EVALUATION OF PERFORMANCE OF CANDIDATE MANAGEMENT PROCEDURES FOR THE NORTH ATLANTIC SWORDFISH MANAGEMENT STRATEGY EVALUATION

A. Hordyk[1], C. Brown[2], T. Carruthers[1,3], R. Coelho[4,5], N. Duprey[6], K. Gillespie[7], A. Hanke[7], S. Miller[8], L. Rueda[9], D. Rosa[4,5], M. Schirripa[2]

SUMMARY

A variety of candidate management procedures (CMPs) were developed and tested in the management strategy evaluation (MSE) of the northern swordfish fishery. The performance of the CMPs was evaluated across a set of nine reference operating models and five robustness models. The reference operating models spanned the uncertainty in the natural mortality rate and the steepness of the Beverton-Holt stock-recruit relationship. The robustness tests considered additional uncertainties, including hyperstability in the indices of abundance, the influence of climate change on future recruitment, and the impact of illegal, unreported, and unregulated fishing. A set of performance metrics (PMs) were defined to calculate the performance of the CMPs. The PMs included metrics related to the biological status of the stock, the probability of breaching the limit reference point, and the magnitude and variability of the total allowable catch recommendations from each CMP. The results from the MSE are presented in an interactive application. This paper provides an overview of the technical specifications of the north Atlantic swordfish MSE and describes the figures and tables that are available in the interactive application.

RÉSUMÉ

Plusieurs procédures de gestion potentielles (CMP) ont été développées et testées dans le cadre de l'évaluation de la stratégie de gestion (MSE) de la pêcherie d'espadon du Nord. La performance des CMP a été évaluée sur un ensemble de neuf modèles opérationnels de référence et cinq modèles de robustesse. Les modèles opérationnels de référence couvraient l'incertitude dans le taux de mortalité naturelle et la pente de la relation stock-recrutement de Beverton-Holt. Les tests de robustesse étudiaient des incertitudes additionnelles, dont l'hyperstabilité des indices d'abondance, l'influence du changement climatique sur le futur recrutement et l'impact de la pêche illicite, non déclarée et non réglementée. Un ensemble de mesures de performance (PM) ont été définies pour calculer la performance des CMP. Les PM incluaient des mesures liées à l'état biologique du stock, la probabilité de dépasser le point de référence limite et l'ampleur et la variabilité du total admissible de captures recommandé par chaque CMP. Les résultats de la MSE sont présentés dans une application interactive. Ce document offre un aperçu des spécifications techniques de la MSE de l'espadon de l'Atlantique Nord et décrit les figures et tableaux qui sont disponibles dans l'application interactive.

RESUMEN

En la evaluación de la estrategia de ordenación (MSE) de la pesquería de pez espada del norte se desarrollaron y probaron diversos procedimientos de ordenación candidatos (CMP). El desempeño de los CMP se evaluó a través de un conjunto de nueve modelos operativos de referencia y cinco modelos de robustez. Los modelos operativos de referencia abarcaban la

[1] Blue Matter Science, 2150 Bridgman Avenue, North Vancouver, BC, Canada, V7P2T9 adrian@bluematterscience.com
[2] NOAA Fisheries, Southeast Fisheries Center, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA.
[3] Institute for the Oceans and Fisheries, AERL, University of British Columbia, Vancouver, British Columbia, Canada
[4] Instituto Portugues do Mar e da Atmosfera (IPMA), 8700-305 Olhão, Portugal
[5] Centro de Ciências do Mar do Algarve (CCMAR), Univ. Algarve, 8005-139 Faro, Portugal
[6] Fisheries & Oceans Canada, Fish Population Science, 200-401 Burrard St, Vancouver, BC, Canada
[7] Fisheries & Oceans Canada, Biological Station, 531 Brandy Cove Road, St. Andrews, NB E5B 2L9, Canada
[8] The Ocean Foundation, 1320 19th St., NW, 5th Floor, Washington, DC 20036, United States
[9] Instituto Español de Oceanografia (IEO-CSIC). Centro Oceanográfico de Málaga. Puerto Pesquero s/n. 29640 Fuengirola

incertidumbre de la tasa de mortalidad natural y la inclinación de la relación stock-reclutamiento de Beverton-Holt. Las pruebas de robustez tuvieron en cuenta incertidumbres adicionales, como la hiperestabilidad en los índices de abundancia, la influencia del cambio climático en el reclutamiento futuro y el impacto de la pesca ilegal, no declarada y no reglamentada. Se definió un conjunto de métricas de desempeño (PM) para calcular el desempeño de los CMP. Las PM incluían métricas relacionadas con el estado biológico del stock, la probabilidad de superar el punto de referencia límite y la magnitud y variabilidad del total admisible de capturas recomendadas por cada CMP. Los resultados de la MSE se presentan en una aplicación interactiva. Este documento ofrece una visión general de las especificaciones técnicas de la MSE del pez espada del Atlántico norte y describe las figuras y tablas disponibles en la aplicación interactiva.

KEYWORDS

MSE, Simulation, Performance Metrics

1 Introduction

The SCRS Swordfish Species Group has been developing a management strategy evaluation (MSE) framework for North Atlantic swordfish (N-SWO) for a decade. In 2009, ICCAT called for development of a limit reference point for swordfish (Rec. 09-02), and the Commission adopted 0.4*B_{MSY} as the interim limit reference point in 2013 (Rec. 13-02). Recommendation 13-02 also tasked the SCRS with development of a harvest control rule for N-SWO. In 2015, the Commission called for adoption of a management procedure (MP) based on an MSE for 8 priority stocks, including N-SWO (Rec. 15-07). In 2017, the SCRS developed an integrated, sized-structured stock assessment model for N-SWO on which a future MSE would be based. Funds were provided by the Commission in 2018 to develop the simulation framework, and following initial work by the SCRS, an MSE expert was contracted in 2019 to develop the N-SWO MSE. MSE development by the SCRS then began in earnest. The Commission adopted conceptual management objectives for N-SWO in 2019 (Res. 19-14) to help guide MSE development. In 2022, the SCRS carried out a new stock assessment in which the base case model was modified to incorporate discard mortality of undersized fish, and the MSE was updated with this new model. MSE development has continued in 2023, incorporating feedback provided by Panel 4 at its March and July meetings.

This paper provides an overview of the MSE process developed for N-SWO, describes the candidate management procedures (CMPs) developed in the MSE and the performance metrics that are used to evaluate the performance of the CMPs, and summarizes the key results.

2 Methods

The technical specifications of the N-SWO MSE process, including details on the conditioning of the operating models, and the assumptions for the projection period, and the definition of the performance metrics, are described in **Appendix A**. In this section we provided a brief overview of the key components of the MSE process, and refer readers to **Appendix A** for further details.

2.1 Operating Models

Operating models for the N-SWO MSE were based on the 2022 stock assessment (Anon., 2022), conducted with the Stock Synthesis 3 (SS3) assessment software (Methot & Wetzel, 2013). The operating models (OMs) were classified into two categories: the Reference Set, which spanned the key uncertainties in the 2022 stock assessment, and the Robustness OMs, a subset of the Reference Set that were modified to account for additional potential uncertainties.

2.1.1 Reference Operating Models

The SCRS Swordfish Species Group (hereafter referred to as Group) identified the natural mortality rate (*M*) and the steepness of the Beverton-Holt stock-recruit relationship (*h*) as the primary axes of uncertainty that had the greatest impact of the estimated stock dynamics and the performance of candidate management procedures (Hordyk, 2021). Three values were selected for each parameter (*M*=0.1, 0.2, 0.3 and *h*=0.69, 0.80, 0.88), and nine operating models were conditioned with these assumed values. These nine OMs are referred to as the Reference OMs. One OM of the Reference Set (M=0.2 & h=0.88) shared identical assumptions with the 2022 stock assessment.

The estimated magnitude of the stock varied considerably across the nine OMs, with almost a six-fold difference between the smallest magnitude (SB0=69,484 t, *M*=0.3, *h*=0.88) and the largest (SB0=416,034 t, *M*=0.1, *h*=0.69; **Table 1**). The estimated stock status in the terminal year (2020) in terms of SB/SB_{MSY} ranged from 1 (M=0.2, h=0.69) to 2.06 (M=0.3, h=0.88; **Table 1**). The estimates of F/F_{MSY} in the terminal year ranged from 0.46 to 0.90 for these same models (**Table 1**).

Each OM in the Reference Set has 50 simulations, with each simulation sharing identical dynamics during the historical period (based on the maximum likelihood estimates of the SS3 model), and stochastic process and observation error during the projection period.

2.1.2 Robustness Operating Models

A set of Robustness OMs were developed to evaluate the impact of additional uncertainties that were not considered in the Reference Set. The fifth OM from the Reference Set (M=0.2, h=0.8; **Table 1**), referred to as R0, was selected to be used as the base case for the development of robustness OMs. This model is the one that uses the middle assumed values in terms of M and *h*. Five Robustness OMs were developed by modifying the assumptions of R0 to consider additional uncertainties for the historical and projection periods. **Table 2** provides a summary of the Robustness OMs. More details on the Robustness OMs are available in **Appendix A**.

2.2 Performance Metrics

The N-SWO MSE currently includes 11 key performance metrics as a benchmark for evaluation of the Commission's selected management objectives (**Table 3**). These performance metrics were developed based on input received from Panel 4 in March and June 2023. Further details on the Performance Metrics are available in **Appendix A**.

The performance metrics are used to summarise the performance of the candidate management procedures. For the Reference Set, the results were combined across the nine operating models and then the performance metrics were calculated. For example, PGK_{short} was calculated by first combining the results from the 50 simulations from each OM in the Reference Set together, resulting in 450 simulations, and then calculating the proportion of data points from the first 10 years of the projection period where $SB > SB_{MSY}$ and $F < F_{MSY}$.

2.3 Candidate Management Procedures

The Group worked collaboratively to develop and test a number of CMPs. All CMPs currently assume a 3-year management cycle and calculate a single total allowable catch (TAC) for the North Atlantic.

The CMPs all use indices of relative abundance and the historical catches as the primary data sources to determine the total allowable catch (TAC) in each management cycle. Eight indices of relative abundance were projected: 7 fleet specific indices (CDN, CHT, JPN, MOR, POR, SPN and USA), as well as an index developed from the combined fleet data (Combined Index). Most CMPs used the Combined Index, which together with the catch data, was lagged by 1-year (e.g., TAC advice generated in 2023 for implementation in 2024 used data up to and including 2022). The CMPs that used the fleet-specific indices used a 2-year lag.

A brief description of the fifteen CMPs developed during the N-SWO MSE process are provided in **Table 4** and a fuller description is provided in **Appendix B**. All the fifteen CMPs were tuned across the Reference Set OMs to three levels for the PGK_{short} performance metric: 0.51, 0.60, and 0.70 referred to as tuning targets a, b, and c respectively.

A subset of five CMPs were selected by the Group as the best candidates that spanned the trade-offs between the various performance metrics (**Table 5**). The three tuning variants for each of these CMPs results in 18 different CMP configurations (CMP + tuning level; **Table 5**).

2.4 Presentation of Results

An interactive Shiny application was developed for examining the MSE results. The app is currently available online (https://shiny.bluematterscience.com/app/swomse). The app can also be run locally by installing the N-SWO MSE R package (https://github.com/ICCAT/nswo-mse) and running the command *Shiny()* after loading the package (*library(SWOMSE)*).

The results of the N-SWO process are summarized as the performance metric values calculated across the Reference Set and the individual Robustness OMs. A series of plots also shows the performance of the CMPs over time during the projection period. The app shows the results for the five selected CMPs (**Table 5**). The results for all CMPs developed in the MSE process are available in the NSWO-MSE R package.

The results in this paper provide an overview of the performance of the five selected CMPs (**Table 5**), and describe examples of the figures and tables that are presented in the app. We refer readers to the app for more information on the MSE results.

3 Results

3.1 Time-Series Plot

Time-series plots show the trends in F/F_{MSY}, SB/SB_{MSY}, and the TAC over the 30-year projection period for each CMP configuration. These plots are useful for providing a graphical interpretation of the performance metrics that are used to summarise the performance of each CMP configuration.

Figure 1 shows a time-series plot for one configuration of the MCC7 CMP for the Reference Set of operating models (MCC7_c). The corresponding performance metric values are shown in a table in the corner of each plot. For example, this variant of the MCC7 CMP was tuned to achieve a value of 0.70 for the probability of being in the green space of the Kobe matrix in the first 10 years (PGK_short; blue dashed line). The probability of remaining in the green space over the medium (green dashed line) and long (red dashed line) time-periods are also reported in the table. The probability of being in the green quadrant of the Kobe matrix over the entire projection period can be calculated as the mean of PGK_short, PGK_med, and PGK_long.

This CMP configuration avoided breaching the limit reference point in any of the simulations (nLRP =1). The TACs that were recommended by this CMP configuration were relatively stable over the projection period. Time-series plots like this are available in the Shiny app for each CMP configuration, as well as for the results from the Robustness operating models.

3.2 Quilt Plot

A quilt plot (or quilt table) provides quantitative values for the performance of CMPs using both probability values of achieving performance metrics as well as absolute values for change between management cycles and TAC within various timeframes, assuming the same set of conditions among all CMPs (**Figure 2**). Colour scale is used to provide a visual guide for performance with darker shades of blue indicating better performance. The Shiny app provides sorting and filtering tools where the user can set probability, TAC, and variance thresholds and then sort CMPs by their chosen performance metric.

The quilt plots can be used to filter CMPs based on minimum performance criteria, or compare the performance of CMPs across the reference and robustness operating models. For example, the 15 configurations of the five selected CMPs had a low probability (≤ 4%) of breaching the limit reference point (LRP) across the nine OMs in the Reference Set (**Table 6**). However, the probability of breaching the LRP was more variable across these CMPs in the robustness tests. For example, the probability of breaching the LRP was highest for Robustness test R3b, which assumed the first 15 years of the projection period had an environmentally driven period of lower than average recruitment (**Table 6**). Under these conditions, only the SPSSFox and CE CMPs when tuned to 0.7 for PGK_{short} had less than 50% probability of breaching the LRP (**Table 6**).

3.3 Kobe Time Plot

Kobe time plots show the percentage of simulations for each year of the projection period that are in each quadrant of the Kobe plot for each CMP in the Reference and robustness operating models. For example, **Figure 3** shows a Kobe plot for one configuration of the CE CMP. In this case, for the Reference operating models, there is greater than 50% probability of being in the green region of the Kobe matrix in all 30-years of the projection period (F**igure 3**). The results from the robustness operating models (R1, R2, R3a, R3b, and R4) can be compared against the results from Reference models, and the baseline robustness OM (R0). This example shows that the CE_a configuration of the CE CMP has a considerably lower probability of remaining in the green region in several of the robustness tests, especially early in the projection period (**Figure 3**).

3.4 Trade-Off Plot

Trade-off plots are used to compare the results of CMPs with respect to two performance metrics in a scatterplot. **Figure 4** provides an example of four trade-off plots showing the trade-offs between the probability of being in the green space of the Kobe matrix (PGK) in the first 10-years of the projection period against the average TAC over this same period (top left), the PGK in years 11 – 20 against the average TAC over this same period (top right), the probability of not breaching the limit reference point against the average TAC in years 11 – 20 (bottom left), and the mean variation in TAC (shown as a negative value so lower values mean more variable) against the median TAC in the medium timeframe (bottom right). In these plots, higher values (further to the right on x-axis or higher on the y-axis) indicate better performance outcomes. This example shows the results from the 15 configurations of the 5 selected CMPs for the Reference operating models. Results for the robustness operating models are displayed in trade-off plots in the Shiny application.

3.5 Violin Plot

Violin plots show the density distribution of simulation outcomes for TAC change between management cycles for each CMP configuration under the conditions of the reference and robustness operating models (**Figure 5**). The width of the violin plot is proportional to the frequency of the absolute change in TAC (i.e., wider areas means value is more common). These plots indicate how reactive a CMP may be to new data and thus be driving change in TAC between management cycles relative to other CMPs given the same set of conditions. For example, a CMP may require a relatively large shift in CPUE data before it changes TAC, whereas another CMP may more closely follow the CPUE trend when generating new TAC advice.

4 Discussion

The candidate management procedures developed for the north Atlantic swordfish MSE use different data sources and different sets of rules to convert these data to a total allowable catch recommendation. Consequently, the performance of the candidate management procedures varies considerably across the different performance metrics, and across the different conditions of the reference and robustness operating models. A considerable challenge in the MSE process is the interpretation of the large amount of output from the analysis, and the identification of a candidate management procedure that is robust to uncertainty and most likely to achieve the management objectives under the range of plausible conditions in the future.

Managers can specify minimum performance criteria for some performance metrics, which allows CMPs that fail these requirements to be identified and removed from the list of options. For example, the managers of the swordfish fishery specified that management procedures must have at least an 85% probability of not breaching the limit reference point, and at least a 51% probability of being in the green space of the Kobe matrix. These criteria were used in the development of the CMPs, and CMPs that fail these minimum requirements are not presented as options to the managers.

It is rare that a MSE process identifies a single CMP that clearly outperforms all other options. The ranking and selection of best performing CMPs can vary across different stakeholders and decision-makers depending on their specific values and objectives for the fishery. More likely, as is the case for swordfish, the CMPs present trade-offs among competing management objectives, such as a desire for high probability of not over-fishing the stock and a desire to maximize the economic output of the fishery. The results presented in the Shiny application allow different groups of decision-makers to evaluate the performance of the CMPs under the conditions of the reference operating models, compare how well these CMPs perform under the more challenging conditions of the robustness tests, and identify the CMP that they consider to be the best candidate for managing the fishery.

5 Acknowledgements

We thank the ICCAT, the SCRS, and the Swordfish Species Group for their support on this work. This work was carried out under the provision of the ICCAT Science Envelope and the ICCAT – EU Grant Agreement – Strengthening the scientific basis for decision-making in ICCAT. The contents of this paper do not necessarily reflect the point of view of ICCAT or other funders and in no way anticipates ICCAT future policy in this area.

References

Anon. (2022). Report of the 2022 ICCAT Atlantic Swordfish Stock Assessment Meeting SCRS/2022/012. Collect. Vol. Sci. Pap. ICCAT, 79(2): 392-564.

Hordyk, A. (2021). Evaluation of the Performance of some Candidate Management Procedures to Prioritize the Key Uncertainties in the North Atlantic Swordfish Operating Models. Collect. Vol. Sci. Pap. ICCAT, SCRS/2021/161.

Methot, R. D., & Wetzel, C. R. (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

Table 1. Summary of the estimated stock dynamics for the nine operating models (OMs) in the Reference Set. The nine OMs spanned uncertainty in the assumed natural mortality rate (*M*) and the steepness of the Beverton-Holt stock-recruit relationship (*h*). The estimated unfished equilibrium spawning biomass (SB0; ton), and the estimated fishing mortality rate (F) and the spawning biomass (SB) relative to their respective values at maximum sustainable yield (MSY) in the terminal year of the assessment (2020) are reported in the table.

OM #	*M*	*h*	*SB0*	F/F_{MSY}	SB/SB_{MSY}
1	0.1	0.69	416,034	0.86	1.14
2	0.1	0.80	359,174	0.86	1.1
3	0.1	0.88	324,022	0.85	1.1
4	0.2	0.69	153,915	0.9	1
5	0.2	0.80	133,478	0.84	1.04
6	0.2	0.88	120,962	0.77	1.14
7	0.3	0.69	85,800	0.63	1.36
8	0.3	0.80	75,196	0.54	1.64
9	0.3	0.88	69,484	0.46	2.06

Table 2. Description of the Robustness operating models (OMs) developed for the North Atlantic swordfish MSE.

Robustness OM	*Purpose*
R0	Reference OM for the Robustness tests. OM 5 from the Reference Set (Table 1)
R1	Evaluate impact of an assumed 1 percent annual increase catchability, that is not accounted for in the standardization of the indices of abundance (historical & projection)
R2	Same as R2, but only for the historical period
R3a	Evaluate impact of cyclical pattern in recruitment deviations in projection period; a proxy for impact of climate change on stock productivity
R3b	Evaluate impact of lower than expected recruitment deviations for first 15 years of projection period; a proxy for impact of climate change on stock productivity
R4	Evaluate impact of illegal, unreported, or unregulated catches

Table 3. Summary of the Management Objectives and corresponding Performance Metrics (PMs) developed for the North Atlantic swordfish MSE.

Category	*Management Objective*	*PM Name*	*PM Description*
Status	The stock should have a [51, 60, 70]% or greater probability of occurring in the green quadrant of the Kobe matrix.	PGK_{short}	Probability of being in Green Zone of Kobe Space ($SB>SB_{MSY}$ & $F<F_{MSY}$) in years 1-10 (2024-2033)
		PGK_{med}	Probability of being in Green Zone of Kobe Space ($SB>SB_{MSY}$ & $F<F_{MSY}$) in years 11-20 (2034-2043)
		PGK_{long}	Probability of being in Green Zone of Kobe Space ($SB>SB_{MSY}$ & $F<F_{MSY}$) in years 21-30 (2044-2053)
		PGK	Probability of being in Green Zone of Kobe Space ($SB>SB_{MSY}$ & $F<F_{MSY}$) over all years (2024-2053)
		PNOF	Probability of Not Overfishing ($F<F_{MSY}$) over all years (2024-2053)
Safety	There should be a [5, 10, 15]% or less probability of the stock falling below B_{LIM} ($0.4*B_{MSY}$) at any point during the 30-year evaluation period.	LRP	Probability of breaching the limit reference point (SB<0.4SBMSY) in any year (2024-2053)
Yield	Maximize overall catch levels.	TAC1	TAC (t) in the first implementation year (2024)
		$AvTAC_{short}$	Median TAC (t) over years 1-10 (2024-2033)
		$AvTAC_{med}$	Median TAC (t) over years 11-20 (2034-2043)
		$AvTAC_{long}$	Median TAC (t) over years 21-30 (2044-2053)
Stability	Any increase or decrease in TAC between management periods should be less than [25]%. [also test no stability limitation]	VarC	Mean variation in TAC (%) between management cycles over all years and simulations

Table 4. Summary of the candidate management procedures that were developed and tested for the North Atlantic swordfish MSE.

Name	*Type*	*Abundance Indicator*	*Description*
AT1	Empirical	CDN, JPN, CHT, MOR, POR, USA, SPN	The indices are smoothed and averaged together using inverse variance weighting. A ratio of the average of the most recent 3 years of the index and the average of the period from 2015 to 2020 dedicates the percentage change in the TAC. TACs are limited to a 25% change.
CE	Empirical	Combined index	Attempts to maintain a constant exploitation rate in the projection period, based on the mean exploitation rate in the recent historical years.
CI1	Empirical	Combined index	The index is smoothed and a ratio of the average of the most recent 3 years of the index and the average of the period from 2015 to 2020 dedicates the percentage change in the TAC. TACs are limited to a 25% change.
EA1	Empirical	MOR, POR, SPN	The indices are smoothed and averaged together using inverse variance weighting. A ratio of the average of the most recent 3 years of the index and the average of the period from 2015 to 2020 dedicates the percentage change in the TAC. TACs are limited to a 25% change.
GSC2	Empirical	Combined index	Focuses on trying to provide a stable TAC and only deviates when the 3-yr average of the combined index increases or decreases more/less than 10% from a historical period. Initial movement away from the base TAC is 1kt, after this larger changes occur.
MCC2	Empirical	Combined index	Mostly Constant Catch 2 (MCC) focuses on trying to provide stable TAC and only deviates when the 3-yr average of the Combined Index increases or decreases by large amount compared to a 3-yr historical average (2018-2020).
MCC3	Empirical	Combined index	Mostly Constant Catch 3 (MCC) focuses on trying to provide stable TAC and only deviates when the 3-yr average of the Combined Index increases or decreases by large amount compared to a 3-yr historical average (2017-2019).
MCC4	Empirical	Combined index	Mostly Constant Catch 4 (MCC) focuses on trying to provide stable TAC and only deviates when the 3-yr average of the Combined Index increases or decreases by large amount compared to a 3-yr historical average (2017-2019). MCC4 differs from MCC3 by implementing smoother for the Combine Index.
MCC5	Empirical	Combined index	Mostly Constant Catch 5 (MCC) focuses on trying to provide stable TAC and only deviates when the 3-yr average of the Combined Index increases or decreases by large amount compared to a 3-yr historical average (2017-2019). MCC5 differs from MCC3 by implementing a set TAC of 5kt when the average Combine Index hits a lower limit.
SPSS	Model	Combined index	Schaefer surplus production model with a harvest control rule that throttles F when estimated biomass is below target level.
SPSSFox	Model	Combined index	A Fox surplus production model with a harvest control rule that throttles F when estimated biomass is below target level.
WA1	Empirical	CDN, USA, JPN, CHT	The indices are smoothed and averaged together using inverse variance weighting. A ratio of the average of the most recent 3 years of the index and the average of the period from 2015 to 2020 dedicates the percentage change in the TAC. TACs are limited to a 25% change.
FX2	Empirical	CDN, JPN, CHT, MOR, POR, USA, SPN	The 20th, 40th, 60th and 80th percentiles of each index are compared to the average of the most recent 3 years of data in order to find the appropriate percentile interval and associated percent TAC change. The average percent TAC change across the 7 indices adjusts a base TAC which varies according to the PGK_short tuning objective.
FX4	Empirical	Combined index	The combined index is subjected to a median smoother of length 3 and then the deciles of the smoothed index are compared with the average of the most recent 3 years of data in order to find the appropriate percentile interval and associated percent TAC change. The percent TAC change adjusts a base TAC which varies according to the PGK_short tuning objective.
C1320	Empirical	NA	A constant harvest scenario where the TAC is fixed at a level that achieves the PGK_short 0.51, 0.60 and 0.70 objectives.

Table 5. The five candidate management procedures that were identified as the best candidates that spanned the trade-off space for the North Atlantic swordfish MSE.

CMP	*Tuned CMP*	*Tuning Target*	*Tuning Parameter*
CE	CE_a	0.51	0.841
	CE_b	0.60	0.828
	CE_c	0.70	0.764
FX4	FX4_a	0.51	1.029
	FX4_b	0.60	0.986
	FX4_c	0.70	0.943
MCC5	MCC5_a	0.51	0.929
	MCC5_b	0.60	0.888
	MCC5_c	0.70	0.850
MCC7	MCC7_a	0.51	0.875
	MCC7_b	0.60	0.834
	MCC7_c	0.70	0.794
SPSSFox	SPSSFox_a	0.51	1.144
	SPSSFox_b	0.60	1.058
	SPSSFox_c	0.70	0.996

Table 6. The probability of breaching the Limit Reference Point (LRP; $0.4SB_{MSY}$) for the tuned versions of the five selected candidate management procedures for the Reference Set and the Robustness Test OMs.

CMP	*Probability of Breaching LRP*						
	Reference Set	*R0*	*R1*	*R2*	*R3a*	*R3b*	*R4*
CE_a	0.04	0.14	0.24	0.08	0.36	0.72	0.26
CE_b	0.03	0.08	0.2	0.02	0.22	0.6	0.14
CE_c	0.02	0.04	0.16	0	0.14	0.48	0.12
FX4_a	0.01	0.02	0.34	0.12	0.52	0.96	0.32
FX4_b	0.01	0.02	0.16	0.08	0.4	0.86	0.1
FX4_c	0	0	0.12	0.04	0.22	0.8	0.04
MCC5_a	0.03	0.1	0.42	0.22	0.58	0.92	0.32
MCC5_b	0.01	0.04	0.30	0.16	0.46	0.86	0.18
MCC5_c	0	0	0.18	0.1	0.22	0.78	0.04
MCC7_a	0.01	0.04	0.36	0.14	0.5	0.88	0.28
MCC7_b	0	0.02	0.24	0.06	0.32	0.8	0.16
MCC7_c	0	0	0.14	0.04	0.2	0.66	0.02
SPSSFox_a	0.03	0.1	0.4	0.04	0.28	0.66	0.2
SPSSFox_b	0.01	0.06	0.18	0	0.18	0.54	0.1
SPSSFox_c	0	0	0.1	0	0.08	0.44	0.1

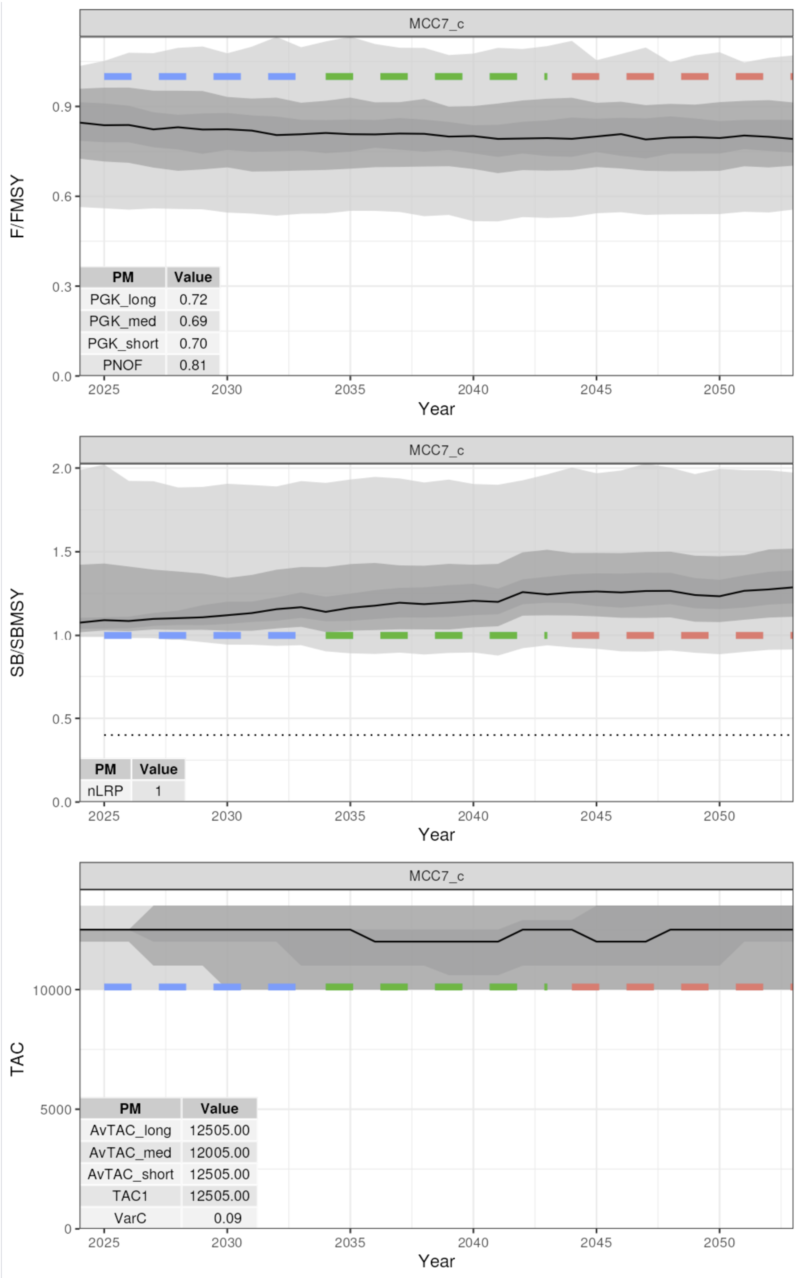

Figure 1. A set of time-series plot for one configuration of the MCC7 CMP, showing the median (black line), 60th, 70th, and 90th percentiles (increasingly lighter shades of grey respectively) for F/F_{MSY} (top), SB/SB_{MSY} (center), and the total allowable catch (TAC; bottom) over the 30-year projection period. This plot shows results for the nine reference operating models. Other plots are available for the robustness models in the Shiny application. The performance metrics associated with this configuration of the MCC7 CMP are shown in tables in the bottom left of each plot.

	MP	AvTAC_long	AvTAC_med	AvTAC_short	nLRP	PGK	PGK_med	PGK_short	PNOF	TAC1	VarC
	All	All	All	All	All	All	All	All	All	All	All
1	CE_a	11655.14	11387.05	13446.71	0.96	0.53	0.51	0.51	0.68	13462.5	0.16
2	CE_b	11651.06	11292.16	12768.65	0.97	0.61	0.59	0.6	0.74	12858.27	0.15
3	CE_c	11555.8	11218.02	12158	0.98	0.69	0.68	0.7	0.79	12247.38	0.15
4	FX4_a	12228.51	12872.12	13515.72	0.99	0.49	0.47	0.51	0.61	13515.72	0.1
5	FX4_b	12324.66	12632.78	12940.89	0.99	0.6	0.57	0.6	0.71	12940.89	0.1
6	FX4_c	12084.33	12379.07	12379.07	1	0.71	0.7	0.7	0.82	12379.07	0.1
7	MCC5_a	11711.75	11711.75	14054.1	0.97	0.48	0.47	0.51	0.57	14054.1	0.06
8	MCC5_b	11188.4	11188.4	13426.08	0.99	0.58	0.56	0.6	0.68	13426.08	0.06
9	MCC5_c	12854.07	12854.07	12854.07	1	0.7	0.68	0.7	0.8	12854.07	0.06
10	MCC7_a	11026.91	11026.91	13783.64	0.99	0.49	0.48	0.51	0.61	13783.64	0.09
11	MCC7_b	11564.15	11564.15	13141.08	1	0.59	0.57	0.6	0.71	13141.08	0.09
12	MCC7_c	12505.21	12005	12505.21	1	0.7	0.69	0.7	0.81	12505.21	0.09
13	SPSSFox_a	11792.19	11819.34	13462.5	0.97	0.53	0.51	0.51	0.67	13462.5	0.17
14	SPSSFox_b	11680.82	11603.5	12753.58	0.99	0.63	0.62	0.6	0.75	13292.91	0.16
15	SPSSFox_c	11571.51	11473.42	12189.85	1	0.72	0.7	0.7	0.82	12521.77	0.15

Figure 2. An example of a quilt plots that are available in the Shiny application that presents the results of the north Atlantic swordfish MSE. This table shows 15 CMP configurations (rows) and 10 performance metrics (columns). The selection of the CMPs and performance metrics can be customized in the Shiny application. The cells are shaded indicating the range of values, with darker colors indicating more desirable outcomes for the various performance metrics.

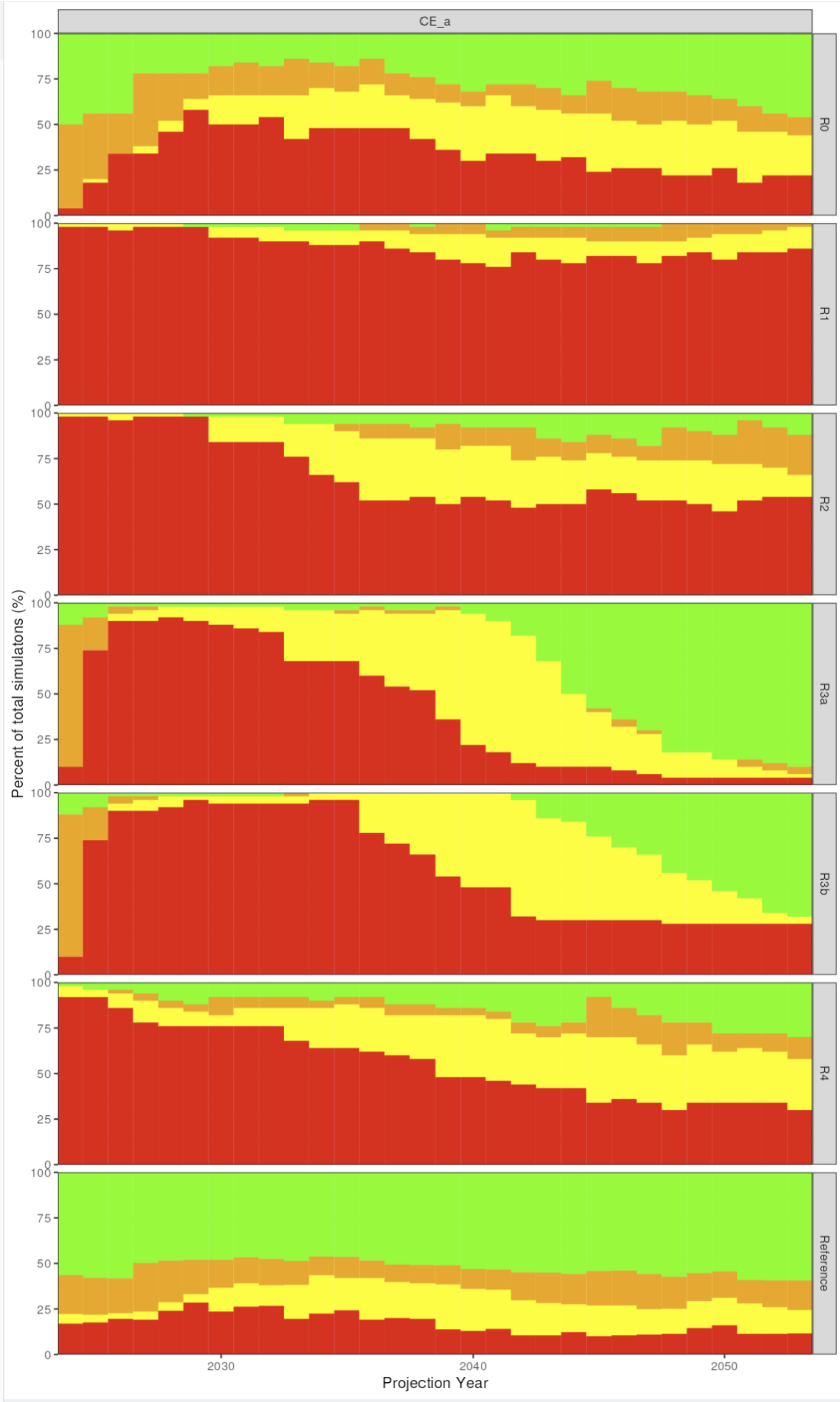

Figure 3. An example of a Kobe time plot for one configuration of the CE CMP, showing the proportion of the simulations in each quadrant of the Kobe matrix in each year of the projection period. The plot on the bottom shows the results for the Reference operating models, and the remaining plots show the results for the baseline (R0) and five robustness models.

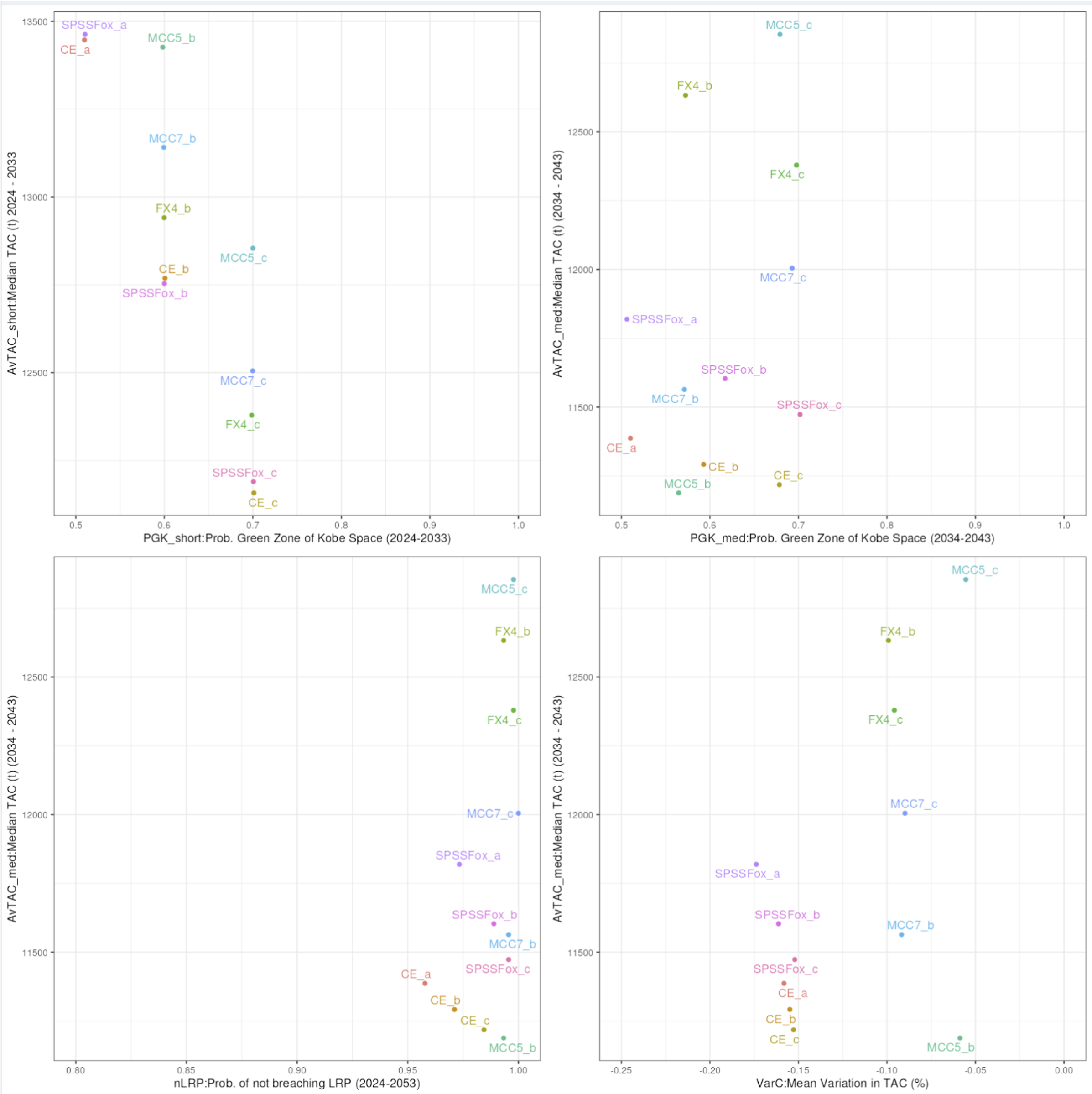

Figure 4. An example of a set of trade-off plots showing the results from 15 configurations of 5 CMPs for the Reference operating models. The plots show the trade-offs between the probability of being in the green space of the Kobe matrix (PGK) in the first 10-years of the projection period against the average TAC over this same period (top left), the PGK in years 11 – 20 against the average TAC over this same period (top right), the probability of not breaching the limit reference point against the average TAC in years 11 – 20 (bottom left), and the mean variation in TAC (shown as a negative value so lower values mean more variable) against the median TAC in the medium timeframe (bottom right).

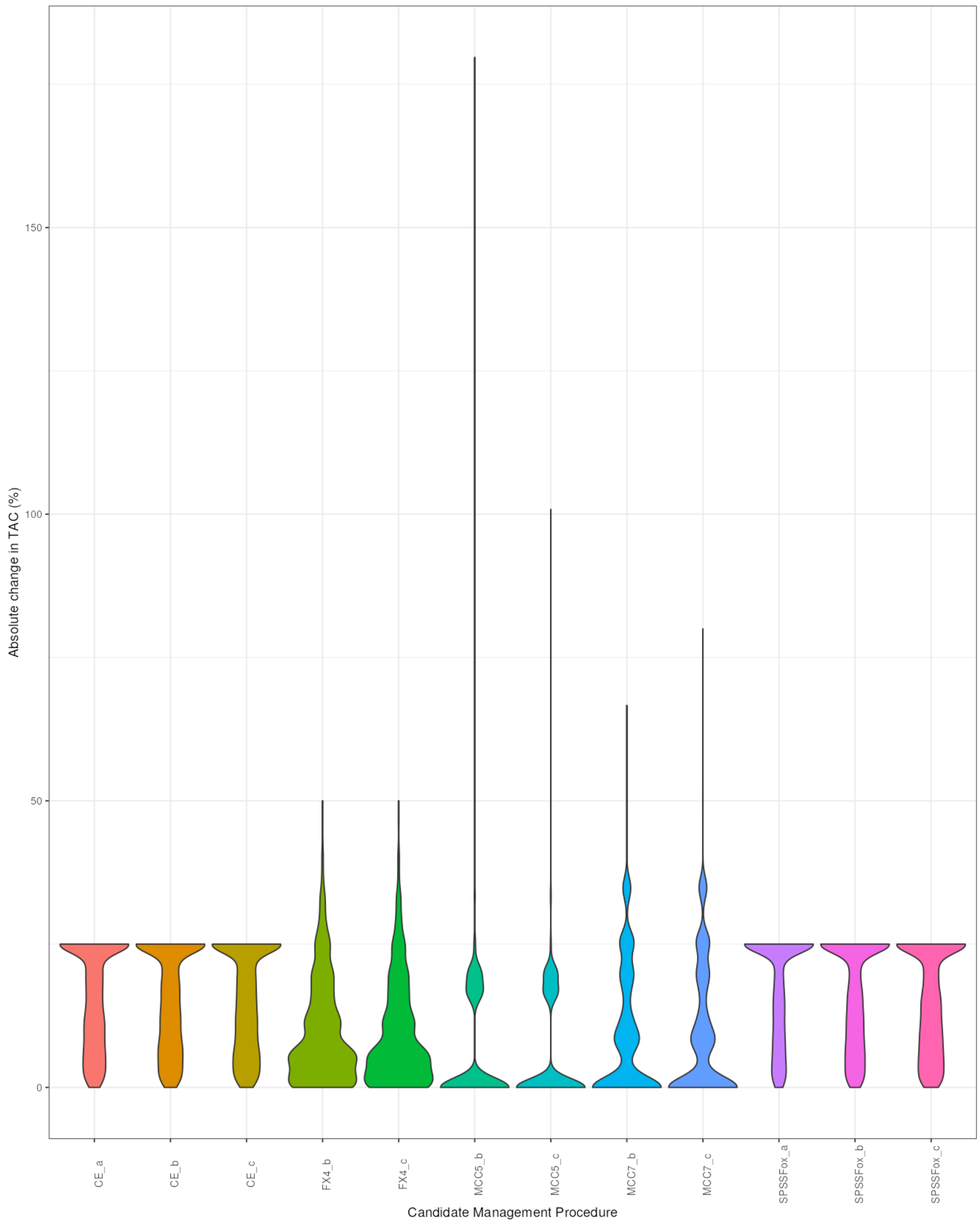

Figure 5. An example of a violin plot showing the distribution of the absolute change in TAC (y-axis) for five CMP configurations (x-axis). The width of the violin plot is proportional to the frequency of the absolute change in TAC (i.e., wider areas mean the value is more common).

Appendix A

Specifications for MSE Trials for North Atlantic Swordfish

The Trial Specifications document for the N-SWO MSE process is available online (https://iccat.github.io/nswo-mse/TS/Trial_Specs.html).

Appendix B

Details of Selected Candidate Management Procedures

1. AT1, CI1, EA1, and WA1

These 4 CMPs have the same underlying structure and decision making process and only differ with respect to the choice of input data (see **Table 4**).

The steps for calculating the dynamic scalar (*deltaTAC*) and new TAC are as follows:

1. Fit an exponential smoothing state space model to each index (R package forecast)
2. Calculate the average of each smoothed series for the years 2015 to 2020 (Target)
3. Calculate the most recent 3 year average from the smoothed series (Current)
4. Calculate the dynamic scalar where *deltaTAC = exp(log(Current/Target)*.1)*
5. TACy+1=TACy *deltaTAC

2. GSC2

The TAC in each management cycle is calculated as:

$$\mathrm{TAC}_y = \mathrm{TAC}_{\mathrm{base}}\Delta_{\mathrm{TAC}} + \mathrm{TAC}_{\mathrm{add}}$$

where $\mathrm{TAC}_{\mathrm{base}} = \theta 13{,}200$, where θ is the tuning parameter that results in achieving the desired short-term PGK (currently tested at 51%, 60%, and 70%), and Δ_{TAC} and $\mathrm{TAC}_{\mathrm{add}}$ are calculated as:

$$\Delta_{\mathrm{TAC}} = \begin{cases} 1.2 & \text{if } I_{\mathrm{rat}} \geq 1.2 \\ 1 & \text{if } 0.75 \leq I_{\mathrm{rat}} < 1.2 \\ 0.625 & \text{if } 0.5 \leq I_{\mathrm{rat}} < 0.75 \\ 0.5 & \text{if } I_{\mathrm{rat}} < 0.5 \end{cases}$$

$$\mathrm{TAC}_{\mathrm{add}} = \begin{cases} 1000 & \text{if } 1.1 \leq I_{\mathrm{rat}} < 1.2 \\ -1000 & \text{if } 0.75 \leq I_{\mathrm{rat}} < 0.9 \\ 0 & \text{otherwise} \end{cases}$$

3. MCC CMPs

The goal of the MCC (Mostly Constant Catch) CMPs is to have the catch remain as constant as possible and only increase if the Combined Index increased substantially and only decrease if the Combined Index declined substantially. The base TAC (constant catch) would be 12,600, this is an approximation of the constant catch that would result in PGK60 and also achieve LRP <15%.

A base TAC ($\mathrm{TAC}_{\mathrm{base}}$) is calculated as:

$$\mathrm{TAC}_{\mathrm{base}} = \theta 12{,}600$$

where θ is the tuning parameter that results in achieving the desired short-term PGK (currently tested at 51%, 60%, and 70%).

$\mathrm{TAC}_{\mathrm{base}}$ is modified by comparing the ratio of the current 3-year average of the Combined Index (I_{curr}) to a historical 3-year average of the Combined Index (I_{base}):

$$I_{\mathrm{rat}} = \frac{I_{\mathrm{curr}}}{I_{\mathrm{base}}}$$

The value of I_{rat} was then used to determine how much $\mathrm{TAC}_{\mathrm{base}}$ should be increased or decreased if at all.

The total allowable catch (TAC) for the following management cycle was then calculated as:

$$TAC_{y+1} = \mathrm{TAC}_{\mathrm{base}}\Delta_{\mathrm{TAC}}$$

where Δ_{TAC} is determined by a set of CMP-specific rules described below.

MCC2

I_{base} is calculated as the average of the Combined Index from 2018-2020, and Δ_{TAC} calculated as:

$$\Delta_{\text{TAC}} = \begin{cases} 1.2 & \text{if } I_{\text{rat}} \geq 1.2 \\ 1 & \text{if } 0.75 \leq I_{\text{rat}} < 1.2 \\ 0.75 & \text{if } 0.5 \leq I_{\text{rat}} < 0.75 \\ 0.5 & \text{if } I_{\text{rat}} < 0.5 \end{cases}$$

MCC3

I_{base} is calculated as the average of the Combined Index from 2017-2019, and Δ_{TAC} calculated using same rules as MCC2.

MCC4

The same historical period is used for I_{base} as in MCC3, but the combined index is given a smoother to help reduce the wide fluctuations between years in the index values. A smoothed index is generated by applying Tukey's Running Median Smoother (stats::smooth R function). The same rules as MCC2 are used to calculate Δ_{TAC}.

MCC5

The same historical period and rules for determining Δ_{TAC} are used as in MCC3, but TAC is set to 5,000 t when $I_{\text{rat}} < 0.5$.

MCC6

A different historical period is used for I_{base} (2015-2017), this is the lowest 3-yr average since 2001. The TAC when $I_{\text{rat}} < 0.5$ is set at 4,000t.

MCC7

The same historical period is used for I_{base} as in MCC3 (2017-2019), but different rules as used to calculate Δ_{TAC}:

$$\Delta_{\text{TAC}} = \begin{cases} 1.35 & \text{if } I_{\text{rat}} \geq 1.35 \\ 1.25 & \text{if } 1.25 \leq I_{\text{rat}} < 1.35 \\ 1.20 & \text{if } 1.20 \leq I_{\text{rat}} < 1.25 \\ 1.10 & \text{if } 1.15 \leq I_{\text{rat}} < 1.20 \\ 1 & \text{if } 0.75 \leq I_{\text{rat}} < 1.15 \\ 0.75 & \text{if } 0.5 \leq I_{\text{rat}} < 0.75 \\ 0.5 & \text{if } I_{\text{rat}} < 0.5 \end{cases}$$

4. CE

The CE management procedure aims to keep a fixed exploitation rate in the projection years. The Combined Index is used to track to relative changes in the population. A smoothed index is generated by applying Tukey's Running Median Smoother (stats::smooth R function).

The historical relative exploitation rate is calculated as:

$$E_{\text{hist}} = \frac{\bar{C}_{\text{hist}}}{\bar{I}_{\text{hist}}}$$

where $\bar{C}_{\text{hist}}$ and $\bar{I}_{\text{hist}}$ are the mean reported catch and smoothed index respectively over the 5 most recent historical years (2016 – 2020).

The current relative exploitation rate is calculated as:

$$E_{\text{curr}} = \frac{\bar{C}_{\text{curr}}}{\bar{I}_{\text{curr}}}$$

where $\bar{C}_{\text{curr}}$ and $\bar{I}_{\text{curr}}$ are the mean reported catch and smoothed index respectively over the 3 most recent projection years.

The target relative exploitation rate is set to E_{hist} but subject to a harvest control rule based on the ratio of the current to historical smoothed index (I_{ratio}) (calculated over same years as above):

$$E_{\text{targ}} = \begin{cases} E_{\text{hist}} & \text{if } I_{\text{ratio}} \geq 0.8 \\ E_{\text{hist}}(-1.4 + 3I_{\text{ratio}}) & \text{if } 0.8 > I_{\text{ratio}} > 0.5 \\ 0.1E_{\text{hist}} & \text{otherwise} \end{cases}$$

The ratio of the target to current relative exploitation rate is calculated:

$$E_{\text{ratio}} = \frac{E_{\text{targ}}}{E_{\text{curr}}}$$

The total allowable catch (TAC) for the following year is then calculated as:

$$TAC_{y+1} = \theta E_{\text{ratio}} TAC_y$$

where θ is a tuning parameter, subject to a constraint where it cannot change by more than 25% from one management cycle to the next.

5. SPSS and SPSSFox

The SPSS and SPSSFox management procedures use a state-space surplus production model to set the TAC. The two CMPs are identical except that SPSS assumes a Schaefer production curve while SPSSFox assumes a Fox production curve.

The Combined Index is used to track to relative changes in the population. A smoothed index is generated by applying Tukey's Running Median Smoother (stats::smooth R function).

The state-space surplus production model from the SAMtool package (SAMtool::SP_SS) is used to fit to the smoothed index and the reported catch.

The following harvest control rule is used to set the target exploitation rate (E_{targ}):

$$E_{\text{targ}} = \begin{cases} E_{\text{prop}} & \text{if } B_{\text{curr}} \geq B_{\text{thresh}} \\ E_{\text{prop}}\left(-0.367 + 1.167\dfrac{B_{\text{curr}}}{B_{\text{thresh}}}\right) & \text{if } B_{\text{thresh}} > B_{\text{curr}} > B_{\text{lim}} \\ E_{\text{min}} & \text{otherwise} \end{cases}$$

where E_{prop} is the proposed harvest rate, calculated as $\theta 0.15$ where θ is the tuning parameter, B_{curr} is the estimated biomass from the surplus production model, B_{thresh} is the estimated biomass corresponding with maximum sustainable yield, B_{lim} is $0.4B_{\text{thresh}}$, and E_{min} is $0.1E_{\text{prop}}$.

The total allowable catch (TAC) for the following year is then calculated as:

$$TAC_{y+1} = E_{\text{targ}} B_{\text{curr}}$$

The TAC is subject to a constraint where it cannot change by more than 25% from one management cycle to the next.

6. FX2

The TAC in each management cycle is updated according to:

$$TAC_{y+1} = \mathrm{TAC}_{\mathrm{base}}\Delta_{\mathrm{TAC}}$$

where $\mathrm{TAC}_{\mathrm{base}}$ is the base TAC that is determined by the tuning parameter, and the dynamic scalar (Δ_{TAC}) is derived from the average TAC change recommended by the seven principle indices of abundance (CDN, CHT, JPN, MOR, POR, SPN, USA):

$$\Delta_{\mathrm{TAC}} = \mathrm{mean}\left(I_{1,j}, I_{2,j}, I_{3,j}, I_{4,j}, I_{5,j}, I_{6,j}, I_{7,j}\right)$$

where each $I_{i,j}$ is selected from the following vector of percent change PC = {0.8, 0.9, 1, 1.025,1.05} based on the relationship between the current 3 year average index value to the ventiles of the index. For example, if the 3 year average was below or equal to the 1st ventile, the percent change would be 0.8, or if were equal to or above the 4th ventile the percent change would be 1.05.

7. FX4

The FX4 cMP bases TAC change on the most recent 30 years of the combined index. The index is first subjected to a running median smoother of length 3. A base TAC, which scales with the tuning objective as shown in the table below, is further scaled by a percent change (*deltaTAC*) selected from the following vector: PC = {0.75, 0.8, 0.85, 0.9,0.95,1, 1.025, 1.05,1.075,1.1, 1.125 }

The relationship between the current 3 year average index value to the deciles of the index determines which value is selected. For example, if the 3 year average was below or equal to the 1st decile, the percent change would be 0.75, or if it were equal to or above the 4th decile the percent change would be 0.95.

8. C1320

This CMP does not rely on data to provide future TAC advice. It represents the largest constant TAC that is tuned to achieve the $\mathrm{PGK}_{\mathrm{short}}$ tuning objectives.