Received: 18 February 2020 | Revised: 12 May 2020 | Accepted: 13 May 2020
DOI: 10.1111/faf.12480

ORIGINAL ARTICLE

FISH and FISHERIES WILEY

Operating model design in tuna Regional Fishery Management Organizations: Current practice, issues and implications

Rishi Sharma[1] | **Polina Levontin**[2] | **Toshihide Kitakado**[3] | **Laurence Kell**[2] | **Iago Mosqueira**[4] | **Ai Kimoto**[5] | **Rob Scott**[6] | **Carolina Minte-Vera**[7] | **Paul De Bruyn**[8] | **Yimin Ye**[1] | **Jana Kleineberg**[9] | **Jo Lindsay Walton**[10] | **Shana Miller**[11] | **Arni Magnusson**[12]

[1]Fisheries and Aquaculture Policy and Resources Division, FAO, Rome, Italy

[2]Centre for Environmental Policy, Imperial College, London, UK

[3]Tokyo University of Marine Science and Technology, Tokyo, Japan

[4]DG Joint Research Centre (JRC), Directorate D - Sustainable Resources , Unit D.02 Water and Marine Resources European Commission, Ispra, Italy

[5]National Research Institute of Far Seas Fisheries, Japan Fisheries Research and Education Agency, Shizuoka, Japan

[6]Oceanic Fisheries Program, SPC (the Pacific Community), Noumea, New Caledonia

[7]IATTC (Inter American Tropical Tuna Commission), La Jolla, CA, USA

[8]IOTC (Indian Ocean Tuna Commission) Secretariat, Victoria Mahé, Seychelles

[9]Kleinberg Design, Liverpool, UK

[10]Sussex Humanities Lab, University of Sussex, Brighton, UK

[11]The Ocean Foundation, Washington, DC, USA

[12]ICES Secretariat, Copenhagen, Denmark

Correspondence
Rishi Sharma, FAO, Fisheries and Aquaculture Policy and Resources Division, Viale delle Terme di Caracalla 00153 Rome, Italy.
Email: rishi.sharma@fao.org

Present address
Iago Mosqueira, Wageningen Marine Research, IJmuiden, The Netherlands
Ai Kimoto, ICCAT (International Convention for the Conservation of Atlantic Tuna) Secretariat, Madrid, Spain

Abstract

The five Regional Fishery Management Organizations dedicated to tunas (tRFMOs) are all either developing or implementing Management Strategy Evaluations (MSEs) to provide advice for the stocks under their competencies. Providing a comparative overview will help tRFMOs to learn from one another and to collaborate on common solutions and may also help to more clearly define the challenges of building decision support tools in contexts of large scientific uncertainty and where management requires cooperation across multiple stakeholders characterized by unequal power and divergent interests. For example, our overview showed that in most cases, a grid-based design with an emphasis on structural uncertainty has been adopted. However, uncertainties such as sampling errors and non-stationarity of important ecological processes, which are of potentially equal significance for demonstrating robustness of management procedures, were not considered. This paper identifies key issues for operating model (OM) design that challenges the tRFMOs, compares how these challenges are being met, summarizes what lessons have been learned and suggests a way forward. Although the current approach of using assessment models as the basis for OM design is a reasonable starting point, improvements should be made to the conditioning of OMs, especially with respect to enabling the inclusion of other important processes and uncertainties that are difficult to account for in stock assessments but that can crucially affect the robustness of advice. Attempts should also be made to improve documentation and communication of uncertainties that are included and those that are excluded from consideration in the process.

KEYWORDS
grid design, Management Strategy Evaluations, model design, model validation, stakeholder involvement, tuna-like species

---

Previous Address: NOAA, Conservation Biology Division, NWFSC, 1,201 Lloyd Boulevard, Portland, OR, USA (Rishi Sharma)

1 | INTRODUCTION

Management Strategy Evaluation (MSE) was pioneered by the Scientific Committee of the International Whaling Commission (IWC) in the late 1980s (Punt & Donovan, 2007). IWC developed what is now known as the MSE framework to rigorously test different management procedures (MPs). Management procedures may refer to a wide-ranging set of practices aimed at promoting sustainability in fisheries, based on ecological, social and economic objectives. Within the MSE simulation framework, MPs are usually simplified representations of the processes of collecting and analysing data and the algorithm for determining catch limits, usually referred to as harvest control rules (HCRs). These were developed initially for commercial whaling of baleen whale species and later also to safely manage aboriginal subsistence hunts, with an aim to ensure the sustainability of the populations being hunted. The rules that make up a HCR are commonly expressed in the form of an algorithm, taking either observations or modelling results as inputs and producing catch or fishing effort advice as outputs. Since the 1980s, MSE has been used to develop MPs and harvest control rules (HCRs) for numerous fisheries in South Africa (e.g. small pelagics sardine (*Sardinops sagax*, Clupeidae) and anchovies (*Engraulis encrasicolus*, Engraulidae), hake (*Merluccius capensis*, *Merluccius paradoxus*, Merlucciidae) and rock lobster (*Jasus lalandii*, Palinuridae) (Bergh & Butterworth, 1987; De Oliveira & Butterworth, 2004; Geromont, De Oliveira, Johnston, & Cunningham, 1999)), Australia (Queensland spanner crab (*Chionoecetes bairdi*, Oregonidae) fishery (Dichmont & Brown, 2010) and the Northern Prawn Fishery (*Pandalus borealis*, Pandalidae) (Dichmont, Deng, & Punt, 2008; Dichmont, Ellis, & Bustamante, 2013)), and other fisheries across the globe (Goethel et al., 2019; ICES, 2019). Recently, the five tuna Regional Fisheries Management Organisations (tRFMOs) have begun developing recovery and long-term management plans for a range of stocks using MSE (Kell et al., 2015) and have already implemented MPs in a few cases (Hillary et al., 2016; Preece, Davies, & Hillary, 2018).

TRFMOs are intergovernmental organizations devoted to monitoring and managing tuna and tuna-like populations. The five main tRFMOs include The International Commission for the Conservation of Atlantic Tunas (ICCAT); The Indian Ocean Tuna Commission (IOTC); the Western and Central Pacific Fisheries Commission (WCPFC); the Inter-American Tropical Tuna Commission (IATTC); and The Commission for the Conservation of Southern Bluefin Tuna (CCSBT). Each tRFMO includes a number of member states, generally coastal states that adjoin the oceans which the regional organization covers (Figure 1). There are also some member states with distant-water fishing fleets that are members of several tRFMOs. Each tRFMO has a convention and process embedded within its articles of operation, and the latter are continually amended based on priorities determined by the respective Commissions. ICCAT was established in 1969 based on a convention signed in 1966. The Commission consisting of the representative of all contracting parties oversees various organizational divisions such as the Standing Committee on Research and Statistics (SCRS), see organigram on ICCAT's website for more details on the institutional structure (https://www.iccat.int/en/organization.html). The SCRS facilitates developing MSEs through awarding short-term contracts to modellers working alongside ICCAT's species-based Working Groups and by supporting the activities of the Joint tRFMO Working Group on MSE that met for the first time in 2016 (http://www.tuna-org.org/mse.htm).

TRFMOs do not have identical institutional structures or histories, but are often subcontracting the same scientists as the number of available and experienced MSE modellers is limited. For example, CCSBT was established in the early 1990s with some members, for example South Africa, joining as recently as 2016; however, CCSBT was the first tRFMO to use MSE. This arose because of certain members advocating different Southern Bluefin Tuna (*Thunnus maccoyii*, Scombridae, SBT) assessments with different management implications, leading ultimately to international litigation. An international scientific panel, introduced to facilitate agreement, proposed that the CCSBT move to an MP so that the associated feedback control mechanism (Chang, 2014) could cater for the uncertainties associated with assessments.

Other tRFMOs, however, were not under such urgent pressures. In some cases (e.g. IOTC skipjack and ICCAT albacore), the industry's desire to pursue Marine Stewardship Council (MSC) certification was the driving factor for MSE. This is reflected in the different stages of transitioning towards MSE among tRFMOs. There is a strong perception that the MSC certification confers economic benefits on a fishery that acquires it (Lallemand, Bergh, Hansen, & Purves, 2016; Stemle, Uchida, & Roheim, 2016). In 2016, when various tRFMO fisheries were certified, a condition was set by the MSC that mandated that a well-defined HCR should be in place. For example, for North Atlantic albacore, the MSE process started in 2013 after the Commission requested the SCRS to develop a limit reference point in response to the desire to obtain certification from the MSC for the North Atlantic albacore artisanal fishery. Similar pressures were exerted by the pole-and-line fishery in the Maldives that was pursuing certification for skipjack tuna in the Indian Ocean.

It is generally accepted that MSE can play an important role in shifting the management of fisheries towards sustainable and risk-averse practices. The MSE process is designed to help find those adaptive approaches (HCRs) that have a high probability of meeting management objectives, which usually include avoiding stock collapse with high probability, ensuring that the biomass is maintained around the level that is sustainable in the long run, while securing high and stable long-term profits for the industry. MSE is also credited with other benefits, for example, in the case of CCSBT, reconciling differences in beliefs among different stakeholders. MSE can be conducted in such a way that it empowers collaborative, transparent and inclusive decision-making, or, in the opposite extreme, it can shift decision-making onto an algorithm, simplifying annual negotiations over quotas. The role that MSE ends up playing very much depends on the specific context of its application. MSE may help move fisheries towards social, economic and environmental sustainability, but unfortunately, it does not guarantee that we get there.

In recent years, MSE has played a role in the provision of advice in both the IOTC and ICCAT. In the WCPFC and IATTC, the developments of MSE frameworks are at earlier stages. Each tRFMO has different management objectives specified in their conventions, so that the MSE process is subject to expressed wishes of respective Commissions that consist of member and cooperating non-contracting members or entities. In summary, CCSBT was the first tRFMO to develop a full MP using MSE, while ICCAT and IOTC started on this path earlier than either WCPFC or IATTC.

Setting up an MSE framework involves reaching agreement on various components and clarifying fishery-specific management objectives. An initial step is to select a reference set of operating models (OMs) which represents the most important uncertainties about the resource and fishery dynamics. OMs are plausible representations of the real world which are used to simulation-test MPs to reveal their likely consequences, risks and trade-offs. There is rarely a consensus on which uncertainties are most important, even within closely knit modelling teams let alone a wider group of stakeholders (Leach, Levontin, Holt, Kell, & Mumford, 2014). At this stage, therefore, it is important that the set of OMs does not represent an overly narrow view of the system, since excluding crucial sources

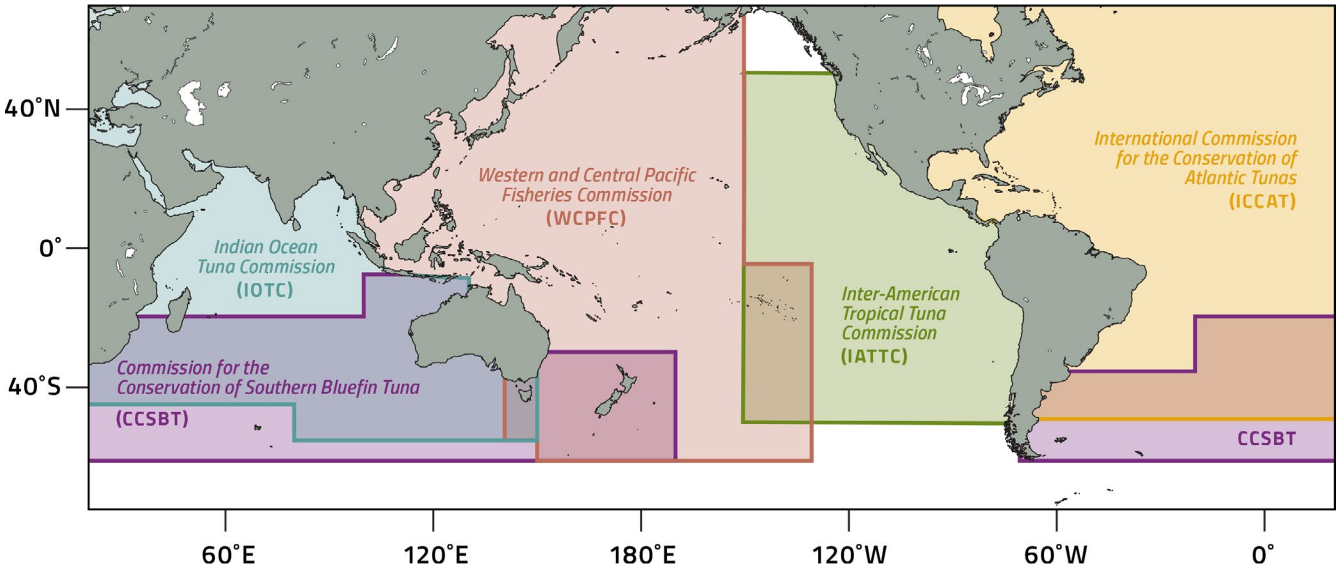


FIGURE 1 Map of tRFMOs jurisdictions

of uncertainty may lead to an MP that cannot meet management objectives, especially those objectives linked to risk values. Second, management objectives must be specified so that the prospects of meeting those objectives can be quantified through summary statistics collected from the simulations. Translating management objectives, especially those related to society rather than biology, into quantities that can be output by the model is not straightforward. The third component of MSE is specifying the candidate management procedures themselves. Management procedures fall into two categories: model-based or empirical, depending on whether restrictions on fishing are based on model outputs or directly on observations. Finally, a set of robustness OMs (Punt, Butterworth, de Moor, De Olivera, & Haddon, 2016) is developed to account for a wider set of uncertainties than covered by the reference set or, more rarely, to address differences in beliefs about the resource by different stakeholders. The order of these steps is by no means crucial, and some can be done in parallel; however, it is important to agree on how to decide which OMs can be included in robustness trials before seeing the simulation results. Otherwise, there is a risk that the plausibility of OMs will be judged on the basis of the performance of MPs, with some OMs rejected as implausible *after* it turns out that the candidate MPs are not robust to the sources of uncertainty these OMs represent.

One of the reasons for using MSE is to simplify negotiations and automate management advice in accordance with a longer term vision for the future; simulation evaluations of MPs were made possible by increases in computational power that have enabled fisheries scientists to develop complex models and conduct robustness studies that are ever more numerically intensive. This trend has led to two important, though seldom addressed, problems:

1. a lack of transparency as complex computer simulation models are harder to understand and document as many are written in high-level programming languages (like C, R and TMB); many “hide” assumptions whose importance is never assessed (this is an especially acute problem for data-limited stocks) as they are inherent properties of the OM that only a select few know about,
2. a lack of access, as only a few highly skilled modellers fully understand and can run such complex models (Hilborn, 2003). This exacerbates problems of representation and equity in the decision-making process, as those with the insider modelling expertise (and hence influence) tend to come from the more developed countries, compounding their economic and political power with greater technical and scientific expertise as well as access to data.

To address a lack of transparency, investment in communication of uncertainties and all other aspects of OMs is essential, for example, by developing suitable visualizations (Levontin et al., 2020). Widening participation in modelling should be an explicit goal of the tRFMOs, education and training opportunities should be offered, with particular attention towards gender and ethnicity representation. More immediate benefits, however, could be obtained by discussing and agreeing on better programming, code-sharing and model validation practices within the small modelling community that is servicing the tRFMOs.

In an effort to improve communication, this paper examines how various uncertainties are being addressed within OM design by the different tRFMOs. It identifies common problems and offers advice on the future improvements to the crucial process of accounting for uncertainties in setting up the MSEs. In the first part of the paper, a brief description of tRFMO approaches to conditioning and selecting of OMs is provided by region and stock, and lessons that can be learned from each are summarized. Then an overview of common issues across tRFMOs is presented, the lessons that could be learned from encountering these issues are discussed, and finally, suggested solutions are explored and summarized.

2 | INDIAN OCEAN TUNA COMMISSION (IOTC)

IOTC accepted MSE as a means of meeting its obligations under the precautionary approach in 2013. The main commercial species of the industrial fleets (albacore (*Thunnus alalunga*, Scombridae) (ALB), bigeye tuna (*Thunnus obesus*, Scombridae) (BET), skipjack (*Katsuwonus pelamis*, Scombridae) (SKJ), yellowfin tuna (*Thunnus albacares*, Scombridae) (YFT) and swordfish *(Xiphias gladius*, Xiphiidae*)* (SWO)) were identified as the initial priorities. Skipjack was given special urgency due to the Maldives pole-and-line fishery MSC certification conditions. Maldives pole-and-line fishery for skipjack first earned an MSC certificate in 2014; the certificate has been briefly suspended over the status of the fishery in 2016; it was recertified in 2017, and according to the first surveillance report in March 2019, the main issues (the existing certification condition) pertain to non-compliance with data reporting protocols, exceeding recommended fishing targets and a lack of agreement over allocation of quotas (fisheries.msc.org). An MSE-tested HCR was adopted for skipjack in 2017 (first quota set in 2018), and others are currently in various stages of development. The quality and quantity of data available for conditioning OMs within IOTC differ from species to species, but include at a minimum: total catches, commercial CPUE abundance indices and catch-at-length. In addition, some tagging data are available for skipjack, bigeye and yellowfin tuna.

Spatial structure is included in the OMs for skipjack, bigeye and yellowfin tuna, but there is assumed to be a single spawning population and biology is identical among regions (except for movement rates). The spatial dynamics is modelled using estimated movement rates from tagging data; however, confidence in these estimates is low because tag releases have been unbalanced, mixing rates within regions are low, and tag reporting rates are low or unquantified for most fleets. Genetic studies suggest that rather than a single population, a multiple stock structure is a more plausible hypothesis, and as a trial, this assumption was modelled for bigeye and yellowfin tuna—these multi-stock/multispecies models are at an exploratory stage and the default assumption for OMs remains that of a single stock.

Further specifics concerning the OMs are summarized below on a species by species basis (IOTC, 2019a, 2019b, 2019c).

2.1 | IOTC albacore

For albacore, a base case OM was conditioned on fishery-dependent data using the Stock Synthesis assessment model (SS, Methot & Wetzel, 2013). Several sources of uncertainty were considered in the form of scenarios: alternative fixed values for hard-to-estimate parameters such as natural mortality (M) or the steepness of the stock–recruitment relationship (h), and several choices for the selection pattern and data weighting (Mosqueira, 2018). A full-factorial design (ensemble of models), or a grid approach, was implemented where all interactions between scenarios were run (see Table 1). Each ensemble represents the maximum posterior density (MPD) estimates assembled from fitting various models selected from a balanced "grid" of interacting assumptions.

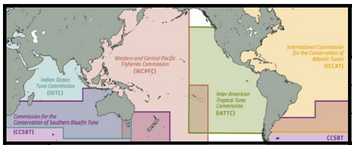
A number of these runs led to unrealistic estimates of virgin biomass (B_0) or trajectories of the historical stock. Clearly unrealistic runs were filtered out based on an upper limit for unfished biomass (B_0), where this limit was estimated based on the relationship between carrying capacity (K) and suitable habitat for all global albacore stocks (Kell & Mosqueira, 2017; see Table 1). Principal component analysis (and sometimes linear modelling) was used to identify those combinations of scenarios in the grid that had appreciable effects, revealing which combinations of assumptions mattered most.

Further, a large proportion of models was rejected based on external data validation criteria, as they predicted that observations available for validation were highly unlikely. It thus became clear that many OMs in the grid lacked the short-term predictive capacity and hence were unsuitable for the MSE. In addition to the external data validation test, models were also assessed for convergence. Many of the operating models showed convergence problems and were excluded on that basis (IOTC, 2019d).

2.2 | IOTC skipjack

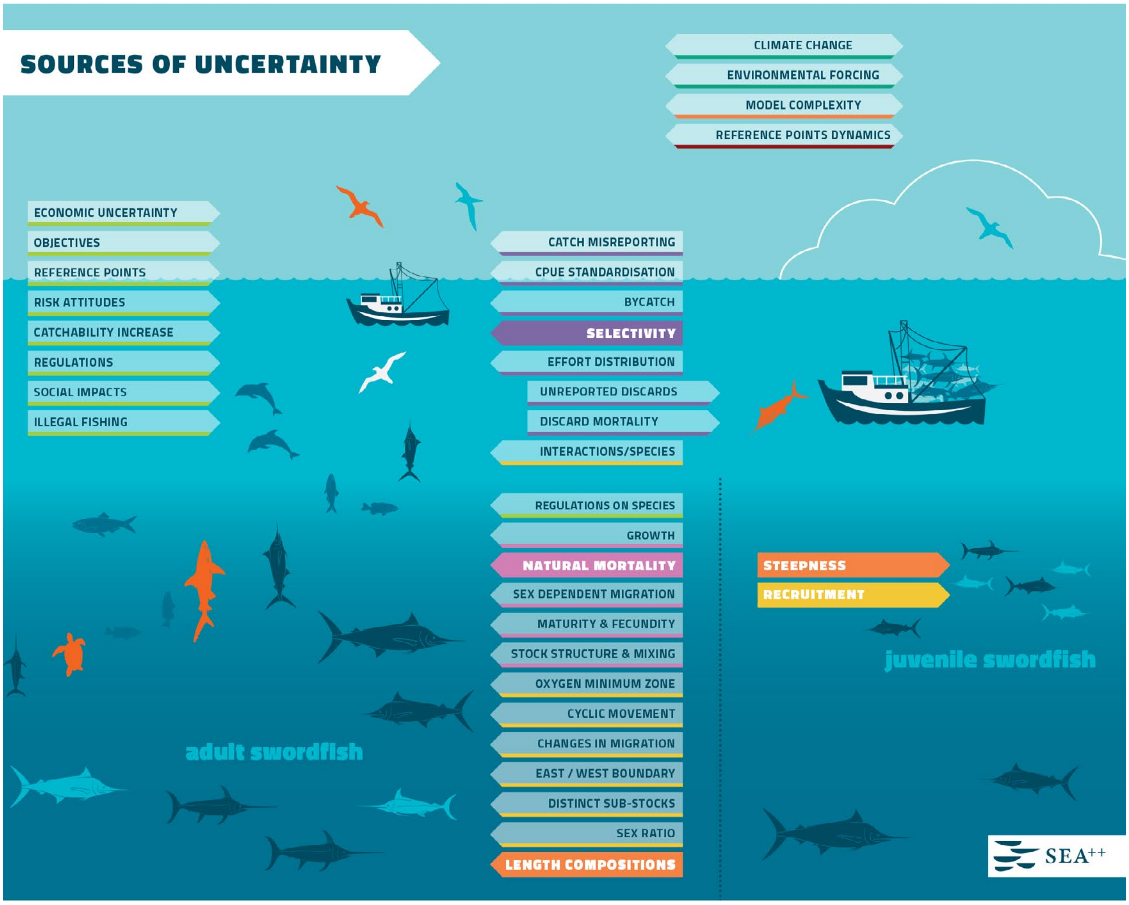
Unlike albacore, whose OM essentially mirrors the stock assessment used at present for management advice, the OM for skipjack tuna was tailor-made and included spatial structure that described the particular dynamics of the Maldives pole-and-line fisheries using different stock structure assumptions. Stock assessment estimates provided the basis for some of the OM parameters (Table 1), while other parameters were represented by prior distributions. These

TABLE 1 Uncertainty sources considered in case-studies across tRFMOs

		IOTC				ICCAT			CCSBT	WCPFC	IATTC	
		ALB	SKJ	YFT, BET	SWO	ALB	BFT	SWO	SBT	SKJ	ALB	BET
		SS3	Custom	SS3	SS3	Multifan-CL	Custom	SS3	Custom	Multifan-CL	SS3	SS3
		Grid	Grid	Grid	Grid	Grid	Grid	Grid	Grid	Grid	Grid	Main
1	Catch								X			
4	CPUE uncertainty over standardisation	X	X	X	X				X			
6	Selectivity	X			X	X		X	X		X	
7	CPUE spatial issues				X							
9	Environmental forcing					X						
10	Growth		X		X						X	X
11	Natural mortality	X	X	X	X	X	X	X	X		X	X
13	Maturity				X		X					
19	Catchability	X		X	X					X		
24	Model complexity									X		
25	Steepness	X	X	X	X		X	X	X	X	X	X
26	Effective Sample Size (ESS)/ Alternative data weights	X		X	X	X	X	X	X			
31	Stock structure											
34	Recruitment variability/regime shifts/S-R residuals (sigma r)	X			X	X		X		X	X	

Note: The model on which OMs were conditioned as well as whether uncertainties were considered in a grid or as main effects only are indicated in head rows. As in Figure 2, uncertainties are categorized and colour coded by type and referenced by a number (first column), consistent with the previous expert elicitation studies (Leach et al., 2014).

UNCERTAINTIES PERTINENT TO THE NORTH ATLANTIC SWORDFISH MANAGEMENT STRATEGY EVALUATION



KEY

UNCERTAINTIES CONSIDERED IN MSE

UNCERTAINTIES EXCLUDED FROM MSE

Colours:
- ■ Catch
- ■ Environmental
- ■ Life History Traits
- ■ Management
- ■ Model
- ■ Population Structure
- ■ Reference Points

CATCH

1. Catch misreporting; catch under-reporting—in particular of juvenile catch in artisanal fisheries
2. Discard mortality
3. Unreported discards
4. CPUE standardisation; Consider CPUE conflicts (by area, NW/NE Atlantic)
5. Bycatch
6. **Selectivity; gear selectivity/catchability changes by fleet (e.g., gear changes, other effects not accounted for in the CPUE standardisation)**
7. Changes in effort distribution: seasonal dynamics (stock/fleet)

ENVIRONMENTAL

8. Climate change; climate change and/or increased variability's potential to change population dynamics
9. Environmental forcing; environmental considerations and behaviour

LIFE HISTORY TRAITS

10. Growth
11. **Natural mortality**
12. Sex dependent migration: spatial sexual segregation of the stock (real or observed)
13. Maturation and fecundity
14. Stock structure and mixing; group dynamics, skipped-spawning, density dependence

MANAGEMENT

15. Economic uncertainty; market and other economic data to be used in assessing the risks
16. Uncertainty over objectives; management objectives
17. Uncertainty over the reference points; reference points and the lack of information on virgin stock levels
18. Risk attitudes of managers
19. Catchability increase
20. Effect of regulations on effort; effect of the minimum size recommendation [Rec. 17-02]; Implementation options (119 cm or 125 cm LJFL)
21. Social impacts on local communities; social impacts of regulations and its effect on small local communities
22. Illegal fishing; regulations that change the balance of effort between legal and illegal fisheries
23. Effect of regulations on species; impacts of regulations and its effect on the apparent global distribution of the species.

MODEL

24. Model complexity
25. **Steepness**
26. **Alternative data weights (length comp); Length compositions effective sample size**

POPULATION STRUCTURE

27. Oxygen minimum zone, i.e., vertical displacement of individuals
28. Cyclic movement of adult swordfish
29. Changes in migration; Environmental factors that influence migration patterns
30. East-West boundary; Location of the current boundary, mixing between East and West stocks
31. Existence of genetically distinct and vulnerable sub-stocks
32. Sex ratio
33. Interactions with other species
34. **Recruitment Variability Recruitment failure of success (cyclic trends/ regime shift); recruitment variability**

REFERENCE POINTS

35. Dynamics of reference points; Stationarity, cohort year effects, density dependence

FIGURE 2 Uncertainty sources considered/not considered in the initial MSE of North Atlantic swordfish (ICCAT). Uncertainties are categorized and colour coded by type, consistent with the previous expert elicitation studies (Leach et al., 2014), see the Key below the main image. For further visualizations, see the app prototype developed for this project (https://pl202.shinyapps.io/Swordfish_MSE_Vis/)

different parameter distributions were resampled and the corresponding stock trajectories were generated. These were then filtered through a set of feasibility criteria, such as "stock has not collapsed." The approach used for skipjack in the Indian Ocean was originally developed for data-poor stocks (Bentley & Langley, 2012), since skipjack is considered to be more data-limited than the other main IOTC commercial species (Bentley & Adam, 2014). Only 7% of the runs passed the feasibility criteria. More generally, there are concerns that such selection criteria can bias the remaining set of OMs towards those that are more optimistic about the stock, truncating the full range of uncertainties and hence biasing conclusions about the robustness of management procedures. These kinds of technical details that might have important consequences for the interpretation of the MSE results need to be discussed across tRFMOs, justified and communicated to stakeholders.

2.3 | IOTC bigeye and yellowfin tropical tunas

An MSE for Indian Ocean bigeye and yellowfin tuna is currently being developed where OMs are conditioned on stock assessment, as in the case of albacore. The OMs consist of an ensemble of models derived from the most recent SS assessments. Similar to Albacore, MPD estimates assembled were generated based on the model ensemble. For both species, the grid included similar scenarios: three levels of natural mortality, three levels of Beaverton–Holt stock recruit steepness and two levels of CPUE catchability trend (no change in catchability; 1% per year increase, Table 1). The yellowfin ensemble also considers three different tag data-related weighting assumptions on the posterior density. The number of uncertainties (additional grid dimensions) considered was expanded in subsequent iterations of the MSE (IOTC, 2019c).

Simple qualitative summary diagnostics were examined to ensure that the individual models meet basic plausibility criteria, including adequate convergence, "reasonable" fit to the data, and no "surprising" outlier dynamics. It was deemed desirable from the MSE perspective that the aggregate ensemble should have a unimodal distribution of characteristics (additional intermediate levels of assumptions were introduced when this was not the case). However, in general, there is no requirement for the aggregate ensemble model distribution to be unimodal. In fact, it may be impossible to represent differences of prior beliefs/specifications of alternative plausible assumptions in a unimodal distribution (IOTC, 2018).

The latest developments in the OM conditioning include (IOTC, 2019a): (a) updating the bigeye MSE to reflect 2019 assessment, (b) updating CPUE series (and alternative catchability scenarios) derived from recent collaborative work on operational data, (c) further exploration of spatial processes and movement estimates and (d) inclusion of alternative catch time series to reflect the data uncertainty in many artisanal fleets. All models are currently weighted equally in the OM ensembles, but differential weighting can be employed in the future. Multispecies OMs taking into account the multi-stock considerations are important to explore as alternative OMs given the nature of yellowfin and bigeye tuna fisheries. But even in current single-species settings, implementation error could be inflated to partially account for the multispecies bycatch considerations.

2.4 | IOTC swordfish

There have been similar early-stage developments in the MSE for the Indian Ocean swordfish stock, with 2,592 OMs run based on a grid approach and conditioned on SS stock assessment (Table 1, IOTC, 2019b). Models were selected based on convergence, residual analysis and plausibility of virgin biomass values. For example, if the gradient in the models were greater than 0.001, these models would not be used, as they had probably not converged. As in the albacore case, if B_0 estimates from the model were above plausible values for the Indian Ocean (as derived from a meta-analysis of other Swordfish populations), those models would also be rejected.

2.5 | IOTC lessons learned

In the IOTC, numerous lessons have been learned in the conditioning process: (a) a number of criteria should be applied to evaluate model plausibility, including tests for numerical convergence, parameters on bounds and outlier behaviour in the quality of agreement between data and model predictions. In some cases, the data are not sufficient to distinguish whether stocks have declined due to fishery effects or because of declining recruitment due to non-fishery-related causes, in which case more subjective plausibility criteria are under exploration (e.g. in relation to albacore habitat, and yellowfin tuna stationarity in recruitment dynamics). Parameter confounding is a problem, as it is difficult to distinguish seasonal movement from changes in catchability. A concept of catch penalty was explored: this corresponded to the level of effort that would be needed to secure the observed catch level, and if the effort implied was so high as to be implausible, the OM was rejected as unrealistic. This approach was developed partly in response to the inability of many of the OMs to explain recent catches without an implausible increase in effort (this was the case in yellowfin tuna OM, IOTC, 2018; IOTC, 2019a).

(b) Options for introducing time-varying fishery selectivity have been implemented, but not parameterized, as finding a suitable set of hypotheses to test is difficult. As such, providing alternative length-based selectivity ogives conditioned on the data could give very different stock trajectories, as dome-shaped selectivity could imply a large cryptic biomass that becomes a lot lower when we use asymptotic selectivity (Butterworth, Rademeyer, Brandao, Geromont, & Johnston, 2014). A similar issue was detected in CCSBT and corrected in the plausible OMs chosen (Hillary et al., 2016).

(c) Given the uncertainty in how quotas will be implemented (and allocated), implementation error is included in the models, with specific scenarios under consideration. Unusually, it is estimated that the bigeye population might be able to sustain considerably higher catches than recently attained (for unknown reasons, possibly related to piracy), such that under-catch implementation error may be important. This is related to point two, as fleet selectivity can have large implications on stock dynamics and long-term trajectories, that is if allocation changes from larger older fish like in a longline fishery to smaller younger fish as in a Fish Aggregating Device (FAD) Purse Seine fishery, the overall allowable catch may need to be reduced as the MSY would decrease as a function of selectivity.

(d) In the case of skipjack, the OM was based on the "feasible stock trajectory" approach of Bentley and Langley (2012), developed for data-poor situations. Simulations consisted of (i) projecting an initial population (with randomly sampled parameters) through the catch history, (ii) evaluating whether the historical projection was plausibly consistent with the historical data (more than 90% of simulations are rejected), and (iii) conducting future projections with the plausible realizations to evaluate expected management performance. Here, parameter priors are based on a combination of assessment inferences and life-history meta-analyses from related populations and species. Thus, choice of priors could have a large influence on the OM dynamics unless conditioned correctly on the data.

(e) OM ensembles have mostly been a uniform sampling of the plausible models in the grid. In the case of yellowfin tuna, it became apparent that the assessment is sensitive to many assumptions. A possible pragmatic approach has been suggested which involves stochastic sampling of the grid to achieve a distribution of stock status inferences (Bcurrent/BMSY and MSY) that is consistent with the mode of the assessment, but subject to an arbitrary and potentially substantial inflation of the variance. This approach has also been attempted for albacore. A lesson here is that the distribution of OMs examined should be internally consistent with the results of the mode of the assessment.

OMs are being developed for albacore, bigeye, and yellowfin that are all derived from the recent integrated stock synthesis (SS) assessments but the assessments themselves are undergoing regular substantive changes as the IOTC scientific community continues to address recognized problematic assumptions and other uncertainties in stock assessments. One of the criticisms of the MSE approach is its dependence on stock assessments. When stock assessments are updated, it is not just that our perception of the current status of the stock that may be altered, but also our understanding of its historic trajectory as well as aspirations for management may need to be revised in turn. Because stock assessments represent our current beliefs about the stock's history, dynamics or possibilities, it is difficult to condition management upon a model that represents a different view. How can we be sure that the MPs in place are still robust under a revised view of reality if we have not retested them? Therefore, there are pressures to redo the MSE every time the stock assessment is seen to revise our beliefs about the stock significantly, which weakens the case for the MSE as the methodology to simplify the management system. One remedy is to set out explicit rules for exceptional circumstances to determine when should the MSE be reconditioned on the updated model. Regardless of whether exceptional circumstances had occurred, the OM and MP should be revisited every 5–10 years, as other parameters that are assumed static may have changed since then.

It should be noted that the OMs for each species are typically coded by a single developer, sometimes in cooperation with another programmer. This creates a problem not only of highly privileged, or inaccessible to other scientists, knowledge but also of mundane coding errors because the specific code is not checked independently. The OM assumptions and plausibility evaluations, however, are presented to one or more IOTC technical working parties each year (methods and species-specific working groups) for feedback, with some level of consultation with the relevant stock assessments which happen in parallel (though not every year). There is also an informal MSE steering committee that monitors progress intersessionally with a dedicated meeting, attended by the developers and other interested parties. In practice, it is not the case even in places where an MP has been implemented that an algorithm for decision-making is agreed upon and then left to run without oversight or revisions for any number of years.

3 | INTERNATIONAL COMMISSION FOR THE CONSERVATION OF ATLANTIC TUNAS (ICCAT)

In ICCAT, three MSEs have been developed so far (ICCAT, 2019), and their details are described below. It has taken a longer time to get agreement on conditioning and performance metrics than desired. One reason for this delay is an institutional complexity on how the Commission interacts with the scientific committees at ICCAT. Although the work may be done by the SCRS, the Commission may not provide any clear guidance on management objectives, thereby delaying the process of reporting on some agreed performance metrics. On the other extreme, the objectives maybe too broad which make it difficult for the SCRS to provide clear guidance.

3.1 | ICCAT North Atlantic albacore

For North Atlantic albacore, by 2013, two stock assessments were in place, Multifan-CL (Fournier, Hampton, & Sibert, 1998) and a biomass dynamic model ASPIC (Prager, 1994). At the time, there were problems with the Multifan-CL assessment due to conflicts in the data, which the working group could not resolve, so TAC advice was based on the biomass dynamic model. However, it was the

Multifan-CL assessment that was subsequently used as a basis for the OM scenarios (as it could be used to evaluate a greater range of hypotheses) that were run as a full factorial design over the grid of uncertainty axes. This grid included a variety of weightings for the different input datasets, priors for steepness and a range of fixed values for natural mortality (Table 1). No attempt was made to reject or weight the different OMs, although there is a consensus that this is not the best practice (the issue of weighting OMs was discussed at the recent joint tRFMO meeting in Seattle, June 2018). While the OM developed for albacore incorporated various alternative structures, it was missing important uncertainties related to non-stationarity in recruitment. The observation error model (OEM) did not replicate the error structure of the historic observations, due to changes in the fishery that were not modelled. This created difficulties in interpreting and ascertaining the robustness of MPs that relied on a simulated index of abundance that in reality had ceased to be informative. In other words, the MP could not be tested with respect to a relevant observation uncertainty as simulated observation error pertained to a Japanese longline by-catch index that became unusable as a part of a MP due to changes in catchability. Nonetheless, an HCR was eventually adopted in 2017 and the MSE work is continuing towards selecting a robust MP (HCR and assessment) that can meet management objectives.

3.2 | ICCAT Atlantic bluefin tuna (*Thunnus thynnus*, Scombridae)

An MSE Trial Specifications document (Anon, 2017) describes the reference and robustness sets of OMs under development for Atlantic bluefin tuna. The principal axes of uncertainty sources considered, include the following: (a) the functional form and steepness of the stock–recruitment relationships; (b) the magnitude and trend in historical stock biomass; and (c) the maturity and natural mortality rate (Table 1). In addition to these axes of uncertainty, operating models for Atlantic bluefin tuna also account for seasonal and age-specific multi-stock mixing. These movement dynamics are considered important since for practical reasons management measures have to be implemented for east and west areas despite conclusive evidence for the mixing of Eastern and Western origin stocks throughout the North Atlantic (Richardson et al., 2016; Rooker et al., 2008). To inform such a model, a wide variety of data was required such as genetic stock assignment data, otolith microchemistry data, electronic and conventional tags (Carruthers, Power, Lauretta, Di Natale, & Kell, 2016). Since the estimation model needed to accommodate novel movement models and likelihood functions, a custom operating model was developed in ADMB referred to as M3 (the Modifiable Multi-stock Model, Carruthers, Kimoto, et al., 2016).

To accommodate a long history of exploitation that extends to the mid-15th century, M3 is a hybrid of two stock assessment types, subtracting catch in a stock-reduction analysis (SRA, Kimura, Balsiger, & Ito, 1984) prior to 1965, after which M3 operates as a conventional statistical-catch-at-age model. Parameter uncertainty is characterized by numerical approximation of the posterior density via Monte Carlo Markov chains (MCMC). The OM is distributed in a fully documented R package (ABTMSE, 2018) that allows users to fit M3 to data, import model estimates to an R MSE framework and conduct MSE analyses for the development and testing of custom MPs. The package is subject to ongoing revision following an MSE workshop in early 2018 (Carruthers & Butterworth, 2018) and is to be updated for further testing. The main challenge for the model is to deal with apparent conflicts in data, reasonable fits to data can only be achieved through specifying strong priors and carefully selecting weightings for various data components.

3.3 | ICCAT swordfish

An exploratory MSE was conducted for North Atlantic swordfish, with an OM conditioned on SS stock assessment (Kell & Levontin, 2019). An interactive website was built to explore swordfish MSE, with a particular emphasis on communicating which uncertainties were considered (Figure 2). A model validation approach was developed, consisting of several steps that can be generalized into a procedure applicable across tRFMOs (Tables 2 and 3). Model validation tests for OM conditioned on stock assessments can reveal problems with the stock assessment model that have gone unnoticed or unaddressed, and this was the case with North Atlantic swordfish. Conflicts in data, especially indices of abundance are common, but the residual tests confirmed the severity of the problem in swordfish stock assessment. Residual tests might also indicate model misspecification, and foretell of problems in simulating data for MPs. One of the desired qualities in the OM is short-term predictive skill. An OM which cannot predict data in the near future would not be considered a plausible representation of reality—OMs are expected to roughly agree with stock assessments and/or be consistent with recently observed data that was not used in the assessment (external data validation). A hindcast procedure can test the ability of OMs to make short-term predictions. For swordfish, a hindcast procedure that removed 1–5 recent years of data in order to see if the model is able to predict them revealed that the OM had poor predictive skill (Kell & Levontin, 2019).

Another test for the plausibility of OM was an implicit comparison with simpler models. Such comparisons can clarify what features of the model are driving dynamics. These tests revealed that although the r and K parameters were in plausible ranges, the shape of the production function was highly skewed. In effect, the OM depicted an extremely resilient population which, in order to be exploited at a maximum sustainable level, would need to be dramatically fished down from its unfished biomass level. The swordfish OM implied that BMSY could be below 20% of the unfished biomass, a depletion level to be avoided with high probability (B_{lim}). Such dynamics of the OM appear to be an artefact of the modelling and not something that can be justified with biological knowledge. Constructing an MSE with a version of reality that implausibly presumes a stock of unnatural productivity and resilience is hardly consistent with the precautionary approach (FAO, 1996; Richards & Maguire, 1998). Conclusions about the robustness of MPs are only as good as the assumptions in the OM (Kell & Levontin, 2019).

3.4 | ICCAT lessons learned

In some cases, conventional stock assessment frameworks may not be sufficiently flexible to accommodate principal sources of structural uncertainty. For Atlantic bluefin tuna, a multi-stock, spatial and seasonal model with age-based movement (M3) was required but could not be assembled straight out of conventional assessment packages. A multi-stock model has the principal challenge of assigning predicted catches to stock in any given time–area strata. As in most assessments, it is the magnitude of catches that scales the relative size of each stock. While stock assignment data (e.g. otolith microchemistry) provide information about the stock composition of catches, these data do not provide information about the exploitation rate. In principal, electronic tagging data provide movement information to infer the spatial distribution of each stock and hence a prediction of spatial biomass from which exploitation rate can be calculated. However, in practice, it was found that these data are sparse and additional information is required to ensure that the model does not estimate unrealistically high or low biomass in any time–area strata (for example hiding 90% of the biomass in the Caribbean with a very low exploitation rate). An early lesson in the development of Atlantic bluefin tuna OM was that it would not be possible to estimate an exploitation rate parameter for every time–area–fleet strata (more than 30,000 parameters). Instead, the model was configured as a "conditional on effort" model: a standardized effort was derived as the observed catch for each time–areas–fleet strata divided by a so-called "master index" for every time–area strata. This standardized effort data could then predict fishing mortality rate according to just a catchability parameter per fleet (15 parameters) and avoid the estimation of unrealistic biomass. The master index was derived by catch rate standardization of nominal CPUE and included the area-specific indices used in the OM conditioning. To provide a less rigid prescription of spatial exploitation rates, a catchability deviation for each season and area (28 additional parameters) was also included. This "conditional on effort" approach was tested by simulation evaluation and found to be comparable to the estimation of all fishing mortality rates and was more than 100 times faster to converge.

Another lesson learned was that the additional complexity and new features of the M3 model meant that conventional wisdom regarding the weighting of likelihood components for various data was not applicable. It was necessary to develop a simulator to derive robust weightings for the various likelihood components.

The principal lesson learned in the development of M3 is that for such complex models it can be extremely hard to gain an intuitive understanding of model behaviour, view all of its various estimates and diagnose estimation problems. For example, just the visualization of age-based movement over seven areas for two stocks is challenging. A recommendation arising from this is to, where possible, consider developing models that implicitly account for movement (e.g. 2-box VPA). Such models might include two areas, and estimate age and seasonal variation in the fractions of each stock by area. Seasonal fishing dynamics and targeting may be approximated implicitly with seasonal fishing selectivity curves. This configuration may be less suited to complex movement and fishery dynamics but is much more parsimonious and may provide a better trade-off between model complexity, ease of interpretation and computational demand when running the MSE.

TABLE 2 Model validation or filtering tests applied in case-studies across tRFMOs to assess the plausibility of OMs

	IOTC				ICCAT			CCSBT	WCPFC	IATTC	
	ALB	SKJ	YFT, BET	SWO	ALB	BFT	SWO	SBT	SKJ	ALB	BET
Convergence/jittering/ maximum gradient			🟪	🟪			🟪	🟪	🟪		
Residual analysis				🟪			🟪	🟪	🟪	🟪	🟪
r and K							🟪				
External data/cross-validation	🟪						🟪		🟪		
Visual examination/outlier dynamics		🟪	🟪			🟪	🟪	🟪			
Plausibility of Virgin Biomass or other estimates	🟪			🟪					🟪		
Plausibility of the historic trajectory (stock has not collapsed)	🟪	🟪	🟪					🟪			
Plausibility of implied effort changes											
The shape of the production function							🟪			🟪	🟪
Likelihood profile on the scaling parameter								🟪	🟪	🟪	🟪
Plausibility of implied spatial dynamics						🟪					

Note: Colour coding—"purple" for model validation—is chosen for consistency with the visualization app (https://pl202.shinyapps.io/Swordfish_MSE_Vis/).

TABLE 3 A four-point metric for measuring OM model suitability/plausibility on the basis of a generic model validation procedure (Kell & Levontin, 2019)

MODEL				
	VERY GOOD ●●●●	GOOD ●●●	POOR ●●	FAILED ○
CONVERGENCE	Maximum gradients <1e-5 for all retrospective runs and Hessian invertible	Maximum gradients <1e-5 and Hessian invertible for last run and Hessian invertible	Maximum gradients <1e-5	Hessian not invertible
RESIDUAL	All Pass	<50%	>50% fail	All fail
R AND K	Both parameters within 1 sd range	The growth parameter r is within 1sd range but K is not	Both parameters are outside the 1 sd range	Both parameters are outside the 2 sd for this species
EXTERNAL DATA/ CROSS VALIDATION	Prediction residual variance ~ Model variance and Mean Absolute Scaled Error test for 1, 2, 3 step ahead	Mean absolute Scaled Error test for 1, 2, 3 step ahead	Residual variance ~ Model variance	None

Note: Whereas the focus in the tRFMOs has been on the binary choice of accepting versus. rejecting OMs, a non-binary approach based on a four-point scale below might be used not only to decide when to reject an OM but also to communicate to stakeholders the limitations of the OMs that are accepted as part of the MSE.

The more complex a model is, the more difficult it is to validate. Not only it is harder to ensure that there are no data conflicts, or that the model is consistent with the available knowledge but it is more difficult, especially with complex tailor-made models, to ensure that the code is free of errors. For instance, coding errors in the Atlantic bluefin OM had been discovered and this had delayed the progress towards adopting an MP for the stock based on the MSE. To overcome this problem, an independent reviewer was appointed to examine the code line by line. Such rigorous code validation practices are a good idea for all types of OMs, even those that are based on standardized packages as errors commonly occur when the code is adapted for specific case-studies or tasks.

A key lesson learned at ICCAT through the MSE on albacore was that even though a wide range of models were fitted, the conditionings were not checked for adequacy or tested using simulations. This may be related to the average predictions over these OMs differing substantially from the output when the associated MP was implemented. More work is needed before MSE can proceed as the conditioning was not done with respect to the actual assessment and framework currently used for management. Hence, one of the lessons learned at IOTC is that conditioning the results to what is currently done with the assessment would possibly have helped ICCAT avoid this issue.

The treatment of multiple conflicting CPUE series was considered a problem for most stocks assessed by ICCAT, but this is also a problem at other tRFMOs and fisheries in general. The ICCAT assessment process tends to prefer blending series in the interests of achieving consensus rather than running multiple assessments and then combining the assessment results when providing advice. For simulation purposes, this requires careful consideration of the potential misspecification of CPUE, particularly when developing stochastic projections. One of the key uncertainties in the CPUE series is how reliable they are as indicators of abundance trends and how their reliability changes over time — many potential changes, both environmental and fishery-dependent, can cause a CPUE index to lose its ability to be informative and hence prevent its future use within an MP.

The model's outcomes, such as historical trajectories for biomass or its predictions for the mixing of bluefin Eastern and Western stocks, are sensitive to data weighting, so that scenarios for weighting data are capable of generating many alternatives for population dynamics: some data weightings are pointing to a large underexploited stock in slow decline, and some to the opposite but more consistent with other stock assessment models—a stock that is several times smaller but that is (too) rapidly recovering. This is a challenge for the MSE. A lesson learned here is that quality of the information used in the assessment and in the conditioning is critical in understanding the simulations of stock dynamics. Arbitrary weighting of alternative sources of data on either the length composition or the abundance index can provide unrealistic stock trajectories; hence, best practices need to be developed that are grounded in statistical sampling theory.

4 | COMMISSION FOR THE CONSERVATION OF SOUTHERN BLUEFIN TUNA (CCSBT)

An MSE was completed for southern bluefin tuna in 2011 (having begun in 2001), resulting in the Commission adopting a fully specified MP called the "Bali Procedure" (Hillary et al., 2016). The Bali Procedure will be replaced by a new MP called "Cape Town Procedure" that will become operational from 2021 (full information is available on ccsbt.org). The new procedure relies on close-kin genetic mark–recapture data for spawning stock estimates, CPUE data as well as juvenile gene-tagging.

The core set of operating models are fitted over a "grid" (i.e. a full factorial design) similar to some of the examples above, and the results are integrated over this set. Originally, likelihood weighting was considered across all factors, but this was changed to prior weights for most, with likelihood weighting remaining only for two of the factors which reflect values defining the natural mortality-at-age vector (Table 1, CCSBT, 2019). The reason was that likelihood weighting led to lower emphasis for some factors than was considered realistic (probably because some of the input data do not conform to the assumptions made to incorporate them in the likelihood). This is an example of complex technical issues that can surface in one case study but have relevance for others, illustrating the need for collaborative approaches across the tRFMOs.

Operating models were also developed for robustness tests. Robustness scenarios incorporate factors considered either of lesser plausibility or with a lesser impact on estimates of population variables than those included in the grid. The robustness scenarios played a role in the MP selection. Specifications for projections are based on estimates from the conditioning on past data, for example for the extent of recruitment variability, so that it is assumed that the future will reflect the same behaviour as seen in the past (including autocorrelation). The approach used in CCSBT assumed that the structural uncertainty should be large enough to include and cover the parameter uncertainty, as estimation uncertainty was not incorporated explicitly.

4.1 | CCSBT lessons learned

The selection of key axes of uncertainty for the OM grid was based on evaluating a wide variety of model assumptions over the course of several years of research, consultations and meetings, that is "forecasting process framework" (Hillary et. al., 2017). CCSBT is one of the few successful cases where all issues were addressed through consensus and involved a high level of involvement from stakeholders (Hillary et al., 2016). This is probably helped by CCSBT only having responsibility for a single stock, unlike the other tRFMOs. A lesson learned here is that by keeping the focus of the Commission on one stock and engaging the key stakeholders in smaller groups could be a faster way to make progress. Whereas parallel processing of several MSEs at once, as in the case of IOTC and ICCAT, is slower. CCSBT also benefited from the following conditions that are not available in any of the other tRFMOs, namely: (a) CCSBT has no in-house expertise; (b) there was continuity of individual member-country scientists (minimal turnover of people familiar with the model and goals); (c) the Scientific Committee was led by an independent Chair throughout the process; and (d) an independent Advisory Panel (with support from a technical consultant) provided (and continue to provide) critical contributions to the coordination and success of the OM and MP testing. The latter in fact provided most of the original assessment modelling and OM development.

5 | WESTERN AND CENTRAL PACIFIC FISHERIES COMMISSION (WCPFC)

Work on the development of OMs has only recently begun in the WCPFC. While many members of the WCPFC have expressed a preference that MPs are developed at a fishery level, much of the initial work to develop OMs has been undertaken at the individual species level, with skipjack (tropical purse seine fishery) and South Pacific albacore (southern longline fishery) being prioritized for initial consideration. Species-specific OM development has proceeded with a view to eventually running the models in parallel so that outputs relevant to multispecies fisheries can be obtained. In addition, the working groups of the International Scientific Committee (ISC) for Tuna and Tuna-like Species in the North Pacific Ocean are developing MSEs for North Pacific albacore and Pacific bluefin tuna. This is detailed in the IATTC section.

An initial suite of OMs has been developed for WCPO skipjack, based on a grid design, using the stock assessment software MULTIFAN-CL (Fournier et al., 1998). The models retain the spatial structure assumed for the most recent stock assessment (McKechnie, Hampton, Pilling, & Davies, 2016) and have been provisionally allocated to reference and robustness sets. The reference set comprises 72 models and includes alternative settings for steepness and the tag mixing period, as well as a variety of weightings for the various input data sets (tag recapture data, length compositions, etc.) and alternative assumptions for recruitment variability (Table 1). The robustness set remains in development but is anticipated to include alternative assumptions of stock movement and the effects of density-dependent catchability.

The assessment of WCPO skipjack relies heavily on tag release and recapture information for which a substantial quantity of data exists. While both SS and MULTIFAN-CL are broadly similar in structure and the way they implement the maximum-likelihood estimation techniques using automatic differentiation, there are some design implications of MULTIFAN-CL that make it easier to also condition the model on the available tagging data, inferred from long-term tagging programs that have been designed specifically for tropical tuna in the Pacific (Sibert, Hampton, Fournier, & Bills, 1999).

Model conditioning was undertaken through integrated fitting to the catch, effort, size and tagging data using a penalized likelihood approach (see McKechnie et al., 2016). Model validation was based on a collection of indicators of the model fits to data, specifically the maximum gradient of the estimated parameters, which indicates the extent to which a model has converged to a solution, and the plausibility of the model estimates (Table 3). It is intended that additional validation tests will be conducted beyond those described above, including retrospective analyses and likelihood profiling of key parameters. However, a practical approach for achieving this is still being investigated due to the computationally intensive nature of the models.

While an initial set of models has been developed for the reference set, a number of sources of uncertainty remain to be investigated, specifically with regard to movement rates that may be subject to environmental forcing. Some of the known uncertainties could be explored in separate, simpler models, in order to assess their likely impact and to inform a discussion about their possible inclusion as OM scenarios. It has also been proposed to use SEAPODYM (a spatial ecosystem and population dynamics model, Lehodey, Senina, & Murtugudde, 2008) to test some of the stock assessment assumptions, in particular, tag mixing rates and likely population movement patterns.

5.1 | WCPFC lessons learned

WCPFC has held extensive discussions in recent years to identify the key management objectives and the performance indicators to report on them (an interactive web application called PIMPLE was developed to compare performance indicators for various HCRs for skipjack, https://ofp-sam.shinyapps.io/pimple/). These objectives are typically focussed more towards social and economic concerns but include stock sustainability, catch stability and other biological and ecosystem-based concerns. Hence OM design should therefore have the ability to evaluate those social and economic criteria that are important considerations for the MSE; however, there are significant challenges both in collecting data and in developing models at the appropriate resolution to effectively calculate these performance indicators. A lesson learned here is that ideally OMs need to account for social and economic criteria, and sufficient time needs to be provided to the Commission to develop these criteria so that these can inform the development of appropriate OMs. More generally, management objectives need to be considered when designing the OM so that sufficient metrics are generated from the OMs and are available for MPs to address criteria relevant to the objectives.

The WCPFC is distinctive among tRFMOs because of the prevalence of exclusive economic zones within the WCPO and the high proportion of catch taken within them. This leads to complex interactions between the Pacific Island countries and territories (many of which are small island developing states) and the distant-water fishing nations. When calculating the performance indicators to report on members' diverse and often competing objectives, a fundamental concern is the spatial structure of the OMs and the fishing fleets that are represented within them. A key lesson learned here was that the spatial concerns of the Pacific Island countries and territories are crucial to consider within the overall objectives, as a large percentage of the catch is taken in coastal exclusive economic zones. These concerns can only increase over time as climate change is predicted to shift the distribution of tuna between international waters and the exclusive economic zones (EEZs) of the Pacific Island countries (Erauskin-Extramiana et al., 2019).

The generation of "pseudo" tag data is not necessarily new or unique but there have been recent developments to MULTIFAN-CL to include observation error in the generation of historical and future tag data. The skipjack assessment, and therefore the MULTIFAN-CL based OM, is most sensitive to alternative assumptions about tag data, specifically mixing periods and overdispersion values (overdispersion effectively weights the influence of the tag data). Hence, a lesson learned here is incorporating tagging data in the conditioning and accounting for the tag weights is a key consideration for the WCPFC.

The geographical area over which the fishery is managed is substantial and the stocks and fisheries can be subjected to large-scale environmental forcing. Specifically, ENSO events affecting the east–west displacement of fisheries which impact on the resource rents for individual members looking to manage their exclusive economic zone. WCPFC is investigating the potential to use alternative modelling platforms (e.g. SEAPODYM) to estimate movement rates under different ENSO conditions for use in the OM. Examining alternative model structures rather than an assessment-based structure is particularly important when examining spatial considerations, another key lesson learned in WCPFC.

One of the key lessons learned here was that the sensitivity of models to inputs or structural assumptions can be part of model validation, but often individual models have idiosyncratic behaviours which only the modeller is partly aware of. Developing a common format whereas the modeller can disclose these unusual behaviours and describe possible implications should be part of a communication strategy.

6 | INTER-AMERICAN TROPICAL TUNA COMMISSION (IATTC)

6.1 | IATTC temperate tunas

A proof-of-concept MSE was carried out for Pacific bluefin tuna using an OM based on the stock assessment model (SS) developed by the Pacific Bluefin tuna working group (Maunder, 2014). Then, the SS stock assessment model was extended to be used in a Bayesian framework, which requires prior distributions for all parameters, some of which are usually fixed in stock assessments, thus naturally encompasses multiple axes of uncertainty and a range of states of nature. This framework, however, has disadvantages compared to other methods when it comes to incorporating structural uncertainty into the OMs. The International Scientific Committee (ISC) plans to continue developing OMs for Pacific bluefin tuna.

The North Pacific albacore working group started an MSE in 2015. Currently, a series of OMs are in development, using SS as a starting point (Table 1). The working group prioritized the scenarios according to their perceived degree of consequence: (a) high priority (autocorrelation in recruitment and alternative values of steepness, several values of natural mortality, several values of growth parameters); (b) medium priority (time-varying age-specific selectivity, linkage of recruitment to environmental indices, sex-specific natural mortality, time-varying management implementation error); low priority (time-varying growth, time-varying catchability, time-varying size selectivity). In addition, high and medium priority scenarios for future fishing effort are also in development, to model known or unknown shifts of effort from the South Pacific, implying a higher F and larger mortality due to fleets (implied F is larger assuming F is proportional to fleet size) shifting targeting from one area/stock to another area which is more profitable; these scenarios could be developed in the projections examined. The working group presented initial MSE results to managers in 2018 and preliminary results in 2019 showing trade-offs between different management objectives and sensitivity of results to various uncertainties (the presentations are available from the iattc.org).

6.2 | IATTC tropical tunas

MSE modelling for tropical tunas started with bigeye tuna (Maunder, Zhu, & Aires-da-Silva, 2015, Tables 1 and 2). Although tropical tuna fisheries in the eastern Pacific Ocean are de facto multispecies, the management is based on single-species stock assessments for bigeye and yellowfin tuna. Since 2002, the management has relied on effort restrictions for the purse-seine fleet, which takes most of the catches, implemented as seasonal closures – other effort management measures have been in place for longer (Maunder & Deriso, 2007). Bigeye tuna has been the species requiring restrictions, thus controlling fishing of other species—this made it a natural candidate to start the MSE modelling process. For that initial MSE modelling, the OM was based on the bigeye tuna stock assessment model (Aires-da-Silva & Maunder, 2016). Four OMs were defined by changing a single assumption about recruitment, growth or natural mortality relative to the base case (Table 1). Currently, a spatial model is in development for the assessment of bigeye tuna, which will also be the basis for an OM. This model will include the results from tagging studies (Schaefer et al., 2015) and allow the inclusion of uncertainty regarding movement and stock structure.

6.3 | IATTC lessons learned

Stock assessment modelling of bigeye tuna illustrates how gaps in understanding assessment model's behaviour can have implications for conditioning the OM and the MSE more generally. It has been suggested by the stock assessment model that the population of bigeye in recent decades had undergone a regime shift towards greater productivity (the model already assumes that recruitment is largely independent of the biomass, using a steepness parameter equal to 1). However, attempts to understand why concluded that such an increase in recruitment is more likely to be an artefact of the model itself than a reflection of reality (IATTC 2019 meeting presentation). Different aspects of the model were suggested to be responsible for the behaviour but no consensus emerged as to how the model should be modified. A general lesson here is that when fisheries models increase in complexity, it becomes more difficult to identify dynamics that are artefacts of modelling and hard to ensure that these artefacts are not biasing conclusions of the MSE.

7 | COMMON ISSUES IN CONDITIONING, SELECTING AND WEIGHTING OF OM ACROSS THE TRFMOS

The problems that are encountered by each tRFMO have much in common, and some problems are generic to all MSEs. Below we look at some of the issues that could benefit from a strategic approach developed jointly and adopted across tRFMOs. The relative progress on tackling these issues is summarized in Table 4, where more saturated colours indicate greater progress. Table 4 shows that greater effort is required to produce a coherent approach on every issue, and that for some of the issues, meaningful progress is yet to be made.

7.1 | Lack of standard filtering, model validation or model weighting methods

In several case-studies, filtering techniques were used to exclude combinations of fixed parameters or scenarios that led OMs to produce implausible dynamics, that is those inconsistent with historical obser-vations of catches or beliefs, about species biology. Model validation techniques are part of the filtering process. Checking whether the mod-els converge, assessing OM's short-term predictive capacity, and look-ing at residuals tests can help to filter out unsuitable operating models.

However, the issue of how to weight the remaining runs remains debatable. In CCSBT, the approach used Bayesian posteriors. However, this possibly under-represented uncertainty, and for the same reason, likelihood weighting of models was discounted for all dimensions except M. It is important to note that likelihoods were the basis of statistical inference and discounting likelihoods may be problematic. However, if the data were incorrect from the outset, using likelihoods could further bias the outcome. The model/data weighting issue is an important one across tRFMOs.

In other cases, such as IOTC albacore, likelihood weighting was not possible due to issues with data (different sample sizes for length frequencies). Similar issues arise with the Akaike information criterion (AIC), since it cannot be used to choose between models (or assign weights) whenever models are conditioned on different sets of data, which occurs whenever OMs are expected to represent uncertainty arising from conflicting or unreliable data sources (Kell, Kimoto, & Kitakado, 2016).

In addition, data and data weighting issues related to representativeness were identified in several case-studies (e.g. the representativeness—in a sense of mimicking the true state of the stock—of the abundance index series was clearly problematic since slight changes can have large impacts on assessments). Weighting the different components of the likelihoods is an important factor to consider when selecting the reference set for initial OM setup, since putting weights on the abundance index versus the length composition data can produce substantially different results.

Other methodologies for weighting OMs using various statistical methods have been suggested. There is a secondary issue of data conflicts, which needs separate models and data set exclusions for consistency. Then, there is the third issue of structurally different models all fitting reasonably, which is in part what the IWC approach was developed to cover (IWC, 2010). For example, IWC provides four levels of plausibil-ity for scenarios: high, medium, low and no agreement. The default is to assign a medium plausibility. A systematic qualitative procedure for weighting OMs could be an alternative, but currently, no common qualitative methods are being used across tRF-MOs. As such, having a standard methodology developed and adapted across tRFMOs for the purpose of weighting or filtering OMs would be a high priority, as validating OMs that are consistent with the current data would be a prerequisite before the HCR and MP are adopted.

More formal methods to prioritize OMs that combine expert elicitation methods with modelling results have been explored by ICCAT (Levontin, Leach, Holt, Mumford, & Kell, 2015). In this approach, the rank of an OM is linked to beliefs among the stakeholders as well as computational evidence of its importance. In simulation studies involving Atlantic bluefin tuna, scenarios involving inflated catch data ranked particularly highly.

Exploratory modelling for swordfish in ICCAT suggested a procedure for model validation, Table 3, and the full technical specification for the procedure can be found in Kell and Levontin (2019). We recommend extending, refining and standardizing such a procedure across the tRFMOs (Kell & Levontin, 2019). For swordfish, a hindcast procedure (that removed 1–5 recent years of data in order to see if the model is able to predict them) revealed that the OM had poor prediction skill. Another test for the plausibility of OM consisted of an implicit comparison with simpler models. Such comparisons can clarify what features of the model are driving dynamics. These tests revealed that although the r and K parameters were in plausible ranges, the shape of the production function was highly skewed. The template for the procedure suggested in Kell and Levontin (2019) includes a guide for a semi-quantitative assessment of model's performance which is summarized in Table 3. The four validation tests correspond to the rows and the performance is converted to a four-point scale—“failed,” “poor,” “good,” “very good” (columns in Table 3)—in order to communicate the results of validation tests to stakeholders.

7.2 | Parameter confounding/Confounding effects of parameters

Confounding effects of parameter choice is another common problem, which makes it difficult to judge the model's reliability for explaining key uncertainties about the fishery. For example, one key uncertainty is the ability of OMs to explain the numbers of older fish in catches, that is whether the reduction in abundance was due to higher natural mortality or a reduction in vulnerability of older fish to fishing (dome-shaped selectivity versus higher M at older ages, Butterworth et al., 2014). Vulnerability can be difficult to estimate, however, since it is confounded with many processes: recruitment, natural mortality, growth, changes in spatial availability, encounter rate with gear and changes in technical management measures (Caruthers et al., 2017). Multiple choices of parameters can produce similar results; hence, it is imperative to have clarity about weighing alternative assumptions that may be all consistent with historical data. CCSBT has taken a lead on this issue by experimenting with prior weights for alternative hypotheses to explore confounding and minimize data conflicts. Eventually, a higher mortality of older fish was selected as a more plausible hypothesis than a dome-shaped selectivity as older fish were thought to be less available to the fishery.

7.3 | Sensitivity to data updates

It is not uncommon for small changes to the input data or model assumptions to result in substantial differences in the output advice (Collette, 2017). For example, an update of the IOTC albacore assessment resulted in a change in the accepted view about the productivity of the stock; it is now believed to be outside the range of scenarios represented by the OM. Instability of assessment models can slow down the MSE process and/or make it so frequent as to defeat its purpose of adopting an automated long-term management strategy. However, if there is an exceptional circumstances clause, then these situations are easily covered as in the case of CCSBT or North Atlantic Fisheries Organization (NAFO). Having some common references for exceptional circumstances clauses would help tRFMOs in development of their OMs.

On this issue also CCSBT is a leader, having been the first of the tRFMOs to codify a procedure for dealing with unexpected updates to the perception of the stock. Suggestions for determining “exceptional circumstances” developed by CCSBT include the following: annually checking whether key observations are not outside the range of values used in the MSE, whether new knowledge or data warrant a rerun of the MSE; every 3 years conduct an in-depth stock assessment and consider if it is still consistent with the OM; every 6 years review the evidence for the performance of the MP and decide whether it is consistent with expectations from the MSE, and if not redo the MSE (CCSBT, 2013).

7.4 | Will the future resemble the past, especially when the climate is changing?

This is a generic problem affecting all models, especially those that are concerned with the future. This is especially pertinent in the century where volatility and changes are expected to be more dramatic than in the observable past. Climate change could affect fisheries in such a way as to render a lot of historical data uninformative about future dynamics and distribution (Erauskin-Extramiana et al., 2019). We want to make MPs robust to all future sources of variability, whereas data about variability in the past may not be representative of future uncertainties. How to construct plausible future scenarios without having the data from the future is a difficult statistical question. One approach is to increase the level of uncertainty moving forward compared to the past, though the level of the increase is subject to debate. Another approach is to invoke “exceptional circumstances” to revise the MP if future data turn out to be outside the range considered in the trials upon which the MP had been selected. Both options imply that a new set of OMs would include these future observations that appear to be outliers.

Having objective criteria for robustness tests, where some scenarios that are plausible but not very likely, could test whether the HCR and MP are robust to extreme events. This could be done in three iterative steps: (a) having explicit guidelines regarding this and standard procedures for testing a “reference set”; trials that reflect the most plausible hypotheses—and hence used to identify management strategies; (b) deciding when the “reference set” needs to be revisited due to exceptional circumstances, and (c) having an alternative, “robustness set” used to assess whether the management strategy behaves “as expected” under unlikely, but still plausible, scenarios. Robustness trials often involve “nasty” combinations of factors, each of which is “somewhat plausible ” (Punt et al., 2016).

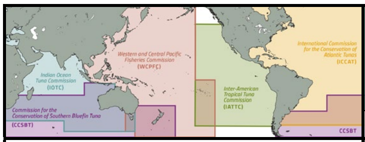
So far none of the tRFMOs have developed methodologies for constructing OMs capable of representing future dynamics under climate change that are based on qualitative futures scenarios that are considered plausible, Table 4. Given increasing evidence that marine ecosystems will be changed in the coming decades by climate change (IPCC) developing an additional set of OM scenarios based on qualitative foresight methods is one option that tRFMOs can pursue. One of the main components of robustness testing should have a climate impact scenario.

7.5 | Under-accounting for uncertainty in data

While data availability/deficiency is examined in some bodies like IWC, most cases in tRFMOs assume catches are known with minimal uncertainty (Table 1). In CCSBT, catch underreporting issues were raised, which meant that the scenarios needed to be rerun, delaying the final outcome. While this was an exceptional case, often such issues are not accounted in other MSEs where issues may be evident as well. For example, the adoption of TACs has led to gross underreporting of catches by some fleets (see discussion on the Chinese Taipei longline fleet, ICCAT, 2015). Methods for accounting for data uncertainty in abundance indices have been developed to some extent, but further work is needed across all tRFMOs. This is one of the key sources of uncertainty driving all assessments, yet most tRFMOs tend to ignore the uncertainty in data, for reasons that range from ontological (“this is how it was always done”) to political (“questioning the data is tantamount to accusing a partner of deception or malpractice”).

Some issues with data are scientific, such as the difficulty of finding out species age. Catch at age is missing across most tRFMOs and length composition data are fairly uncertain. In most cases, this is

TABLE 4 Progress on the outstanding issues identified across the tRFMOs, the table below summarizes how far various case-studies have progressed towards a unified approach for addressing the main issues identified in this review

	IOTC				ICCAT			CCSBT	WCPFC	IATTC	
	ALB	SKJ	YFT, BET	SWO	ALB	BFT	SWO	SBT	SKJ	ALB	BET
1. Filtering, validation, weights of OM	■	■	■	■		▒	■	■	■	■	■
2. Parameter confounding, data conflicts								■			
3. Sensitivity to data updates								■			
4. Climate change/Future dynamics											
5. Underaccounting for uncertainty											
6. Addressing non-stationarity	■	▒	▒	■	▒	▒		▒	▒	▒	
7. Spatial and substock structures		▒	▒			▒					
8. Economics											
9. Process errors							▒				
10. Social, cultural and ecological issues											
11. Communication and inclusive modelling							▒	▒			

Note: The more saturated colour corresponds to the case-studies that are leading the way. Empty cells indicate a lack of attention to the issue or an absence of a tangible solution to the problem.

down-weighted in the grids being developed in the OM. Parameter uncertainty in the OMs (the error in estimated values of key parameters due to inadequate data) may result in strong misrepresentations of the actual stock dynamics. While this is a key issue, such is the case in most fisheries assessments data. Acknowledging this uncertainty by increasing the range of uncertainty through Bayesian priors with a wider range of uncertainty could address this to some extent.

Resolving uncertainties in data is often a sensitive and intrinsically political issue. Methods to smooth over disagreements by blending or aggregating potentially conflicting or misleading data can jeopardize conditioning of OMs and the MSE process as a whole. MSE should be used to resolve disagreements, not to avoid confronting them. A common approach to interrogating data and dealing with data conflicts can be considered across the tRFMOs. A possible approach is specifying data quality criteria, collected based on sampled proportions. In addition, the quality of abundance indices can be evaluated with an objective criterion such as the one used in ICCAT (ICCAT, 2019).

Figure 2 illustrates the gap between the number of the sources of uncertainties that would be important to investigate (35) and an actual number (5) that are feasible to consider within an MSE. Note, this does not imply that we should do 35 axes of uncertainty, but that we should be transparent about the process. Under-accounting for uncertainty is a challenge that none of the tRFMOs had yet managed to address (Table 4). In the near future, the most pragmatic way forward is to contextualize results of the MSE to make under-accounting of uncertainty transparent and understandable to the audiences.

7.6 | OMs not addressing the non-stationarity of processes

Modelling time-varying biological and ecological processes are normally analysed in the stock assessment for several stocks, but it should also be standard for OM design (Table 1). Bjornstad, Nisbet, and Fromentin (2004) show how the dynamics in certain population trajectories (East Atlantic bluefin) can be mimicked with time-varying changes in recruitment exhibiting resonant cohort effects over time. While some tRFMOs (CCSBT and IATTC) are trying to incorporate time-varying dynamics, the majority of OMs do not represent existing non-stationarity in most of the biological/ecological parameters in a consistent manner.

Many however consider time-dependent factors in the standardization of the CPUE indices and in the modelling of recruitment time series (Table 1). Non-stationarity is an issue related to uncertainty in future dynamics related to climate change, and hence, the two issues should be addressed in a linked process.

7.7 | Spatial structure and substocks

Spatial structure of the fishery or substock structure of the species can represent important sources of uncertainty, but these are difficult to model and parameterize. Nevertheless, several OMs have tried to include these, as in the case of Indian Ocean yellowfin tuna, bigeye and skipjack or Atlantic bluefin tuna (Tables 1 and 4). Incorporating spatial resolution into OM design is a new advancement. The Atlantic bluefin example is complicated, but it can be adapted to as fine a resolution as needed to define and model the stock/multi-stock structure and its interaction with the fisheries. Spatial complexity is typically difficult to assess, and ultimately, it is a subjective choice made by the managers and modellers. The only data available to infer movement are tagging data. If a tagging program is not designed appropriately and tag mixing is not achieved, then inferences derived from tagging data on movement can bias the analysis in several ways. For example, movement estimates can bias recruitment estimates by area. A possible way forward is to test a range of scenarios with different spatial structure assumptions. If it is found that spatial assumptions affect the relative performance of the MPs then further work on deciding the appropriate spatial scale for conditioning would be warranted. At the very least then, it would be justified to include alternative spatial assumptions into the set of the robustness trials.

7.8 | Excluding economics

Fishing is an economic activity, and most tRFMOs do not collect economic data. This is problematic as it is important to include economic information in OMs. Economics drives fishing behaviour, and fishers may respond to management in ways that were not considered when the MSE was conducted. It is, therefore, an issue that economics is largely missing currently from the models. The primary interactions with fisheries are modelled using different selectivity functions and catchability. Temporal trends of increasing catchability are used in most tRFMOs; however, such trends present a challenge in conditioning of OMs with regard to profitability and fleet efficiencies (Tidd, Blanchard, Kell, & Watson, 2018). Again, while consensus views were not available on this topic, alternative views were of the opinion that proxies from other sources could cover the need to represent this adequately in the process (Pons, Melnychuk, & Hilborn, 2018). Examples of such proxies are statistics on the variability of catches (the coefficient of variation or interannual variability) that correspond to the socioeconomic objective of *stability* that is often one of the explicit performance criteria for management procedures in the MSE (e.g. ICCAT albacore and swordfish or IOTC skipjack MSEs).

Dealing with economics in the OMs is a challenge for all of the tRFMOs (Table 4), as mentioned above, partial consistent approaches are possible and a common communication strategy with respect to economic uncertainties can and should be developed (Sethi & Thompson, 2002). For example, supply of a product can inadvertently drop prices and as such profit margins if the biomass can become too large and flood a market with product; conversely, the behaviour of fleets to continue fishing a depressed stock at high rates is a behaviour that is market driven (e.g. Atlantic bluefin in the Asian economies can inadvertently get less profitable if the supply is larger than the market capacity, and vice versa).

7.9 | Process errors

Closely related to the non-stationarity issue is the problem of understanding process errors in the model due to temporal and spatial variability in dynamic populations and fisheries. Process errors are key to understanding stock dynamics, to diagnosing problems with models and to constructing OMs. Table 4 illustrates different approaches chosen in different tRFMOs. Simulations where simulated data approach the variability of real data indicate that most assessment models produce imprecise results (Quinn, 2003). It is likely that an MSE that fails to enforce a realistic amount of process error will underestimate the risks associated with MPs, potentially providing support for unsuitable and risky MPs. One should note that process error is a key product of simulations that is often scrutinized; thus, standardizing this process of scrutiny would be worthwhile endeavour across tRFMOs.

7.10 | Excluding ecological and social or cultural aspects

The management remit of the tRFMOs includes ecological and social dimensions, more urgently for fisheries seeking MSC certification that also in theory requires that fisheries are managed in a culturally sensitive manner. There can be a common approach among the tRFMOs towards integrating MSE within a wider risk management framework and that might mean in the future expanding OMs to account for new sources of complexity and including relevant proxy parameters, as with economics. Additionally, there could be a common approach to communicate what is not modelled by the OM, as ICCAT tried with swordfish (Kell & Levontin, 2019).

7.11 | Involving and communicating with stakeholders

Involving stakeholders, especially at the stage of conditioning OMs, is crucial to ensuring that the MSE as a process contributes to managing fisheries in a way that respects both ecological and social aspects of sustainability (Dankel et al., 2012; Goethel et al., 2019; Kell et al., 2015; Miller et al., 2018). Operating in a context where there are many stakeholders with divergent interests, unequal power and degrees of organization makes it extra difficult to ensure fair representation. TRFMOs can learn from each other and from examples of best practice of involving stakeholders in the MSE in other fisheries (Goethel et al., 2019; Rockmann et al., 2012). Currently, stakeholder involvement and communication in the tRFMOs are far from the best practice standards, for example, The New England Fishery Management Council had used open invitation and public workshops for input OMs used in the MSE leading to improved management outcomes (Feeney et al., 2018). CCSBT is ahead of other tRFMOs on stakeholder involvement, but inclusiveness and communication can be significantly improved in all cases, Table 4. A management procedure dialogue is held at the CCSBT with the key stakeholders. First, the OM and MP are developed collectively by the scientists (from contracting and non-contracting parties). Then, the results of the testing of MPs are presented to the Commissioners and an iterative dialogue with the different stakeholders is initiated during which a fine-tuning of MPs takes place. Finally, an agreement is reached through negotiations on what the eventual MP would be.

8 | SUMMARY AND RECOMMENDATIONS

This section discusses some of the overarching principles drawn from the examples above. The first principle is that the uncertainty of the OMs can cover a wide array of scenarios (Table 1) and be conditioned on a variety of methods: either data poor, as in the case of the Indian Ocean skipjack, or data-rich, as in the case of the Indian Ocean albacore. Even in data-rich cases, data conflicts may exist and so more data can increase uncertainty rather than reduce it. As Punt et al. (2016) indicate, the OM conditioning should cover a plausible range of uncertainty, and does not need to accurately reflect the assessment. This arises more from legitimate alternative interpretations of scientific analysis outcomes so that the final “robustness” trials can be judged in the context of all the known uncertainty as exhibited by the OM design.

Using age-structured assessment models (SS, MULTIFAN-CL, VPA) for conditioning the OMs is the most common approach pursued by tRFMOs, with a few examples (such as IOTC skipjack and ICCAT bluefin tuna) using variations on this approach (Table 1). This is based on the concept of conditioning the alternative states of nature represented by different model structures and/or values for the parameters on the available data in order to avoid unrealistic scenarios. Although robustness is required across a range of uncertainties, there is no need to demonstrate that MPs are robust to implausible OMs. The need to appropriately constrain uncertainty ranges is often the reason why the “current best assessment” models are used as OMs as these are consistent with data.

It is important to explicitly address each of the major sources of uncertainty or at least indicate how the modelled uncertainties were selected; this should be a transparent and inclusive process. Punt et al. (2016) state that as best practice, minimally, these uncertainties should be considered (a) process uncertainty, in particular, variation in recruitment about the stock–recruitment relationship; (b) parameter uncertainty relating both to productivity and to the overall size of the resource; and (c) observation error in the data used when applying the management strategy. While points (a) and (b) are explicitly dealt with within most tRFMO setups, point (c) has been addressed to varying degrees in ICCAT, CCSBT and IOTC. Common best practice guidelines for observation error in the data used when setting up the management strategy would help tRFMOs address this issue in a consistent manner. Further, it is crucial that the conditioning of the OMs is adequate to ensure that there is no evidence of systematic misfits to data, unless a particular data source itself is highly uncertain. It is not uncommon for data used in assessments to be in conflict with one another, providing contradictory information. Creative uses of data weighting might be necessary to enable the models to mimic plausible dynamics, but given their potential influence on the conclusions of the MSE such manipulations of data should be documented, examined and communicated to stakeholders. We recommend establishing communication guidelines and developing visualization techniques not only for dealing with data weighting problems but for all sources of uncertainty that can influence conclusions about the robustness of management procedures.

Using a "current best assessment" for conditioning OMs is not integral to MSE, and alternative bottom-up approaches such as the Atlantic bluefin case may be more palatable. Atlantic bluefin is one of the few OMs not based directly on the assessment model used to provide advice, although it is conditioned on historical data and so is itself a candidate for an alternative stock assessment model. In the conditioning OM, outputs were compared, however, to those from the assessment as a form of ad hoc validation. As another example of an OM not based on the current assessment, the Indian Ocean skipjack OM resembles a stock-reduction analysis (that could in principle be used as a data-poor assessment method), but explicitly incorporates spatial structure in the projections. Providing explicit guidelines for data-poor versus data-rich OM scenarios would help tRFMOs ensure best practices under different data constraints.

Process error, mainly through modelling recruitment variability, dominates tuna dynamics. It is crucial to capture this in the OMs in order to realistically project uncertainty. Using an assessment model is pragmatic and allows advancement in the MSE, but the drawback is that the assessment even if it tells us something relevant about the past might be misleading when it comes to the evolution of uncertainties in the future. These issues can be dealt with as stated before using an inclusive process that revisits the "reference set" with an exceptional circumstances clause, "robustness trials" as well might have to be revisited as the OM would be reconditioned.

Basing the OM on the current assessment has the lowest demands on additional data. Attempts made to move away from current assessment models in order to produce more risk comprehensive OMs suffered from lack of evidence to inform scenarios, such as spatial and stock structures. Further, "current best assessment" models allow for the computation of reference points that are familiar to managers, and reference points often play a role in specifying harvest control rules as well as in evaluating whether management procedures are meeting management objectives. When reference points are used in conjunction with OMs that are not based on the same "current best assessment" that was used to calculate stock status, it might be difficult to interpret results of the MSE. For example, if MSY conditions based on "current best assessment" are lower than the MSY implied by the OM which is meant to represent "true" dynamics, the results of the MSE might only tell you which management procedure satisfies biased management objectives, where we perceive an overexploited stock to be "sustainable." An alternative would be to ensure consistency between the criteria for judging the performance of management procedures and OMs.

Similar issues can occur as new data are obtained; for example, new data have been known to lead to major updates in beliefs about the stock carrying capacity or stock–recruitment relationship, transforming quantitatively the reference points and thus shifting targets for management. It might be argued that if new data have implications for reference points, that conditioning of OMs and possibly the entire MSE process should be reviewed. TRFMOs would greatly benefit from a standardized set of explicit guidelines for exceptional circumstances that would call for the OMs to be revised. This would help in providing guidelines that would be consistent across tRFMOs at least for similar species in different oceans.

However, updating an assessment should not necessarily invalidate work done on comparing alternative MPs since the MP should have been tested against a range of uncertainties and shown to be robust (e.g. the MP should be able to prevent overfishing even in the presence of invalid inputs or stressful environmental conditions). If every time an assessment is rerun the MP has to be re-evaluated, then there is no point in conducting MSE. What circumstances should trigger a review of OM conditioning is yet to be worked out and harmonized across the tRFMOs, though CCSBT may provide a good model for how to proceed, as it is becoming the international norm in setting the MPs.

Similarly, there is a need for harmonizing filtering procedures that are used to discard OMs, especially when a grid approach is used. Such approaches for classifying OMs and highlighting those features that seem to be influencing the conclusions of robustness trials for a variety of MPs would enable minimum sets for robustness trials to be determined. The question remains, however, whether the robustness sets can be sufficiently generic across different stocks for this approach to be useful.

So far, the experience of the five tRFMOs has offered many lessons related to model design. A process is underway to create a common set of approaches to deal with various issues that have been identified as specified by the joint tRFMO meeting held in June 2018. And although there is progress on many technical aspects of conditioning, selecting and weighting OM, much less attention has been focused on making the process of enhancing understanding of key aspects of the process by stakeholder groups. Some headway has been made on communicating these results across different forums (Miller et al., 2018), but an understanding of key aspects of the process by stakeholders is still largely missing.

However, some tRFMOs have made substantial advances in this process, like IOTC, and this may be a model to pursue across tRFMOs. In particular, in the case of IOTC bigeye and yellowfin tuna, the OM development is iteratively presented to the working party on methods and tropical tunas. These annual meetings are open to all members and attended by most of them. Feedback (OM-focused) is provided in plenary with guidance for future iterations documented in the meeting reports. Irregular capacity building exercises have been held, and a science-management dialogue precedes Commission meetings to further provide updates and solicit feedback (primarily on management objectives and performance trade-offs). Similar processes are employed by CCSBT, and using these as common approaches could benefit ICCAT, IATTC and WCPFC. More consultation is always better than less, specifically if some countries have lesser capacity than others, but this may possibly be a way forward as improved communication and a shared understanding of the stock status and uncertainties in the OM is often identified as one of the key goals and benefits from MSE, and is a key step in subsequent process of MSE and MP development (Levontin et al., 2017, 2020). In order to increase transparency and improve stakeholder input, it is desirable to establish intermediary bodies that include the participation of scientists, managers and stakeholders. These intermediary bodies can address conceptual and technical aspects of the MSE, in both formal and informal settings, to increase stakeholder engagement in—and support of—the process (Miller et al., 2018).

We conclude by noting that this review revealed a challenging but quickly evolving landscape of OM developments across all of the tRFMOs. The progress is uneven, and most of the urgent issues we identified and analysed are yet to be addressed in a consistent manner, see Table 4. We are hopeful, however, that the process of harmonizing and standardizing methodologies for conditioning OMs across the tRFMOs will be catalysed by this review that will help assimilate the lessons learned and to focus on the key issues that we identified.

ACKNOWLEDGEMENTS

We would like to thank the ABNJ process in FAO (in particular, Alejandro Anganuzzi) for funding the original meetings across tRFMOs to discuss issues relevant to these assessments. In addition, we would like to thank FAO-ABNJ for supporting two small group meetings to complete key analyses for the development of this paper. We thank Dr Pons (University of Washington) for help with Figure 1. Finally, we thank Dr Andre Punt, Dr Doug Butterworth, Dr Dale Kolody, Dr Campbell Davies and Dr Tom Carruthers for invaluable comments on an earlier version of this document; however, the views expressed in this document reflect those of the authors alone.

DATA AVAILABILITY STATEMENT

The work presented here is primarily a review of analysis done across tRFMOs. The data that support the findings of this study are available on request from the identified tRFMOs directly. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Rishi Sharma https://orcid.org/0000-0002-3818-9393
Polina Levontin https://orcid.org/0000-0001-8229-4054
Yimin Ye https://orcid.org/0000-0001-9902-3699
Arni Magnusson https://orcid.org/0000-0003-2769-6741

REFERENCES

ABTMSE. (2018). https://github.com/ICCAT/abft-mse/find/master

Aires-da-Silva, A., & Maunder, M. (2014). Status of Bigeye Tuna in the Eastern Pacific Ocean in 2013 and Outlook for the Future. Inter-American Tropical Tuna Commission. *Stock Assess. Rep.*, *15*, 25–39. Retrieved from http://www.iattc.org/StockAssessmentReportsENG.htm

Anon. (2017). *Specifications for MSE trials for bluefin tuna in the North Atlantic. GBYP Core Modelling Group. ICCAT Atlantic Wide Research Programme for Bluefin Tuna*. Retrieved from https://github.com/ICCAT/abftmse/tree/master/Manuals_and_design_documents/TrialSpecifications.docx

Bentley, N., & Adam, S. (2014). *Management Strategy Evaluation of Indian Ocean Skipjack: First steps. IOTC 2014-WPTT 16-39*. Retrieved from https://www.iotc.org/documents/management-strategy-evaluation-indian-ocean-skipjack-tuna-first-steps

Bentley, N., & Langley, A. (2012). Feasible stock trajectories: A flexible and efficient sequential estimator for use in fisheries management procedures. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 161–177. https://doi.org/10.1139/f2011-143

Bergh, M. O., & Butterworth, D. S. (1987). Towards rational harvesting of the South African anchovy considering survey imprecise and recruitment variability. *South African Journal of Marine Science*, *5*, 937–951. https://doi.org/10.2989/025776187784522702

Bjornstad, O. M., Nisbet, R. M., & Fromentin, J. M. (2004). Trends and cohort resonant effects in age-structured Populations. *Journal of Animal Ecology*, *73*, 1157–1167. https://doi.org/10.1111/j.0021-8790.2004.00888.x

Butterworth, D. S., Rademeyer, R. A., Brandao, A., Geromont, H. F., & Johnston, S. J. (2014). Does selectivity matter. *Fisheries Research*, *158*, 194–204. https://doi.org/10.1016/j.fishres.2014.02.004.

Carruthers, T., & Butterworth, D. S. (2018). ATL-MSE: An R Package for Atlantic Bluefin Tuna management Strategy Evaluation. ICCAT SCRS 74: 3553–3559. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Carruthers, T., Kimoto, A., Powers, J. E., Kell, L., Butterworth, D. S., Lauretta, M. V., & Kitakado, T. (2016). Structure and estimation framework of Atlantic Bluefin Tuna Operating Models. *ICCAT SCRS*, 72: 1782–1795. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Carruthers, T., Power, J. E., Lauretta, M. V., Di Natale, A., & Kell, L. (2016). A summary of data to inform operating models in management strategy evaluation of Atlantic Bluefin Tuna. *ICCAT SCRS*, 72: 1796–1807. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Caruthers, T., Kell, L., & Palma, C. (2017). Accounting for uncertainty due to data processing in virtual population analysis using Bayesian multiple imputation. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*, 883–896. https://doi.org/10.1139/cjfas-2017-0165.

CCSBT (2013). *Attachment 10 of the Report of the Eighteenth Meeting of the Scientific Committee*. Retrieved from https://www.ccsbt.org/en/content/reports-past-meetings

CCSBT (2019). *Report of the Tenth Operating Model and Management Procedure Technical Meeting*. Retrieved from https://www.ccsbt.org/en/content/reports-past-meetings

Chang, X. H. (2014). *Robust output feedback H-infinity control and filtering for uncertain linear systems*. Berlin, Heidelberg: Springer Science & Business.

Collette, B. B. (2017). Bluefin tuna science remains vague. *Science*, *358*, 879–880. https://doi.org/10.1126/science.aar3928.

Dankel, D. J., Aps, R., Padda, G., Rockmann, C., van der Sluijs, J. P., Wilson, D. C., & Degnbol, P. (2012). Advice under uncertainty in the marine system. *ICES Journal of Marine Science*, *69*, 3–7. https://doi.org/10.1093/icesjms/fsr179.

De Oliveira, J. A. A., & Butterworth, D. S. (2004). Developing and refining a joint management procedure for the multispecies South African pelagic fishery. *ICES Journal of Marine Science*, *61*, 1432–1442. https://doi.org/10.1016/j.icesjms.2004.09.001.

Dichmont, C. M., & Brown, I. W. (2010). A case study in successful management of a data-poor fishery using simple decision rules: The Queensland spanner crab fishery. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science*, *2*, 1–13. https://doi.org/10.1577/C08-034.1.

Dichmont, C. M., Deng, A., & Punt, A. E. (2008). Beyond biological performance measures in Management Strategy Evaluation: Bringing in economics and the effects of trawling on the benthos. *Fisheries Research*, *94*, 238–250. https://doi.org/10.1016/j.fishres.2008.05.007.

Dichmont, C. M., Ellis, N., & Bustamante, R. H. (2013). Evaluating marine spatial closures with conflicting fisheries and conservation objectives. *Journal of Applied Ecology*, *50*, 1060–1070. https://doi.org/10.1111/1365-2664.12110.

Erauskin-Extramiana, M., Arrizabalaga, H., Hobday, A. J., Cabré, A., Ibaibarriaga, L., Arregui, I., … Chust, G. (2019). Large-scale distribution of tuna species in a warming ocean. *Global Change Biology*, *25*, 2043–2060. https://doi.org/10.1111/gcb.14630.

Feeney, R. G., Boelke, D. V., Deroba, J. J., Gaichas, S., Irwin, B. J., & Lee, M. (2018). Integrating management strategy evaluation into fisheries management: Advancing best practices for stakeholder inclusion based on an MSE for Northeast US Atlantic herring. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*, 1103–1111. https://doi.org/10.1139/cjfas-2018-0125.

Food and Agriculture Organization (FAO) (1996). Precautionary Approach to Capture Fisheries: Part 1. Guidelines on the precautionary approach to capture fisheries and species introductions. Retrieved from http://www.fao.org/3/W1238E01.htm#ch1

Fournier, D. A., Hampton, J., & Sibert, J. R. (1998). MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*, 2105–2116. https://doi.org/10.1139/f98-100.

Geromont, H. F., De Oliveira, J. A. A., Johnston, S. J., & Cunningham, C. L. (1999). Development and application of management procedures for fisheries in southern Africa. *ICES Journal of Marine Science*, *56*, 952–966. https://doi.org/10.1006/jmsc.1999.0536.

Goethel, D. R., Lucey, S. M., Berger, A. M., Gaichas, S. K., Karp, M. A., Lynch, P. D., & Walter, J. F. (2019). Recent advances in management strategy evaluation: Introduction to the special issue "Under pressure: Addressing fisheries challenges with Management Strategy Evaluation". *Canadian Journal of Fisheries and Aquatic Sciences*, *76*, 1689–1696. https://doi.org/10.1139/cjfas-2019-0084.

Hilborn, R. (2003). The state of the art in stock assessment: Where we are and where we are going. *Scientia Marina*, *67*, 15–20. https://doi.org/10.3989/scimar.2003.67s115.

Hillary, R., Preece, A. L., Davies, C., Kurota, H., Sakai, O., Itok, T., … Branch, T. A. (2016). A scientific alternative to moratoria for rebuilding depleted international tuna stocks. *Fish and Fisheries*, *17*, 469–482. https://doi.org/10.1111/faf.12121.

Hillary, R. M., Preece, A. L., Davies, C. R., Takahashi, N., Sakai, O., & Itoh, T. (2017). Reconditioning of the CCSBT Operating Model in 2017. CCSBT-ESC/1708/14. Retrieved from www.ccsbt.orgESC22_14_AUandJP_ReconditionOM

ICCAT (2015). *Report of the 2015 ICCAT bigeye tuna data preparatory meeting*. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

ICCAT (2019). *Report of the 2019 ICCAT working group on stock assessment methods (WGSAM)*. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

ICES (2019). Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2). *ICES Scientific Reports* 1:33. 162 pp. https://doi.org/10.17895/ices.pub.5331

International Whaling Commission (IWC) (2010). Report of the Scientific Committee, Annex D. Report of the Sub-committee on the Revised Management Procedure. *Journal of Cetacean Research and Management*, *12*, 89–116.

IOTC. (2018). *Update on IOTC Yellowfin Tuna Operating Model Development October 2018*. IOTC-WPM 09–10. 46 pp. Retrieved from: https://www.iotc.org/documents/WPM/09/10

IOTC (2019). *Update on IOTC Bigeye Tuna MSE Operating Model Development October 2019. IOTC-2019-WPM10-08*. Retrieved from https://www.iotc.org/meetings/10th-working-party-methods-wpm10

IOTC (2019). *Updates on the Indian Ocean swordfish management strategy evaluation: Initial testing of candidate management procedures. IOTC-2019-WPM10-12_Rev2*. Retrieved from https://www.iotc.org/meetings/10th-working-party-methods-wpm10

IOTC (2019). *Update on IOTC Yellowfin Tuna Management Strategy Evaluation Operating Model development Oct 2019. IOTC-2019-WPM10-09*. Retrieved from https://www.iotc.org/meetings/10th-working-party-methods-wpm10

IOTC (2019d). *Report of the 10th Session of the IOTC Working Party on Methods*. Pasaia, Spain 17–19 October 2019. IOTC–2019–WPM10–R[E]: 36 pp. Retrieved from https://www.iotc.org/meetings/10th-working-party-methods-wpm10

Kell, L. T., Kimoto, A., & Kitakado, T. (2016). Evaluation of the prediction skill of stock assessment using hindcasting. *Fisheries Research*, *183*, 119–127. https://doi.org/10.1016/j.fishres.2016.05.017.

Kell, L. T., & Levontin, P. (2019). Modelling approaches: support to ICCAT North Atlantic Swordfish MSE Process. Collect. *Vol. Sci. Pap. ICCAT*, 76(3): 114–152. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Kell, L. T., Levontin, P., Davies, C. R., Harley, S., Kolody, D. S., Maunder, M. N., … Sharma, R. (2015). The quantification and presentation of risk. In C. T. T. Edwards, & D. J. Dankel (Eds.), *Management science in fisheries: An introduction to simulation-based methods* (p. 460). Oxford, UK: Routledge.

Kell, L. T., & Mosqueira, I. (2017). Conditioning operating models on data and knowledge and rejecting and weighting of hypotheses. *Collect. Vol. Sci. Pap. ICCAT*, 73(9): 3000–3008. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Kimura, D. K., Balsiger, J. W., & Ito, D. H. (1984). Generalized Stock reduction Analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *41*, 1325–1333. https://doi.org/10.1139/f84-162.

Lallemand, P., Bergh, M., Hansen, M., & Purves, M. (2016). Estimating the economic benefits of MSC certification for the South African hake trawl fishery. *Fisheries Research*, *182*, 98–115. https://doi.org/10.1016/j.fishres.2016.02.003.

Leach, A. W., Levontin, P., Holt, J., Kell, L. T., & Mumford, J. D. (2014). Identification and prioritization of uncertainties for management of Eastern Atlantic bluefin tuna (Thunnus thynnus). *Marine Policy*, *48*, 84–92. https://doi.org/10.1016/j.marpol.2014.03.010.

Lehody, P., Senina, I., & Murtugudde, R. (2008). A spatial ecosystem and populations dynamics model (SEAPODYM)–Modeling of tuna and tuna-like populations. *Progress in Oceanography*, *78*, 304–318. https://doi.org/10.1016/j.pocean.2008.06.004.

Levontin, P., Walton, J. L., Kleineberg, J., Barons, M., French, S., Aufegger, L., … Houssineau, J. (2020). *Visualising uncertainty: A short introduction*. London, UK: AU4DM.

Levontin, P., Baranowski, P., Leach, A. W., Bailey, A., Mumford, J. D., Quetglas, A., & Kell, L. T. (2017). On the role of visualization in fisheries management. *Marine Policy*, *78*, 114–121. https://doi.org/10.1016/j.marpol.2017.01.018.

Levontin, P., Leach, A. W., Holt, J., Mumford, J. D., & Kell, L. T. (2015). Specifying and weighting scenarios for MSE robustness trials. Collect. *Vol. Sci. Pap. ICCAT*, 71(3): 1326–1343. Retrieved from https://www.iccat.int/en/pubs_CVSP.html

Maunder, M. N. (2014). *Management strategy evaluation (MSE) implementation in Stock Synthesis: Application to Pacific Bluefin Tuna. Inter-American Tropical Tuna Commission, Scientific Advisory Committee. 5th meeting. Document SAC-05-10b*. Retrieved from: http://www.iattc.org/Meetings/Meetings2014-2018ENG.htm

Maunder, M., & Deriso, R. (2007). Using indicators of stock status when traditional reference points are not available: Evaluation and application to skipjack tuna in the Eastern Pacific Ocean. *IATTC Stock Assessment Report Stock assessment report*, *8*, 229–246. Retrieved from https://www.researchgate.net/publication/237603497_USING_INDICATORS_OF_STOCK_STATUS_WHEN_TRADITIONAL_REFERENCE_POINTS_ARE_NOT_AVAILABLE_EVALUATION_AND_APPLICATION_TO_SKIPJACK_TUNA_IN_THE_EASTERN_PACIFIC_OCEAN

Maunder, M. N., Zhu, J., & Aires-da-Silva, A. (2015). *Preliminary management strategy evaluation to evaluate the IATTC interim reference points and proposed harvest control rules. Inter-American Tropical Tuna Commission, Scientific Advisory Committee. 5th meeting. Document SAC-05-10b*. Retrieved from: http://www.iattc.org/Meetings/Meetings2014-2018ENG.htm

McKechnie, S., Hampton, J., Pilling, G. M., & Davies, N. (2016). *Stock assessment of skipjack tuna in the western and central Pacific Ocean. WCPFC-SC12-SA-WP-04*. Retrieved from: https://www.wcpfc.int/node/27490

Methot, R. D., & Wetzel, C. (2013). Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, *142*, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012.

Miller, S. K., Anganuzzi, A., Butterworth, D. S., Davies, C. R., Donovan, G. P., Nickson, A., … Restrepo, V. (2018). 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.

Mosqueira, I. (2018). *Update on the development of MSE analysis of candidate Management Procedures for Indian Ocean albacore. IOTC-2018-SC21-17_Rev1*. Retrieved from https://www.iotc.org/meetings/21st-scientific-committee-sc21

Pons, M., Melnychuk, C. M., & Hilborn, R. (2018). Management effectiveness of large pelagic fisheries in the high seas. *Fish and Fisheries*, *19*, 260–270. https://doi.org/10.1111/faf.12253.

Prager, M. H. (1994). A suite of extensions to a nonequilibrium surplus-production model. *Fishery Bulletin*, *92*, 374–389.

Preece, A. L., Davies, C. R., & Hillary, R. M. (2018). Report on the tuna RFMO MSE working group meeting. CSIRO, Australia. Retrieved from www.ccsbt.orgsystemfilesESC23_21_AU_TRF

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Olivera, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best Practices. *Fish & Fisheries*, *17*, 303–334. https://doi.org/10.1111/faf.12104.

Punt, A. E., & Donovan, G. (2007). Developing management procedures that are robust to uncertainty: Lessons from the International Whaling Commission. *ICES Journal of Marine Science*, *64*, 603–612. https://doi.org/10.1093/icesjms/fsm035.

Quinn, T. (2003). Ruminations on the development and future of population dynamic models in fisheries. *Natural Resource Modelling*, *16*, 341–392. https://doi.org/10.1111/j.1939-7445.2003.tb00119.x.

Richards, L. J., & Maguire, J. J. (1998). Recent international agreements and the precautionary approach: New directions for fisheries management science. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*, 1545–1552. https://doi.org/10.1139/f98-043.

Richardson, D. E., Marancik, K. E., Guyon, J. R., Lutcavage, M. E., Galuardi, B., Lam, C. H., … Hare, J. A. (2016). Discovery of spawning grounds reveals diverse migration strategies in Atlantic bluefin tuna (*Thunnus thynnus*). *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3299–3304. https://doi.org/10.1073/pnas.1525636113.

Rockmann, C., Ulrich, C., Dreyer, M., Bell, E., Borodzicz, E., Haapasaari, P., … Pastoors, M. (2012). The added value of participatory modelling in fisheries management: What has been learnt? *Marine Policy*, *36*, 1072–1085. https://doi.org/10.1016/j.marpol.2012.02.027.

Rooker, J. R., Secor, D. H., Metrio, G. D., Schloesser, R., Block, B. A., & Nielsen, J. D. (2008). Natal homing and connectivity in Atlantic Tuna Bluefin Populations. *Science*, *332*, 742–744. https://doi.org/10.1126/science.1161473.

Schaefer, K., Fuller, D., Hampton, J., Caillot, S., Leroy, B., & Itano, D. (2015). Movements, dispersion, and mixing of bigeye tuna (*Thunnus obesus*) tagged and released in the equatorial Central Pacific Ocean, with conventional and archival tags. *Fisheries Research*, *161*, 336–355. https://doi.org/10.1016/j.fishres.2014.08.018.

Sethi, S., & Thompson, G. (2002). *Optimal Control Theory. Applications to Management Science and Economics*. Boston: Springer Verlag.

Sibert, J., Hampton, J., Fournier, D., & Bills, P. J. (1999). An advection–diffusion–reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Canadian Journal of Fisheries and Aquatic Sciences*, *56*, 925–938. https://doi.org/10.1139/f99-017.

Stemle, A., Uchida, H., & Roheim, C. A. (2016). Have dockside prices improved after MSC certification? analysis of multiple fisheries. *Fisheries Research*, *182*, 116–123. https://doi.org/10.1016/j.fishres.2015.07.022.

Tidd, A., Blanchard, J. L., Kell, L., & Watson, R. A. (2018). Predicting global tuna vulnerabilities with spatial, economic, biological and climatic considerations. *Scientific Reports*, *8*, 10572. https://doi.org/10.1038/s41598-018-28805-z.

How to cite this article: Sharma R, Levontin P, Kitakado T, et al. Operating model design in tuna Regional Fishery Management Organizations: Current practice, issues and implications. *Fish Fish*. 2020;00:1–22. https://doi.org/10.1111/faf.12480