying on a narrowly defined operating model. This explicit consideration of uncertainty is fundamental to designing MPs that are robust to incomplete knowledge and natural variability. The MSE then tested multiple candidate procedures and their robustness to a number of scenarios, in alignment with documented best practices. The management framework includes provisions for periodic review (by 2030) and an Exceptional Circumstances Protocol (ECP), demonstrating a commitment to adaptive management that allows the system to respond to new information or unforeseen developments.

While MP1 Harvest represents a significant advancement for IOTC, incorporating quantitative objectives and explicit uncertainty consideration, several areas may warrant future investigation and further development to ensure long-term sustainability and performance of the BET MP. These include improving the characterization and communication of complex uncertainties, such as OM plausibility; implementing realistic levels of implementation error and incorporating explicit data uncertainties, and ensuring the long-term institutional commitment to high-quality

data collection and standardised analysis programs that are fundamental to the MP's operation. Developing a range of objective, and ideally simulation-tested indicators for the Exceptional Circumstances Protocol will enhance the robustness of the MP.

The report expands on these findings and provides evidence-based recommendations for potential future enhancements to the MP framework to ensure it aligns with recognised best practice and matures towards a robust long-term management tool for BET in the Indian Ocean. Many of the recommendations in this review represent potential refinements, rather than fundamental issues with the MP or MSE that would warrant immediate re-evaluation or changes. Nevertheless, the suggestions may contribute to increase long-term performance and robustness of the BET MP.

1. Introduction

Bigeye tuna (*Thunnus obesus*; BET) is a highly valuable, migratory species of significant economic and ecological importance in the Indian Ocean. It is exploited by diverse fishing fleets from numerous nations, employing various gears, predominantly longline and purse seine. The complex nature of this fishery, involving multiple jurisdictions and gear interactions, presents substantial management challenges.

The Indian Ocean Tuna Commission (IOTC) is the Regional Fisheries Management Organisation (RFMO) mandated with the conservation and sustainable management of tuna and tuna-like species within its area of competence. Historically, fisheries management within RFMOs, including the IOTC, often relied on integrated stock assessments to determine stock status and annually negotiated quotas. Such approaches can be susceptible to short-term socio-political pressures and may not fully or transparently account for the pervasive uncertainties inherent in fisheries systems. The evolution towards Management Procedures (MPs), developed and tested through management strategy evaluation (MSE), represents a paradigm shift (Butterworth, 2007). This shift aims to establish more robust, precautionary, and transparent decision-making frameworks, which are particularly crucial for managing complex transboundary stocks.

The scientific advice provided by the IOTC Scientific Committee (SC) and its subsidiary bodies, such as the Working Party on Tropical Tunas (WPTT) and the Working Party on Methods (WPM), is fundamental to the development, implementation, and review of management measures, including MPs. These bodies play a critical role in conducting stock assessments, developing MSE frameworks, and providing recommendations to the Commission.

The adoption of MP1 Harvest by the IOTC aligns with a global movement by RFMOs towards more structured, pre-agreed, and uncertainty-robust decision-making frameworks. This trend is particularly vital for complex transboundary stocks managed by RFMOs, and developing MPs is widely recognized as best practice for RFMOs (Gutteridge et al., 2024; Lodge et al., 2007; Nakatsuka, 2017). The inherent difficulties of RFMO decision-making, often characterized by varying interests and capacities, make such structured and pre-evaluated approaches even more critical for achieving long-term sustainability and reducing management conflicts.

Through Resolution 22/03, the IOTC formally adopted MP1 Harvest for the management of BET. This resolution signifies a commitment to a systematic and scientifically tested approach for setting catch limits (TACs) for BET. The full text of Resolution 22/03 provides the legal and operational basis for this management procedure.

Management strategy evaluation (MSE) is the core scientific process that underpins the development and testing of MP1 Harvest (Kolody & Jumppanen, 2021; Kolody et al., 2020). MSE is a simulation-based framework that assesses the likely long-term performance of different pre-defined management strategies (or MPs) against a set of agreed management objectives, across a wide range of explicitly defined uncertainties regarding stock dynamics, fishery operations, and data quality. The MSE process for MP1 Harvest was a multi-year endeavour involving extensive technical work. This included the development of a suite of Operating Models (OMs) representing plausible alternative states of nature for the Indian Ocean bigeye tuna stock, the explicit characterization of uncertainties, and the simulation testing of nine candidate MPs to evaluate their ability to meet the IOTC's management objectives.

This review aimed to conduct a critical and detailed evaluation of the IOTC's MP1 Harvest and its associated MSE process against internationally recognized best practices in fisheries management and

MSE. The review identifies key strengths, weaknesses, and areas for future improvement in the IOTC's approach to managing bigeye tuna, with the aim of contributing to the long-term sustainability of the fishery and providing recommendations for enhancing the MP frameworks used for BET and other species in the IOTC. The detailed Terms of Reference for the review are provided in Appendix A.

2. The IOTC bigeye tuna MP

The following provides a summary of the BET MP, its testing and application. This section serves to provide a reflection of the basis on which the review and recommendations in the following section are based.

2.1. Core components of MP1 Harvest (IOTC Resolution 22/03)

The IOTC, through Resolution 22/03, adopted MP1 Harvest as the management procedure for bigeye tuna in its area of competence. This resolution details the objectives, mechanisms, and operational aspects of the MP.

2.1.1. Management objectives

The overarching aims of MP1 Harvest, as stipulated in IOTC Resolution 22/03 , are to maintain the spawning biomass (SB) of bigeye tuna in the green zone of the Kobe plot (i.e., $SB > SB_{MSY}$ and $F < F_{MSY}$), thereby ensuring the stock is not overfished and not subject to overfishing. Concurrently, the MP aims to maximize average catch and reduce inter-annual variability in the Total Allowable Catch (TAC).

Specific, quantifiable probabilistic targets underpin these goals:

- A 60% probability of the spawning biomass achieving the level capable of producing Maximum Sustainable Yield (SB_{MSY}) by the period 2034-2038.

- Avoidance of the interim limit reference point (LRP), defined as 20%SB_0 (20% of unfished spawning biomass), with high probability.

The formal adoption of clear, quantitative, and probabilistic management objectives is a significant strength of the IOTC's approach, aligning with international best practices. This explicitness facilitates objective performance evaluation and enhances transparency. In addition, the evaluation of alternative targets during the MSE testing phase provides some contrast about the interaction between targets and MP performance.

2.1.2. Harvest control rule (HCR)

MP1 Harvest employs a Pella-Tomlinson biomass dynamic model as its assessment component. This model is fitted to historical data series of total catch and a standardized, spatially aggregated longline CPUE index. The Pella-Tomlinson model is a generalized surplus production model that allows the biomass level at which maximum surplus production occurs to be a fraction of the carrying capacity (K), determined by a shape parameter 'm' (or 'n'). This flexibility allows it to represent a range of production curve shapes, including the Schaefer model (where m=2 or n=1) as a special case. Its parameters typically include K, intrinsic rate of growth (r), and the shape parameter (m).

The HCR itself is a "hockey stick" rule that adjusts fishing mortality (F) based on the estimated biomass depletion (B_y/K). The specific formulation from Resolution 22/03 is:

$$HCR_{\text{mult}} = \begin{cases} 1 & \text{if } B_y/K \geq 0.4 \\ (B_y/K - 0.1)/0.3 & \text{if } 0.1 < B_y/K < 0.4 \\ 0.0001 & \text{if } B_y/K \leq 0.1 \end{cases}$$

The overall fishing mortality used to estimate the TAC is calculated as:

$$F_{target} = F_{MSYratio} \cdot HCR_{mult} \cdot F_{mult},$$

where $F_{MSYratio}$ is the ratio of fishing mortality to the value which produces MSY, and F_{mult} is a tuning parameter. The new TAC is then calculated using this F_{target} and the estimated current biomass (B_y): $TAC_{new} = B_y(1 - exp(-F_{target}))$.

A critical constraint within the HCR is that TAC changes are capped at a maximum of 15% (increase or decrease) relative to the TAC from the previous period.

2.1.3. Data inputs

The operation of MP1 Harvest depends on consistent and accurate data inputs:

- Annual total catches of bigeye tuna in the IOTC Area of Competence, from 1980 to the most recent year of data available. These are collated by the IOTC Secretariat.
- A standardized, spatially aggregated longline CPUE series, from 1980 to the most recent year of data available. This series is derived from the joint standardization analysis approach described in Hoyle et al. (2019).

2.1.4. TAC setting and Review

The process for setting and reviewing the TAC under MP1 Harvest is structured as follows:

- The IOTC Scientific Committee (SC) is responsible for running the MP using the latest available data and advising the Commission on the resulting TAC.
- The first TAC derived from MP1 Harvest was applied to the years 2024-2025. Subsequent TACs are set for three-year blocks.
- A formal review of the performance of MP1 Harvest by the Commission and its subcommittees is scheduled for 2030. This review will assess whether the MP is performing as expected and if conditions warrant reconditioning OMs, retuning the MP, or considering alternative MPs.

2.1.5. Exceptional circumstances protocol (ECP)

An ECP is a critical component of an MP framework (de Moor et al., 2022), designed to address situations where the system behaves in ways not anticipated during the MSE testing, or where the MP is clearly failing to meet objectives. Guidelines for exceptional circumstances were adopted in IOTC SC (2021), with ongoing work mandated as part of Resolution 22/03 to refine these guidelines and develop specific management responses. The IOTC Scientific Committee was requested to review and refine the ECP guidelines by 2024, and the Technical Committee on Management Procedures

(TCMP) to develop guidance on management responses by 2025. If exceptional circumstances are triggered, the pre-existing TAC remains in place until a new TAC or other management action is agreed upon by the Commission.

Table 1 summarizes the key elements of IOTC Resolution 22/03.

Table 1: Key Elements of IOTC Resolution 22/03 for Bigeye Tuna MP1 Harvest

Element	**Specific Detail (Reference: IOTC Res. 22/03 ,)**
Management Objectives	Maintain SB in Kobe green zone (not overfished, not subject to overfishing); Maximize average catch; Reduce inter-annual TAC variability. 60% probability of SB achieving SB$_{MSY}$ by 2034-2038. Avoid interim LRP (SB$_{20\%SB0}$) with high probability.
Harvest Control Rule (HCR)	MP1 Harvest: Pella-Tomlinson biomass dynamic model fitted to catch biomass and standardized longline CPUE. "Hockey stick" rule adjusting fishing mortality based on estimated biomass depletion (B_y/K). TAC changes capped at a maximum of 15% (increase or decrease) from the previous TAC.
Data Inputs	Annual total catches (1980-present). Standardized, spatially aggregated longline CPUE series (1980-present), based on Hoyle et al. (2019).
TAC Setting	IOTC Scientific Committee runs the MP and advises on the TAC. First TAC applied to 2024-2025. Subsequent TACs set for three-year blocks.
Review Schedule	Formal review of MP performance by the Commission and its subcommittees by 2030.
Exceptional Circumstances Protocol (ECP)	Guidelines adopted. Ongoing refinement of guidelines (SC by 2024) and development of management responses (TCMP by 2025). Pre-existing TAC remains if ECP triggered, until new TAC/action agreed.

2.2. The management strategy evaluation (MSE) process

The development of MP1 Harvest was underpinned by a comprehensive MSE process, designed to test robustness and performance across a range of plausible future scenarios.

2.2.1. Operating models (OMs)

The OMs are intended to represent the underlying dynamics of the bigeye tuna stock and fishery, incorporating key uncertainties. The OMs were based on the 2019 IOTC bigeye tuna stock assessment, a spatially structured Stock Synthesis (SS3) model. The OMs were conditioned using historical data on CPUE, size composition, and tagging information. Details on OM conditioning can be found in Kolody et al. (2020).

2.2.2. Uncertainty characterization

A critical component of MSE is the explicit incorporation of uncertainty; the IOTC BET MSE addressed this through several key "uncertainty axes". These axes represent major sources of uncertainty in the understanding of stock dynamics and data interpretation (Table 2). From over 400 potential combinations of these uncertainty levels, a reference set of 72 OMs was selected using a fractional factorial design (a statistical technique allows for an efficient exploration of the multi-dimensional uncertainty space without the computationally prohibitive task of running all possible OM combinations).

Table 2: Key Uncertainty Axes in the IOTC Bigeye Tuna MSE Operating Models

Uncertainty Axis	Brief Description
Stock-recruit steepness	Parameter (h) defining the relative recruitment level of the stock at low biomass levels; higher steepness implies faster recovery potential.
Natural mortality	The instantaneous rate of mortality due to all causes other than fishing; uncertainty in its absolute level and potential age-specific patterns.
Tag recapture weighting	The relative influence or weight assigned to tag return data in the stock assessment model, affecting estimates of mortality, growth, and movement.
CPUE catchability trends	Potential systematic changes over time in the efficiency of the fishing gear (e.g., longline) used to derive abundance indices (CPUE), which could bias abundance trends if not accounted for.
Regional CPUE scaling	Method used to combine or scale CPUE data from different geographical regions, particularly relevant for spatially structured stock assessment models.
Longline selectivity	The pattern of vulnerability of different age or size classes of bigeye tuna to capture by longline gear.
Size data effective sample sizes (ESS)	The statistical information content of length or age composition data; influences the precision of estimates derived from these data.

2.2.3. Candidate MPs

The MSE process evaluated nine candidate MPs. These spanned a range of complexities and approaches, including:

- Constant catch strategies
- CPUE-based (or empirical) MPs
- Model-based MPs (such as the adopted Pella-Tomlinson based MP1 Harvest).

2.2.4. Tuning objectives

Candidate MPs were tuned and evaluated against specific performance objectives defined by the IOTC's Technical Committee on Management Procedures (TCMP). A key objective was achieving a 0.6 or 0.7 probability of the stock being in the green zone of the Kobe plot (i.e., not overfished and not subject to overfishing) by the period 2034–2038.

2.2.5. Technical implementation

The computational framework for these simulations is hosted in the niMSE-IO-BET-YFT project repository on github, including R and C++ code used to simulate from operating models and evaluate

MPs.

3. Methodology of evaluation and review

3.1. Framework for evaluating the IOTC BET MP

The evaluation framework employed in this report is anchored in established MSE best practices. Key guidance documents informing this framework included the guide by Punt et al. (2016) titled "Management strategy evaluation: best practices", and discussion in scientific publications, such as Gutteridge et al. (2024), Lodge et al. (2007), and Nakatsuka (2017). In particular, recent revisions of the MSC fisheries standard, for example, directly refer to these best practices, and provide practical expectations associated with the development and implementation of harvest strategies (Gutteridge et al., 2024).

Seven core MSE best practice areas form the primary structure for this critical assessment. These are:

1. Clear Definition of Management Objectives and Quantitative Performance Metrics
2. Identification of System Knowledge and Key Uncertainties
3. Development of a Plausible Set of Operating Models (OMs)
4. Design and Simulation of Candidate Management Procedures (MPs) and Harvest Control Rules (HCRs)
5. Rigorous Simulation Testing and Iteration Across Operating Models
6. Transparent Presentation of Results, Trade-off Evaluation, and Stakeholder-Informed Decision-Making
7. Periodic Review and Adaptation

In addition, in line with the terms of reference for the present work, the MSE software and evaluation code were reviewed and key results reproduced. The code for the BET MSE was downloaded (cloned via git clone) and a test script developed with the help of the main software developer for the MSE (P. Jumppanen). The script was developed to reproduce the MSE for the accepted MP as closely as possible, in order to explore some potential issues identified with the MP detailed below. While exact reproducibility could not be achieved, a very close match to the documented MSE was achieved, and it is assumed that the results from this evaluation closely mirror those from the original MSE.

3.2. Literature review

The foundational document for this review is Kolody and Jumppanen (2021). The review process involved a systematic integration of findings and insights from a broad range of sources. These sources included:

- Official IOTC documents: Resolutions (especially Resolution 22/03), reports from the Scientific Committee (SC), Working Party on Tropical Tunas (WPTT), Working Party on Methods (WPM), and Technical Committee on Management Procedures (TCMP). These reports are helpfully catalogued by the authors of the MP and MSE in the corresponding github repository.

- Peer-reviewed scientific publications on MSE theory and application, stock assessment methodologies (e.g., Stock Synthesis, Pella-Tomlinson model), uncertainty analysis, stakeholder engagement in fisheries, and tuna biology and ecology.
- Reports and guidance documents from other RFMOs (e.g., IATTC, ICCAT, WCPFC) and international fisheries bodies (e.g., FAO).

A primary focus of the review was to critically examine and elaborate upon the potential shortcomings and further considerations.

4. Evaluation against MSE best practices

4.1. General observations

This section evaluates the IOTC bigeye tuna MP (MP1 Harvest) and its underlying MSE process against established MSE best practices, as well as technical issues identified during the review.

Overall, most decisions could be traced back to meeting decisions through appropriate referencing in the respective documents - for example, the decisions around management objectives. However, the broader implementation context was difficult to follow and evaluate from available documents. The development of the BET MP and MSE spanned six years, and decisions and rationale are found across a large number of meeting minutes and meeting papers. The overall process is therefore somewhat difficult to reconstruct, especially with respect to the stakeholder engagement and process for development of MP tuning objectives. For example, the TCMP minutes for the 2019 meeting suggest that an interactive tool built for WCPFC stakeholder engagement was used during the process, but details are vague, and it is not clear if this type of interactive exploration of MSEs was further pursued.

Recommendation 4.1.1. *A summary report of the MP development process, including a description of the nature and breadth of stakeholder engagement, and its interaction with technical elements of the MSE (possibly organised as part of a "retrospective" workshop on the process) would be a helpful exercise to capture lessons and possible improvements for future iterations of MSEs for BET, other IOTC stocks and MSEs in RFMOs in general (see also de Moor et al., 2022). Such "retrospectives" (not to be confused with retrospective patterns in derived quantities stock assessments) could form an integral part of the MSE process for any stock, and enable explicit linkages with MSEs from other stocks by exploring commonalities and differences.*

4.2. Definition of management objectives and quantitative performance metrics

A foundational step in any MSE is the collaborative definition of clear, quantifiable management objectives (biological, ecological, socio-economic) and associated performance metrics. These should include elements to be achieved, timeframes, and acceptable risk levels.

The IOTC's approach, particularly through Resolution 22/03, demonstrates a strong alignment with this best practice by outlining quantitative management objectives for MP1 Harvest. These include specific probabilistic targets for stock status (e.g., 60% probability of $SB_{2034-2038}$ achieving SB_{MSY}), defined timeframes, and considerations of risk (e.g., avoiding the LRP $SB_{20\%}$). The MSE also utilized performance metrics for catch, biomass, and TAC variability.

A technical inconsistency can be noted at this point; Resolution 22/03 specifically refers to a target of a 60% probability of $SB_{2034-2038}$ achieving SB_{MSY}, while all MSE documents refer to a target biomass and fishing mortality levels, aliased by referring to the probability of being the "the green quadrant" of the Kobe plot: "The stock status is in the Kobe green quadrant over the period 2030-2034 exactly 60% of the time (averaged over all simulations)" (Kolody & Jumppanen, 2021). Resolution 22/03 includes a superscript in the biomass objective that refers to appendix text about fishing mortality - however, this superscript being associated with a biomass reference point, it does not convey the intent that fishing mortality is an explicit target with the same weight as the stated biomass target. While the reference to the Kobe plot may be applying too much jargon, an equivalent formulation should include an explicit reference to fishing mortality.

Recommendation 4.2.1. *Ensure complete consistency in wording and meaning between MSE process and agreements to avoid confusion and potential inconsistencies between the MSE process and quantitative tuning objectives used in the process, and the communicated management objectives via the corresponding CMM.*

While socio-economic aspects like catch maximization and TAC stability are mentioned, the degree to which explicit socio-economic performance metrics formally traded-off against biological ones in the MSE could be more explicit (see Punt, 2017). MP1 Harvest aims to maximize average catch and reduce TAC variability (both having socio-economic implications), a more explicit framework for quantifying and systematically evaluating trade-offs between socio-economic objectives and biological objectives (e.g., stock status, risk to LRPs) would strengthen the MSE process.

Recommendation 4.2.2. *Trade-off (spider-) plots, and/or coloured decision tables are useful in visualising trade-offs between MPs and should be employed to provide simple visual aids to reading performance and trade-offs (Miller et al., 2019).*

In the context of the alternative management procedures presented and evaluated in Kolody and Jumppanen (2021), a quantitative target was only attached to probabilities of achieving biomass and fishing mortality targets, while catch stability was enforced by a 15% TAC change constraint, requested by the TCMP. This cap aims to provide stability for the fishing industry, a common socio-economic objective in MPs. However, such caps can also slow the MP's responsiveness to rapid changes in stock status if not carefully evaluated across a wide range of scenarios, potentially delaying necessary strong action if the stock declines quickly or preventing full utilization of a rapidly recovering stock.

In addition, catch stability acts as a limit constraint rather than a target. Although it was noted in the context of the application of the MP that the catch constraint was frequently encountered during MSE (IOTC SC, 2025), it may not have been obvious during MSE testing and related discussions that the MP acts mainly via catch changes at the imposed 15% limit, as shown in Figure 1. This means that catch stability is only achieved via the 15% constraint, and is not a feature of MP in terms of its functional form or reflected in tuning parameter values.

By presenting metrics around catch stability in largely relative terms, it is difficult to judge if a control rule leads to desirable behaviour in absolute (realised) terms — reporting the CV or variance of catch provides a useful indicator to compare MPs in relative terms, it may be too abstract to communicate behaviour and trade-offs within a given MP (Butterworth & Punt, 1999), and lead to consideration of trade-offs around mean catch and risk only. In the case of MPs explored for BET, most exhibited a behaviour of maximal catch variation (reductions or increases) for all but the

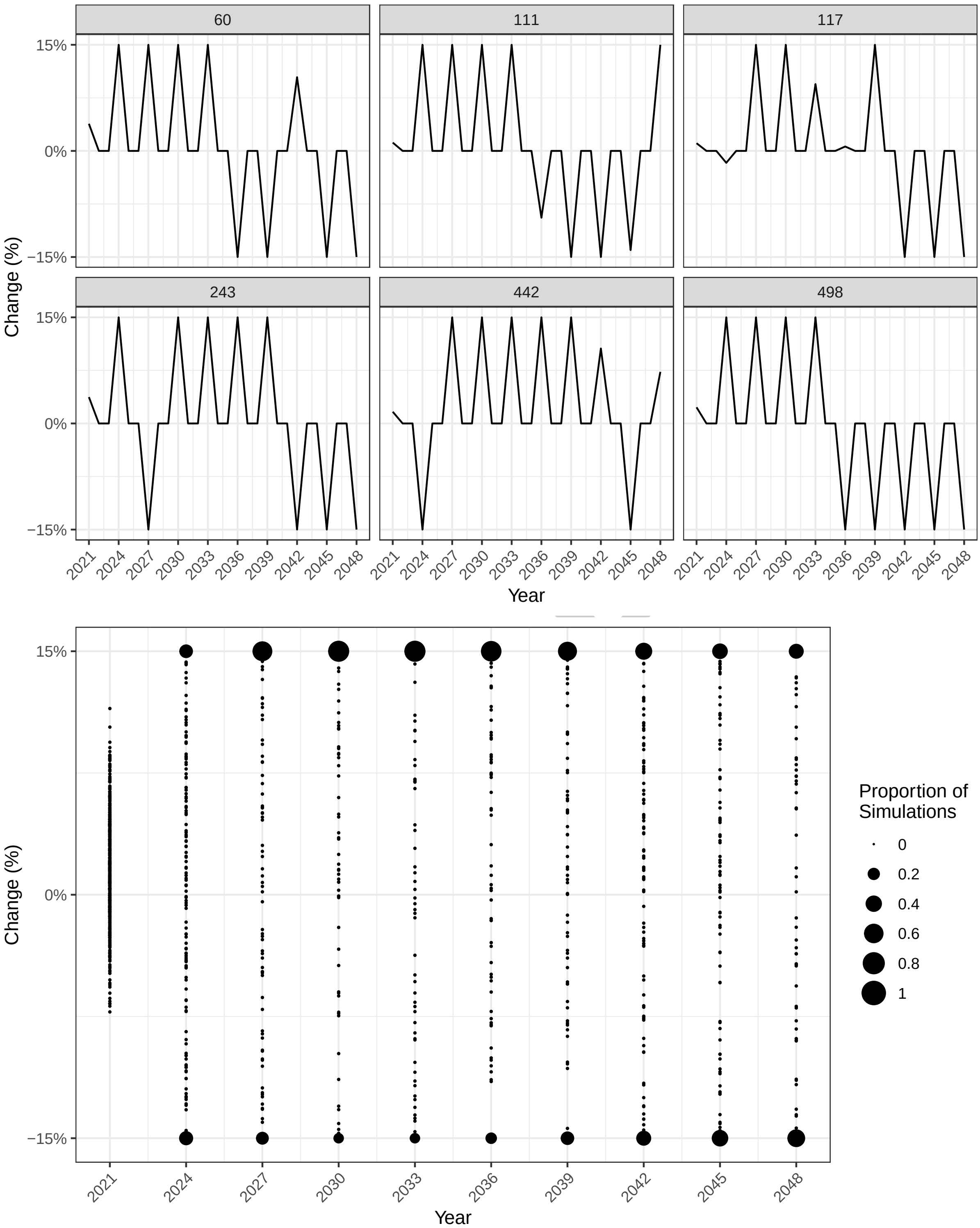

Figure 1: Simulated TAC change for a random draw of simulations under a reproduced MSE for MP1 Harvest for Bigeye tuna (top), and magnitude of simulated TAC changes across the MSE in terms of proportion of MSE simulations (bottom). Simulations were reproduced from provided code, but do not represent an exact replication of simulations used for original MSE advise (see text for detail).

first simulated TAC (Figure 1). This behaviour was not evident from the commonly used "annual average variation in yield" metric employed for the bigeye MSE because the rule is only applied every three years, and the interim variability in TACs is zero, which means the actual variation is likely understated by a factor of three (i.e., the actual average variability between subsequent TAC settings is close to the 15% limit). It is possible to include more explicit targets around stability in the tuning, as shown by Carruthers et al. (2016).

Recommendation 4.2.3. *A more thorough exploration of catch stability implications of tuning objectives in conjunction with alternative tuning settings (e.g., including targets for average variability in yield) may yield MPs that have more desirable behaviour in terms of catch stability, while maintaining desired over-all catch and risk levels.*

The behaviour of the MPs under consideration to drive maximal TAC change is likely in part due to the limited information contained within CPUE over short time-scales (see discussion in Kolody et al. (2020)) as well as the lag inherent in the management system. But it is also likely linked to an interplay between OM assumptions and management objectives used for tuning. The initial trends in the OM for the projection period are likely driven by a period of poor recruitment between 2009 and 2015, which is exhibited by all OM models, followed by a period of average recruitment for the last three years that was forced during OM conditioning (Kolody et al., 2020). Simulated recruitment then starts from an assumption of average recruitment (with noise; Figure 2). By focusing on near-term objectives only, the catch performance measures largely measure the ability of the MP to extract the increased production from the assumed influx of three years of average (i.e., better than all estimated between 2009 and 2015) recruitment. The relative performance of MPs is therefore likely to be a reflection of a single set of circumstances, defined by low estimated recruitment and average assumed recruitment for three years before the start of simulations — effectively assuming that these circumstances are close to reality.

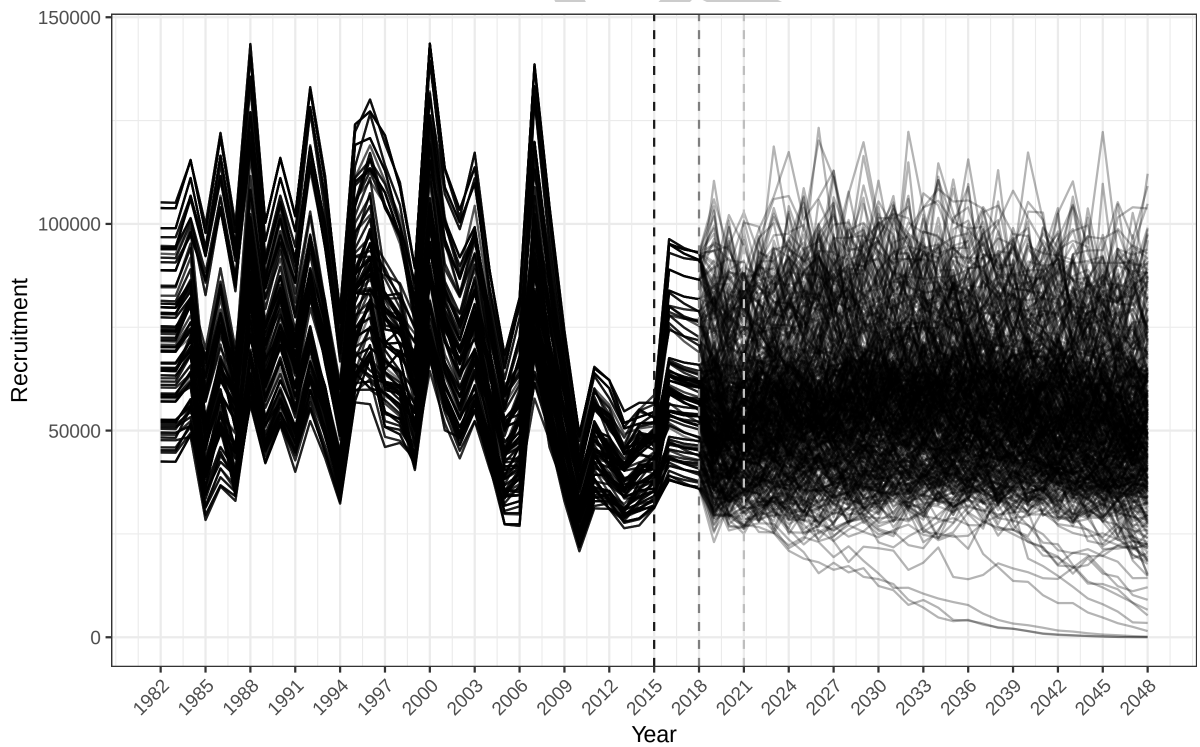

Figure 2: Estimated (before 2015), assumed (2016–2018), and simulated (after 2018) recruitment for IOTC bigeye tuna from the OM used to evaluate MP1 Harvest. Simulations were reproduced from provided code, but do not represent an exact replication of simulations used for original MSE advise (see text for detail).

A longer term evaluation of MP behaviours (and possibly longer-term tuning objectives) may have shown that the MPs tuned to the assumed reality do not perform well in other circumstances; for example, extending the simulations beyond 2040 shows that MP1 Harvest, while meeting the assigned management objectives and maximising short term catch from the assumed recruitment, ultimately leads to overfishing and an overfished stock (Figure 3). This suggests that the MP may be overly tuned to an assumed near-term reality, as opposed to providing desirable performance for managing the BET tuna stock across a range of alternative plausible realities, both in the near and long-term. Indeed, exploring longer-term tuning targets as part of this evaluation suggested progressively lower values for F_{mult}, the parameter that scales the HCR TAC response relative to the assessed status, because the MP settings become less dominated by short-term recruitment trends. Although MPs are to be re-evaluated periodically, a narrow focus on short- to medium-term behaviour has the potential to mask undesirable MP behaviour that may come to bear if assumptions about recent productivity turn out to have been incorrect. A cautionary example to this effect is given by de Moor et al. (2022) in the context of South African anchovy, where the MSE was dominated by a recent survey that was found to be biased in hindsight, leading to a suspension of the MP.

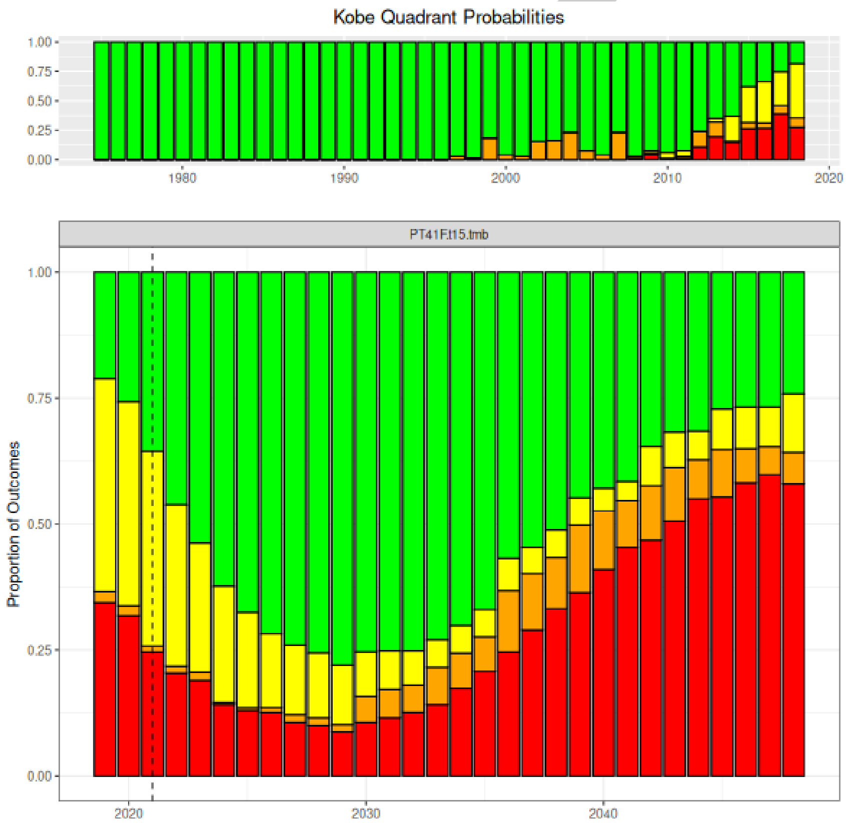

Figure 3: Proportion of simulations in each of the Kobe quadrants over time for MP1 Harvest. Historical estimates are included in the top panel. The lower panels are projections, with the first MP application indicated by the broken vertical line (2021). Captions reproduced from Kolody and Jumppanen (2021); simulations were reproduced from provided code, but do not represent an exact replication of simulations used for original MSE advise (see text for detail).

Recommendation 4.2.4. *Combining short-medium term targets such as the ones used for BET MSE with exploration and/or explicit evaluation of longer-term performance and management targets may yield important insights into the ability of a given MP to perform well across a range of scenarios.*

4.3. Identification of system knowledge and key uncertainties

MSE requires a thorough compilation of information about the fishery system and systematic identification of all major uncertainties (e.g., process, parameter, model, observation, assessment, implementation) to ensure a robust MP design (Punt et al., 2016).

The IOTC bigeye tuna MSE incorporated multiple axes of uncertainty in the testing of the bigeye MP, including those related to stock productivity (e.g., steepness, natural mortality), data interpretation (e.g., tag data weighting, CPUE trends, size data effective sample size), and fishery dynamics (e.g., selectivity). This demonstrates a strong effort to characterize structural uncertainty, which is crucial for developing robust MPs.

A potential issue with the MSE relates to the elimination of alternative CPUE series in the MSE. Kolody et al. (2020) noted " We eliminated the alternative CPUE standardization series. This decision was taken following recognition that the CPUE group did not produce the alternate CPUE series that were used in the previous MSE iteration. Furthermore, following discussion with the consultant for the CPUE group (Simon Hoyle, pers. comm.), it was decided that the CPUE uncertainty in the MSE is too important to simply use an ad hoc decision as was done in 2018. Accordingly, the decision was taken to only use the (implicitly "best") CPUE series from the assessment at this time. It has been recommended to the Secretariat and CPUE working group that the Terms of Reference for the CPUE group should be expanded in 2020 to include the provision of explicit MSE recommendations."

"CPUE uncertainty in the OM is retained in other dimensions (alternative regional scaling factors, catchability trends) and the alternative weighting of different data sources. The CPUE uncertainty introduced by the catchability trend is shown in Figure 4 and regional-scaling factors in Figure 5. The catchability trend is clearly very influential."

The MSE authors clearly attempted to retain data uncertainties. However, across the OM grid, the use of a single CPUE series likely led to nearly identical recruitment time series between models (aside from scaling differences; Figure 2), thereby exacerbating potential issues highlighted above with a strong reliance on recent estimated and assumed recruitment in the context of relatively near-term tuning targets. Alternative CPUE and/or catch time series can lead to differences in estimated recruitment trends and (potentially) productivity, which, in turn, influence both short- and long-term MP performance.

Recommendation 4.3.1. *Incorporating important data uncertainties may lead to a wider range of estimated recruitment trends and over-all productivity that may be as (or more) important for MP robustness and performance than parameter uncertainty. It is recommended that, for future iterations, MSE developers work with CPUE and catch data experts to explore alternative plausible time series to include in OM conditioning.*

The empirical basis for specific assumptions, like the 1% annual increase in longline catchability in 50% of OM runs, warrants ongoing scrutiny due to its significant impact on projections. Assumptions about trends in fishery catchability (effort creep) are highly influential in stock assessments and MSEs, yet are always difficult to estimate precisely. A fixed percentage increase, such as 1% per annum, may not adequately capture the potentially episodic or varying nature of technological advancements, learning behavior in fisheries, or changes in targeting practices. Given its acknowledged impact, this assumption should be subject to sensitivity analyses, and alternative models of catchability change (e.g., step-changes, non-linear trends) could be explored in robustness tests.

The explicit modeling of other implementation uncertainties (e.g., catch reporting inaccuracies, TAC overages) could be expanded. Implementation uncertainty refers to the difference between the intended management action (e.g., a TAC set by the MP) and what actually occurs in the fishery. This can arise from various sources, including incomplete catch reporting, misreporting, discarding practices, and TAC overages by some fleet segments. Failing to adequately model plausible levels of such implementation errors can lead to an overestimation of MP performance and an underestimation of risk to the stock. While the 15% TAC constraint in MP1 Harvest does address one element of implementation (by smoothing management responses), it does not account for potential deviations from the advised TAC.

Only TAC overages were explored during the BET MSE robustness trials, simulating robustness trials with 10% over-catch and various levels of over-catch reporting. In reality, implementation error may be strongly driven by abundance of the species in question (BET) as well as other targets in mixed fisheries. In fisheries where bigeye tuna are a bycatch, abundance-driven implementation error is especially likely, more so given the lag of the TAC setting with respect to most recent data used in MPs (3 years). This lag likely means that pronounced increases or declines in available biomass will lead to difficulties for certain sectors to remain within the TAC (if abundance is high relative to TAC) or with catching the TAC (if abundance is low).

Recommendation 4.3.2. *Implementation errors specifically linked to implementation lag and biomass trends could be explored in future iterations, possibly informed by observed catch relative to MP TACs from the first years of operating the MP.*

4.4. Development of a plausible set of operating models (OMs)

A cornerstone of MSE is the development of a set of OMs that represent plausible alternative hypotheses or "realities" of the fishery system. These OMs are ideally conditioned on historical data and encapsulate key uncertainties to robustly test MP performance, while maintaining a sufficiently narrow focus on plausible realities in order to not induce overly conservative MPs (de Moor et al., 2022). The IOTC BET MSE utilized a reference set of 72 OMs, based on the 2019 IOTC bigeye tuna stock assessment (a spatially structured Stock Synthesis model). These OMs were conditioned with historical data (CPUE, size composition, tagging data) and covered a fractional factorial design of identified uncertainties. Robustness tests were also conducted beyond this reference set.

The process for assigning plausibility weights to different OMs or uncertainty levels, beyond their inclusion in the reference set, could be further explored, as suggested by the MSE authors. While the MSE used a fractional factorial design to select a reference set of 72 OMs, assigning weighting (plausibility) to these OMs, or to the levels within each uncertainty axis, is a critical step that influences the interpretation of MP performance.

Best practices advocate for transparent and justifiable weighting schemes (Neubauer et al., 2023), which can be based on statistical fit to data, predictive performance, expert judgment, or stakeholder consensus. Initial exploration of the OM grid found that the catch likelihood might be able to distinguish between models that were able to take the assumed catch and those that couldn't, but this avenue of reasoning was later dropped as it became evident that it was largely related to maximum allowed fishing mortality (Kolody et al., 2020).

Without a clear articulation of how plausibility was assessed or assigned, the aggregate performance of an MP across the OM set might be biased by OMs that are, in reality, less likely. This is particularly important if certain OMs (e.g., those representing extreme states of nature) disproportionately drive the outcomes of the MP evaluations. In this context, one might argue that it is important to construct the OM grid (and adjust the weighting) in such a way that it represents an explicit prior distribution over assumptions, and that OM assumptions are weighted according to such a prior. In the context of productivity assumptions, for example, assigning equal weight to progressively lower estimates of natural mortality is problematic in the context of an OM as a lower M suggests lower productivity, and is probably not as likely as the base assumption. A one-sided departure from base assumptions, especially with equal weight, will likely make the OM unduly pessimistic or optimistic.

In the context of the BET MSE, the assumption of catchability increase over time was retained, and is clearly influential. It represents such a one-sided departure of base assumptions, and drives, to a large extent, the lower end of stock status estimates for the OM. It seems highly unlikely that the reference scenario of no catchability increase is true in reality. Equally, high levels on compounding increases in catchability are equally unlikely, as reflected by inclusion of the 3% catchability increase in the robustness trials. It would seem warranted to construct a similar prior over catchability increases, and use this prior in a consistent way in the OM (Figure 4; see Neubauer et al. (2023) for simulations and elaboration of this issue).

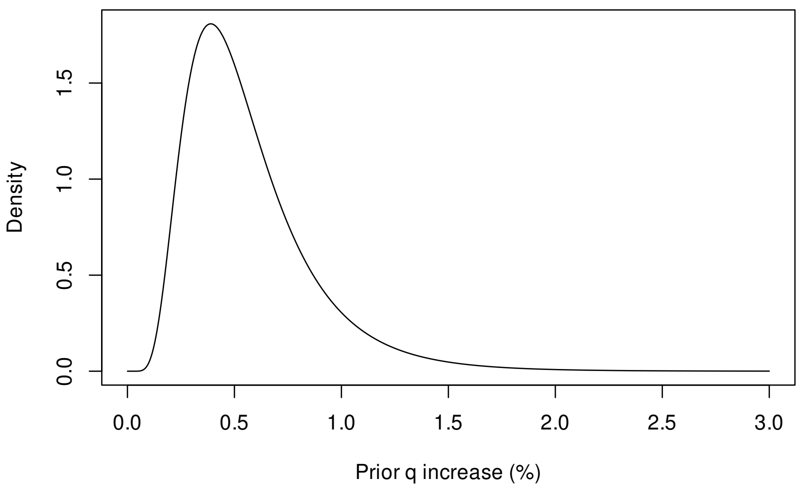

Figure 4: Illustration of a possible prior on catchability trends, used to weight OM axes.

Recommendation 4.4.1. *Where possible, weight OM axes according to explicit priors, to ensure that the MP performance is calculated from an ensemble of models weighted according to their plausibility. If explicit data-driven characterisation of uncertainty is possible (e.g., with estimated M), this uncertainty should be translated into the OM.*

4.5. Design and simulation of candidate management procedures (MPs)

Candidate MPs encompass data collection protocols, analysis/assessment methods, and HCRs, should be fully specified, realistic, implementable, and simulated accurately, including the potential effects of observation and implementation errors. The IOTC MSE process involved the development and testing of nine candidate MPs, which varied from simple (e.g., constant catch) to more complex model-based approaches. The adopted MP1 Harvest features a clearly specified HCR (based on a Pella-Tomlinson model with a hockey-stick adjustment rule), defined data inputs (total catch and standardized longline CPUE), and a TAC change constraint (15% cap).

A consideration is the structural mismatch between the complex, age-structured OMs (Stock Synthesis) and the simpler biomass dynamic model within MP1 Harvest (Pella-Tomlinson). While simpler MP assessment models are often preferred for operational reasons (e.g., lower data requirements, ease of implementation, transparency), their ability to correctly interpret signals of stock status from a complex underlying reality (as represented by the more complex OMs) is a key vulnerability. One such issue may be that the current MP only uses current status to estimate the multiplier on fishing mortality, but does not incorporate recent recruitment trends. The MP therefore cannot deal with expected trends in the forthcoming years, especially in the lag years between data availability and TACs coming into effect.

The alternative MP option presented to the IOTC commission used a projection-based MP that effectively incorporates recent process error (i.e., recent productivity deviations) into TAC recommendations by finding the TAC that reaches a target status. It was found that this MP had lower levels of catch variation (Kolody & Jumppanen, 2021), and did not lead to overfishing in the last evaluation years. A more thorough exploration of this projection-based method, and it's potential advantages over the more "retrospective" MPs such as MP1 Harvest would have been useful (acknowledging that the projection-based MP was a last minute addition and there was little time to fully explore its advantages).

Recommendation 4.5.1. *A more thorough exploration of projection-based MPs is desirable given its ability to use recent process error (productivity deviations) to set TACs, which may provide better performance given the current 3-year lag between data availability and TAC setting.*

The reliable operation of MP1 Harvest, with its specific requirements for annual total catch data and a standardized longline CPUE series dating back to 1980, necessitates a clearly defined, resourced, and committed long-term data collection program. The long-term success of any data-driven MP hinges on the stability, quality, and consistency of its input data streams. Within RFMOs like the IOTC, ensuring such consistency across all member fleets and over extended periods can be challenging due to varying national capacities, reporting standards, and resource availability. Any degradation of the quality or availability of these crucial data series could compromise the performance of MP1 Harvest. Thus, the robustness of the MP is intrinsically linked to the robustness and sustainability of the underlying data collection and processing systems.

4.6. Simulation testing

Candidate MPs must be rigorously tested by simulating their application across each OM over many stochastic replicates and a sufficient projection period. The MSE process should be iterative, allowing for refinement of MPs, OMs, or objectives based on initial results. The evaluation of nine candidate MPs across the reference set of 72 OMs suggests a significant simulation effort within the IOTC BET MSE. Meeting notes for the IOTC WPM and SC indicate an iterative process of development and refinement of the MPs and OMs.

Robustness trials covered a range of scenarios, from data (CPUE CV), implementation error, process error (catchability increase, recruitment shock) to software implementation. Beyond statements above about implementation uncertainties and the importance of recent recruitment, it may be worth explore, within simulations, the ability of exceptional circumstance criteria to reduce the likelihood of poor outcomes under such exceptional circumstances (e.g., if recruitment failure occurs between stock assessments, the MP does not explicitly model recruitment and this failure may not be evident and trigger an exceptional circumstance until the assessment is run). Criteria related to the estimated process error for the MP's Pella-Tomlinson model, for example, may be able to provide guidance on useful quantities to monitor in this context (see also more specific discussion below regarding exceptional circumstances). Such metrics might also provide insights into the potential for directional change in productivity.

Recommendation 4.6.1. *If the scenarios in the robustness trials are considered plausible, then a reasonable action would be to develop indicators that can provide monitoring with regards to key future uncertainties. The development of such indicators could be considered as part of future MSEs.*

Long-term, climate impacts will affect all oceans, and corresponding changes to marine ecosystems are inevitable. As scientifically plausible scenarios emerge about the future of tuna populations in a changing climate, these will need to be incorporated in simulations (Punt et al., 2016). A question then arises whether such scenarios are to be treated as robustness or plausible operating models. In addition, MPs may need to be evaluated in the context of dynamic reference points in this context as assumptions about static reference points will become increasingly unrealistic, and references to unfished biomass levels under changing productivity may need to be incorporated (Punt, 2017). However, regardless of how climate scenarios are treated, they will undoubtedly be relevant in developing and choosing future MPs.

Recommendation 4.6.2. *Develop plausible climate and/or ecosystem change scenarios that can inform about the robustness of MPs in a (directionally) changing environment.*

4.7. Presentation of results and trade-off evaluation

MSE results must be presented transparently and in an understandable manner to decision-makers and stakeholders, facilitating the evaluation of trade-offs between competing objectives (Miller et al., 2019). The final selection of an MP is a management decision informed by this stakeholder consultation and dialogue. MSE results for the IOTC bigeye tuna MP were generally appropriately presented and discussed within IOTC scientific and management advisory bodies (WPM, SC, TCMP). The adoption of Resolution 22/03 indicates a decision based on these MSE outputs, and the resolution acknowledges the trade-offs involved (e.g., balancing stock health, catch levels, and TAC stability).

There is a need for continued efforts in developing effective communication strategies for complex MSE results, assumptions (e.g., OM pessimism), and trade-offs to non-expert audiences. MSE outputs are complex and probabilistic, making them difficult for non-specialists to interpret. Best practices in science communication for MSE involves using clear language, effective visualizations (e.g., trade-off plots, risk-reward graphics), and interactive tools to help stakeholders understand the implications of different MP choices (Miller et al., 2019). This is not merely about better figures, but about fostering a shared understanding of the MSE process and its outcomes.

Recommendation 4.7.1. *Develop broadly accessible interactive tools that allow stakeholders to develop an intuitive understanding of key trade-offs between key quantities across a range of MPs.*

Ensuring broad stakeholder consensus on the specific probability levels chosen for objectives is important for the legitimacy of the MP. These probability levels directly reflect the acceptable level of risk associated with achieving or failing to achieve management targets. The selection of these levels should ideally emerge from a consultative process where stakeholders understand the implications of different risk tolerances. Without such consensus, the chosen probability levels might be perceived as arbitrary or as unduly favouring certain interests, potentially undermining support for the MP. As mentioned above, the process by which the TCMP arrived at the management objectives is currently not transparent, and could be made explicit in a retrospective report as outlined above.

4.8. Periodic review and adaptation

MSE is not a static, one-time exercise. The implemented MP and the MSE framework itself (including OMs and assumptions) should undergo periodic review (e.g., every 3-5 years or as triggered by specific events). An Exceptional Circumstances Protocol (ECP) should be in place and well-defined to address unforeseen situations or significant MP under-performance. IOTC Resolution 22/03 schedules a comprehensive review of MP1 Harvest's performance by 2030. The resolution also mandates the refinement of ECP guidelines and the development of management responses to exceptional circumstances. The IOTC WPM has also discussed criteria for OM update frequency.

The effectiveness of the ECP will depend on the specificity, objectivity, and pre-tested nature of its triggers. An ECP is designed to act as a safety net when the fishery system deviates significantly from what was simulated in the MSE, or when the MP is clearly failing to meet the objectives laid out at its implementation (de Moor et al., 2022; Preece et al., 2021). Generic or vaguely defined ECP triggers may fail to activate when needed or may activate inappropriately due to normal system variability, leading to unnecessary interventions or, conversely, a failure to act in a true crisis.

The current ECP guidelines referenced in IOTC SC (2021) Appendix 6A define EC as "... circumstances (primarily related to future monitoring data falling outside the range covered by management strategy evaluation (MSE) simulation testing) where overriding of the output from a Management Procedure should be considered...". Their initial definition is deliberately loose "...so that a range of unanticipated issues can be addressed" (Preece et al., 2021). With regard to BET CMM 22/03 specifies that (§12)"The Scientific Committee is requested to review, and if necessary, further develop and refine (not later than 2024), the exceptional circumstances guidelines (adopted by SC24 and S26), taking into account, inter alia, the need for an appropriate balance between specificity versus flexibility in defining exceptional circumstances, and the appropriate level of robustness to ensure that exceptional circumstances are triggered only when necessary." and (§13) "The IOTC, through the Technical Committee on Management Procedures, is requested to review the need for, and if necessary, develop at latest by 2025, guidance on a range of appropriate management responses should those exceptional circumstances be found to occur."

Currently, EC evaluation is focused on recent trends (stock status or CPUE in relation to 90% confidence bounds of simulated outcomes under the MSE), alongside reported changes in catch reporting and understanding of the stock biology. The latter have been evaluated qualitatively, while the former are quantitative evaluations (Preece & Williams, 2024, 2025).

With respect to changes in population dynamics and biology, these could occur both temporally, or in terms of new evidence that alters the scientific understanding of productivity. While productivity parameters like natural mortality (M), growth, and steepness are not parameters considered in the MP (as pointed out by Preece & Williams, 2024), these parameters will define the productivity in the OM grid used to tune and evaluate the MP. A significant change, either temporally or in terms of new evidence towards alternative productivity assumptions, towards values beyond those included in the OM, may be akin to observing a significant departure from simulated values for CPUE or stock status: the MP was tuned and evaluated for a set of OM productivity assumptions. If these are no longer assumed to be correct, the suitability of the MP may also not be given anymore, especially if productivity is revised downward.

Recommendation 4.8.1. *It is important both ensure that the OM includes a broad-enough set of assumptions around plausible levels of productivity, and to monitor productivity changes explicitly with respect to levels of productivity assumed in the OM. For directly applicable parameters, this may be achieved by comparing any new productivity values to those explored in the OM grid, and/or by exploring the impact on the assumed production function, both within the OM and the MP assessment model.*

The idea of monitoring the production function can be extended to monitoring of observable quantities from the simulations, beyond the currently employed percentile bounds on simulation time series of stock status and CPUE. For instance, monitoring the production function of the Pella-Tomlinson state-space surplus production model applied as part of the MP may highlight temporal trends in production, which can be assessed against ranges of productivity estimated during simulations (for which productivity is assumed constant within the OM grid). Other observable quantities, such as the number of sequential declines in biomass and catch witnessed during the application relative to simulations can inform if the MP is working as intended. Deviations from simulated distributions (e.g., observing three successive declines in catch when 90% of simulations of the MP reversed stock declines and associated TAC reductions within two MP cycles) would indicate that the MP is not functioning as expected, even if the over-all stock and CPUE values might be within the 90% bounds of simulations.

Recommendation 4.8.2. *Develop a more refined set of automated indicators that inform about potential consistent changes in productivity and deviations of MP performance relative to simulated outcomes.*

Developing effective ECP triggers and appropriate management responses may benefit from its own dedicated simulation testing—essentially an "MSE for the ECP"—to ensure they are sensitive to genuine exceptional circumstances but robust to noise. This involves defining precisely what constitutes an "exceptional circumstance" (e.g., recruitment failure over multiple years, dramatic unexplained changes in CPUE, significant changes in key biological parameters) and testing the performance of various triggers across the OMs.

Recommendation 4.8.3. *Test the performance of ECP indicators and triggers using the MSE simulations for future MSEs.*

Input data are a critical component of the MP, yet diagnostics to ensure consistency in CPUE especially are currently only sparsely available for scrutiny. Provided that CPUE inputs are driving the MP and associated TACs, the inconsistency in CPUE analyses used to inform the MP process, relative to those specified by Williams et al. (2022) is concerning. This was acknowledged by the 1st Special Session of the IOTC Scientific Committee, the SC recommended that "a fixed set of CPUE standardization code is developed for each MP to ensure that it is developed following the specifications of the MP". While this suggestion provides a first step, a more precise specification of the analysis is probably worthwhile.

Recommendation 4.8.4. *To ensure reproducibility, code and APIs to provide data should*

1. *Include a suite of tests that ensure consistent data input formats over time, including expectations around field types and contents (e.g., number of total vessels in the provided data cannot decline year-on-year).*

2. *Be containerised using tools such as Docker containers. This ensures that no unintended software changes will influence analysis outcomes (e.g., changes in packages that could transform variables during reading operations from numeric to factor levels).*

3. *Be version controlled to ensure that all changes are documented, and can be audited to ensure long-term traceability of key decisions and changes.*

In addition to the CPUE code, some standardised tools should be developed to critically assess the utility of CPUE as an index of abundance on an ongoing basis. For this, tools such as influence plots and coefficient-distribution-influence (CDI) plots (Bentley et al., 2012) are invaluable tools. In addition, spatio-temporal aspects, such as over-all effort distribution in space over time should be summarised at an acceptable level to allow scrutiny of changes and their potential impact on CPUE.

Recommendation 4.8.5. *A consistent set of diagnostics and criteria for judging the adequacy of the CPUE for use in the MP should be developed as a priority, and included in evaluations of ECs.*

4.9. Software implementation and reproducibility

Two aspects of the implementation can be distinguished. First, the software tool itself (niMSE), and the use of the tool to evaluate the MPs. niMSE was cloned from github, and, with some help from the developer, was reasonably straightforward to install and use. The niMSE code is well organised in an object-oriented structure, and easy to follow and test. A script was developed on the basis of supplied code as well as code archived for earlier iterations of the MSE, which provided guidance about the use of the code. Overall the MSE software appears fit-for-purpose, relatively easy to generalise (beyond BET and YFT). The software code is version-controlled on github, allowing for open access of the tools.

The currently archived code allows for relatively straightforward emulation of the MSE. This took only a couple of days to get to a reasonable point where it was possible to extract useful information and run alternative ideas through the software; the script can be provided on request. However, exact reproducibility of the MSE is currently not given as i) the most recent analyses appear not to be fully archived ii) software changes since the original MSE require a few changes in the code, and iii) consistency of the runtime environment is not given as the analysis is not containerised and scripted end-to-end. Such a reproducible pipeline for analysis is an important step towards ensuring continuity in analyses and more precise scrutiny of outputs by peer-review (Middleton et al., in press).

Recommendation 4.9.1. *Future iterations of the MSE should ensure full reproducibility by using containers and version-controlled build scripts that produce all necessary assets for the MSE reports and tools.*

4.10. Summary of recommendations

Table 3: Summary of IO BET MP and MSE Alignment with Best Practices

MSE Best Practice	Recommendation
General observations	**Recommendation 4.1.1.** *A summary report of the MP development process, including a description of the nature and breadth of stakeholder engagement, and its interaction with technical elements of the MSE (possibly organised as part of a "retrospective" workshop on the process) would be a helpful exercise to capture lessons and possible improvements for future iterations of MSEs for BET, other IOTC stocks and MSEs in RFMOs in general (see also de Moor et al., 2022). Such "retrospectives" (not to be confused with retrospective patterns in derived quantities stock assessments) could form an integral part of the MSE process for any stock, and enable explicit linkages with MSEs from other stocks by exploring commonalities and differences.*
Definition of management objectives and quantitative performance metrics	**Recommendation 4.2.1.** *Ensure complete consistency in wording and meaning between MSE process and agreements to avoid confusion and potential inconsistencies between the MSE process and quantitative tuning objectives used in the process, and the communicated management objectives via the corresponding CMM.*
	Recommendation 4.2.2. *Trade-off (spider-) plots, and/or coloured decision tables are useful in visualising trade-offs between MPs and should be employed to provide simple visual aids to reading performance and trade-offs (Miller et al., 2019).*
	Recommendation 4.2.3. *A more thorough exploration of catch stability implications of tuning objectives in conjunction with alternative tuning settings (e.g., including targets for average variability in yield) may yield MPs that have more desirable behaviour in terms of catch stability, while maintaining desired over-all catch and risk levels.*

Continued on next page

Table 3 – continued from previous page

MSE Best Practice	Recommendation
	Recommendation 4.2.4. *Combining short-medium term targets such as the ones used for BET MSE with exploration and/or explicit evaluation of longer-term performance and management targets may yield important insights into the ability of a given MP to perform well across a range of scenarios.*
Identification of system knowledge and key uncertainties	**Recommendation 4.3.1.** *Incorporating important data uncertainties may lead to a wider range of estimated recruitment trends and over-all productivity that may be as (or more) important for MP robustness and performance than parameter uncertainty. It is recommended that, for future iterations, MSE developers work with CPUE and catch data experts to explore alternative plausible time series to include in OM conditioning.* **Recommendation 4.3.2.** *Implementation errors specifically linked to implementation lag and biomass trends could be explored in future iterations, possibly informed by observed catch relative to MP TACs from the first years of operating the MP.*
Development of a plausible set of operating models (OMs)	**Recommendation 4.4.1.** *Where possible, weight OM axes according to explicit priors, to ensure that the MP performance is calculated from an ensemble of models weighted according to their plausibility. If explicit data-driven characterisation of uncertainty is possible (e.g., with estimated M), this uncertainty should be translated into the OM.*
Design and simulation of candidate management procedures (MPs)	**Recommendation 4.5.1.** *A more thorough exploration of projection-based MPs is desirable given its ability to use recent process error (productivity deviations) to set TACs, which may provide better performance given the current 3-year lag between data availability and TAC setting.*

Continued on next page

Table 3 – continued from previous page

MSE Best Practice	Recommendation
Simulation testing	**Recommendation 4.6.1.** *If the scenarios in the robustness trials are considered plausible, then a reasonable action would be to develop indicators that can provide monitoring with regards to key future uncertainties. The development of such indicators could be considered as part of future MSEs.* **Recommendation 4.6.2.** *Develop plausible climate and/or ecosystem change scenarios that can inform about the robustness of MPs in a (directionally) changing environment.*
Simulation testing	**Recommendation 4.6.1.** *If the scenarios in the robustness trials are considered plausible, then a reasonable action would be to develop indicators that can provide monitoring with regards to key future uncertainties. The development of such indicators could be considered as part of future MSEs.*
Presentation of results and trade-off evaluation	**Recommendation 4.7.1.** *Develop broadly accessible interactive tools that allow stakeholders to develop an intuitive understanding of key trade-offs between key quantities across a range of MPs.*
Periodic review and adaptation	**Recommendation 4.8.1.** *It is important both ensure that the OM includes a broad-enough set of assumptions around plausible levels of productivity, and to monitor productivity changes explicitly with respect to levels of productivity assumed in the OM. For directly applicable parameters, this may be achieved by comparing any new productivity values to those explored in the OM grid, and/or by exploring the impact on the assumed production function, both within the OM and the MP assessment model.* **Recommendation 4.8.2.** *Develop a more refined set of automated indicators that inform about potential consistent changes in productivity and deviations of MP performance relative to simulated outcomes.*

Continued on next page

Table 3 – continued from previous page

MSE Best Practice	Recommendation
	Recommendation 4.8.3. *Test the performance of ECP indicators and triggers using the MSE simulations for future MSEs.* **Recommendation 4.8.4.** *To ensure reproducibility, code and APIs to provide data should* 1. *Include a suite of tests that ensure consistent data input formats over time, including expectations around field types and contents (e.g., number of total vessels in the provided data cannot decline year-on-year).* 2. *Be containerised using tools such as Docker containers. This ensures that no unintended software changes will influence analysis outcomes (e.g., changes in packages that could transform variables during reading operations from numeric to factor levels).* 3. *Be version controlled to ensure that all changes are documented, and can be audited to ensure long-term traceability of key decisions and changes.* **Recommendation 4.8.5.** *A consistent set of diagnostics and criteria for judging the adequacy of the CPUE for use in the MP should be developed as a priority, and included in evaluations of ECs.*
Software implementation and reproducibility	**Recommendation 4.9.1.** *Future iterations of the MSE should ensure full reproducibility by using containers and version-controlled build scripts that produce all necessary assets for the MSE reports and tools.*

5. Conclusion

The adoption of MP1 Harvest for Indian Ocean bigeye tuna represents a significant advancement in the fisheries management of IOTC's BET stock. This structured framework moves away from reactive management towards a pre-agreed management procedure and decision rule. Such a framework should allow for more certainty and stability in the resource long term.

The review found that MP1 Harvest and the associated MSE process align with many best practices. The MP is explicitly linked to quantitative, probabilistic management objectives related to stock status and overfishing. This allows for explicit tuning of the MP's performance against these pre-agreed benchmarks. MP1 Harvest includes a clearly specified HCR, based on a Pella-Tomlinson model and a hockey-stick rule, which dictates how TACs are adjusted in response to changes in estimated stock status. A strength of the underlying MSE for MP1 Harvest is the identification and incorporation of a number of biological and fishery-related uncertainties into a set of operating models, as opposed to relying on a narrowly defined operating model. This explicit consideration of uncertainty is fundamental to designing MPs that are robust to incomplete knowledge and natural variability. The MSE then tested multiple candidate procedures and their robustness to a number of scenarios, in alignment with documented best practices. The management framework includes provisions for periodic review (by 2030) and an Exceptional Circumstances Protocol (ECP), demonstrating a commitment to adaptive management that allows the system to respond to new information or unforeseen developments.

While MP1 Harvest represents a significant advancement for IOTC, incorporating quantitative objectives and explicit uncertainty consideration, several areas may warrant future investigation and further development to ensure long-term sustainability and performance of the BET MP. These include improving the characterization and communication of complex uncertainties, such as OM plausibility; implementing realistic levels of implementation error and incorporating explicit data uncertainties, and ensuring the long-term institutional commitment to high-quality data collection and standardised analysis programs that are fundamental to the MP's operation. Developing a range of objective, and ideally simulation-tested indicators for the Exceptional Circumstances Protocol will enhance the robustness of the MP.

The report expands on these findings and provides evidence-based recommendations for potential future enhancements to the MP framework to ensure it aligns with recognised best practice and matures towards a robust long-term management tool for BET in the Indian Ocean. Many of the recommendations in this review represent potential refinements, rather than fundamental issues with the MP or MSE that would warrant immediate re-evaluation or changes. Nevertheless, the suggestions may contribute to increase long-term performance and robustness of the BET MP.

References

Bentley, N., Kendrick, T. H., Starr, P. J., & Breen, P. A. (2012). Influence plots and metrics: Tools for better understanding fisheries catch-per-unit-effort standardizations. *ICES Journal of Marine Science*, *69*(1), 84–88.

Butterworth, D. (2007). Why a management procedure approach? some positives and negatives. *ICES Journal of Marine Science*, *64*(4), 613–617.

Butterworth, D., & Punt, A. (1999). Experiences in the evaluation and implementation of management procedures. *ICES Journal of Marine Science*, *56*(6), 985–998.

Carruthers, T. R., Kell, L. T., Butterworth, D. D. S., Maunder, M. N., Geromont, H. F., Walters, C., McAllister, M. K., Hillary, R., Levontin, P., Kitakado, T., & Davies, C. R. (2016). Performance review of simple management procedures. *ICES Journal of Marine Science*, *73*(2), 464–482.

de Moor, C. L., Butterworth, D. S., & Johnston, S. (2022). Learning from three decades of management strategy evaluation in south africa (E. Jardim, Ed.). *ICES Journal of Marine Science*, *79*(6), 1843–1852.

Gutteridge, A., Melnychuk, M., Longo, C., Currey, R., & Jardim, E. (2024). The msc approach for developing harvest strategies and harvest control rules for regional fisheries management organisations (rfmos). *Marine Policy*, *161*, 106004.

IOTC SC. (2021). Report of the 24th session of the IOTC scientific committee. *IOTC-2021-SC24-RE_Rev1*.

IOTC SC. (2025). Report of the 1st special session of the IOTC scientific committee. *IOTC-2025-SSC01-R*.

Kolody, D., & Jumppanen, P. (2021). Indian ocean bigeye tuna management procedure evaluation update march 2021. *IOTC-2021-PM12(MSE)-04*.

Kolody, D., Jumppanen, P., & Day, J. (2020). Indian ocean bigeye tuna management procedure evaluation update march 2020. *IOTC MSE Task Force*.

Lodge, M. W., Anderson, D., Willock, A., & Royal Institute of International Affairs (Eds.). (2007). *Recommended best practices for regional fisheries management organizations: Report of an independent panel to develop a model for improved governance by regional fisheries management organizations*. Chatham House.

Middleton, D. A. J., Thompson, F. N., Langley, A. D., & Neubauer, P. (in press). Implementing reproducible fisheries research: A decade of experience with the kahawai reporting system. *Journal of the Royal Society of New Zealand*.

Miller, S. K., Anganuzzi, A., Butterworth, D. S., Davies, C. R., Donovan, G. P., Nickson, A., Rademeyer, R. A., & Restrepo, V. (2019). Improving communication: The key to more effective MSE processes. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(4), 643–656. https://doi.org/10.1139/cjfas-2018-0134

Nakatsuka, S. (2017). Management strategy evaluation in regional fisheries management organizations — how to promote robust fisheries management in international settings. *Fisheries Research*, *187*, 127–138. https://doi.org/10.1016/j.fishres.2016.11.018

Neubauer, P., Kim, K., A'mar, T., & Large, K. (2023). Addressing uncertainty in wcpfc stock assessments: Review and recommendations from wcpfc project 113. *WCPFC-SC19-2023/SA-WP-12*.

Preece, A., & Williams, A. (2024). Consideration of exceptional circumstances for the bigeye tuna MP 2024. *IOTC-2024-WPM15-10*.

Preece, A., & Williams, A. (2025). An update on consideration of exceptional circumstances for the bigeye tuna MP 2025. *OTC WPM_MSE 2025*.

Preece, A., Williams, A., & Hillary, R. (2021). Consideration of exceptional circumstances processes in the IOTC. *IOTC-2021-WPM12-17*.

Punt, A. E. (2017). Strategic management decision-making in a complex world: Quantifying, understanding, and using trade-offs. *ICES Journal of Marine Science*, *74*(2), 499–510. https://doi.org/10.1093/icesjms/fsv193

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, *17*(2), 303–334. https://doi.org/10.1111/faf.12088

Williams, A., Preece, A., & Hillary, R. (2022). Specifications of the IOTC bigeye tuna management procedure. *IOTC-2022-WPM13-11_Rev1*. https://iotc.org/sites/default/files/documents/2022/10/IOTC-2022-WPM13-11_Rev1.pdf

A. Terms of Reference

The Indian Ocean Tuna Commission is an intergovernmental organization established under article XIV of the FAO constitution. The IOTC is mandated to manage 16 tuna and tuna-like species in the Indian Ocean with its primary objective the conservation and optimum utilization of the stocks for long-term sustainability. The scientific advice on the status of major IOTC tuna stocks are based upon the results of quantitative stock assessment models and analyses of available information. Stock assessment is critical to enhance scientific elements in the conservation and sustainable exploitation of these valuable fishery resources.

IOTC Resolution 15/10 requested the IOTC Scientific Committee to develop and assess, through the management strategy evaluation (MSE) process, the performance of Management Procedures (MP) and Harvest Control Rules (HCRs), to achieve Target Reference Points (TRPs) on average and avoid the Limit Reference Points (LRPs) with a high probability taking into account the levels of uncertainty in the stock assessments for the priority species listed as Skipjack, Bigeye tuna (BET), Yellowfin tuna, Albacore tuna, and swordfish.

The Working Party on Methods (WPM), in conjunction with other Working Parties, has been developing the MSE for different IOTC species and has identified the need for external peer-review of the MSE process. As such, the 2021 MSE task force of the WPM meeting discussed the merits of external peer-review and agreed that external review at both the technical (e.g. code inspection) and process can greatly benefit the MSE. The MSE task force of the WPM suggested that the external review could be conducted independently and in parallel to the TCMP process, preferably to species for which the MP evaluation is close to completion (such as BET), and when resources and expertise are available.

Moreover, the 2021 WP on Methods identified as a priority in its workplan the need of an external peer review based on Terms of Reference agreed to by the WPM and the Scientific Committee and following the schedule recommended in its workplan (Appendix VI of IOTC-2021-WPM12(MSE)-04). The WPM also identified as a research priority in its workplan the external peer-review for Bigeye MSE to be completed by 2023. The WPM also noted that clear instructions and terms of references should be provided for the external review given the complexity of the MSE process. Moreover, the WPM highlighted that arranging an external peer review will not be trivial due to the complexity of the process, the limited number of experts able to conduct the work as well as the cost and, thus, the WPM suggested that the scope should be agreed and terms of references developed.

The absence of an independent review should not preclude the Commission from the initial adoption of a Resolution for an MP. As part of the MP resolution, the review of the performance of the MP (e.g. after 6-9 years of implementation) by the SC and TCMP should also be specified. To this end, a short assignment is required to conduct the review of the bigeye MSE operating model (OM) and the MPs evaluated through MSE Based on the review work provide recommendations for improving the bigeye MSE, including the operating model, harvest control rules and management procedures.

The key areas for consideration by the peer review of the Indian Ocean Bigeye Tuna management strategy evaluation (IOTC-2021-WPM12(MSE)-04) are listed below:

- Review the operating model conditioning and the range of uncertainties included in reference and robustness sets

- Review the evaluation of the adopted Management Procedure

- Review the robustness scenarios tested through the MSE
- Review the exceptional circumstances guidelines endorsed by the Commission
- Advise on adequacy of communication in the reports of MSE results, the trade-offs between various management procedures, and the ranking of management procedures.
- Provide recommendations for future developments to improve the bigeye MSE and MP, if applicable.

It is envisaged that the external expert will review the bigeye MSE documentation and be able to replicate the results from the BET MP code. To carry out the peer-review the external expert will:

- work with the developer of the bigeye MSE, the WPTT, the SC Chair, IOTC Secretariat and IOTC Stock Assessment expert, to review the OM and performance of candidate MPs. The contractor will also suggest ways to improve the current MSE framework.
- review the process followed to adopt a MP, by reviewing relevant documents, including SC documents, reports from SC or Commission meetings, IOTC Recommendations, etc. and to provide comments on the appropriateness of the selection (or omission) of OMs and MPs considered, as well as on the involvement of scientists and/or stakeholders in the process.
- document the quality control procedures followed by the reviewer to test the code, the outcomes of the review process, and the comments on the MSE process, with suggestions for the future work of the SC.
- participate in the WPM meeting other relevant IOTC MSE meetings and discussions where feasible.