SCRS/2022/090 Collect. Vol. Sci. Pap. ICCAT, 79(4): 216-230 (2022)

PRELIMINARY CLOSED-LOOP SIMULATIONS FOR NORTHEAST PORBEAGLE: ILLUSTRATING THE EFFICACY OF ALTERNATIVE MANAGEMENT PROCEDURES AND ASSESSMENT FREQUENCY

Nathan G. Taylor[1], Mauricio Ortiz[2], Ai Kimoto[2], and Rui Coelho[3]

SUMMARY

Porbeagle shark populations are listed on CITES Appendix 2 so it may be hard to obtain indices of abundance for stock assessment. We conduct a preliminary series of closed-loop simulations to determine yield, conservation, and variability in effort performance of different Management Procedures (MPs) and assessment frequency. We test 106 MPs with different data requirements. We condition Operating Models, on catch and CPUE time series of the northeast stock using life history and other parameters from the 2020 ICCAT assessment for the western porbeagle stock. There are many MPs that could meet status reference points equivalent to CITES listing criteria as well as B_{MSY} and F_{MSY} criteria; within those, there is a large variability in catch performance. The effect of assessment frequency on performance depends on the metric and on the MP. Future refinements of OMs are needed to provide a more challenging set of hypotheses to test MP performance and to match the properties of the eastern porbeagle stock and fishery closely.

RÉSUMÉ

Les populations de requins-taupes communs étant inscrites à l'annexe 2 de la CITES, il peut être difficile d'obtenir des indices d'abondance pour l'évaluation des stocks. Nous avons réalisé une série préliminaire de simulations en boucle fermée afin de déterminer la production, la conservation et la variabilité des performances de l'effort de différentes procédures de gestion (MP) ainsi que la fréquence d'évaluation. Nous avons testé 106 MP avec différentes exigences en matière de données. Nous avons conditionné les modèles opérationnels sur les séries temporelles de captures et de CPUE du stock du Nord-Est en utilisant le cycle de vie et d'autres paramètres de l'évaluation de 2020 du stock de requin-taupe commun de l'Ouest de l'ICCAT. De nombreuses MP pourraient atteindre des points de référence d'état équivalents aux critères d'inscription à la CITES ainsi que des critères de BPME et FPME ; parmi ceux-ci, il existe une grande variabilité dans les performances de capture. L'effet de la fréquence des évaluations sur les performances dépend de la mesure et de la MP. Des améliorations futures des OM sont nécessaires pour fournir un ensemble d'hypothèses plus stimulantes pour tester la performance des MP et pour correspondre étroitement aux caractéristiques de la pêcherie et du stock de requin-taupe commun de l'Est.

RESUMEN

Las poblaciones de marrajo sardinero se incluyen en el Apéndice 2 de CITES, por lo que podría ser difícil obtener índices de abundancia para la evaluación del stock. Se llevó a cabo una serie preliminar de simulaciones de círculo cerrado con el fin de determinar el rendimiento, la conservación y la variabilidad en el desempeño del esfuerzo de diferentes procedimientos de ordenación (MP) y la frecuencia de las evaluaciones. Se probaron 106 MP con diferentes requisitos de datos. Se condicionaron los modelos operativos a las series temporales de capturas y CPUE del stock del nordeste utilizando el ciclo vital y otros parámetros de la evaluación de ICCAT de 2020 para el stock del marrajo sardinero del oeste. Muchos MP podrían cumplir los puntos de referencia del estado equivalentes a los criterios de inclusión en CITES, así como los criterios de B_{RMS} y F_{RMS}; dentro de ellos, hay una gran variabilidad en el desempeño de las capturas. El efecto de la frecuencia de evaluación sobre el desempeño depende de la medición y del MP. Es necesario perfeccionar en el futuro los OM para proporcionar un conjunto de

[1] nathan.taylor@iccat.int ICCAT Secretariat. Calle Corazón de Maria 8, Madrid Spain 28002.
[2] ICCAT Secretariat. Calle Corazón de Maria 8, Madrid Spain 28002.
[3] IPMA - Portuguese Institute for the Ocean and Atmosphere. Av. 5 de Outubro s/n, 8700-305 Olhão, Portugal

hipótesis más exigente que permita comprobar el desempeño de los MP y que se ajuste a las propiedades de la pesquería y el stock de marrajo sardinero del oeste.

KEYWORDS

Simulation, Stock Assessment, Depleted stocks, Shark fisheries, Tuna fisheries, Population dynamics, By catch, Fishery sciences, Fishing effort

1. Introduction

The CITES listing of Porbeagle shark in Atlantic Ocean on CITES Appendix 2 and no-retention for several fleets, creates uncertainty about which data might be available to assess the stock. For the northwest porbeagle stock, it was not possible to generate a CPUE series to use for stock assessment (ICCAT 2020). For the eastern stock one index of abundance, the Spanish longline index may be available for future stock assessment. While there are many catch, effort, index, and length-based Management Procedures (the set of decisions about data, stock assessment model, and harvest control rule MPs) that do not require indices of abundance (Carruthers and Hordyk 2018), it is not clear what the consequences of using such methods are in terms of the conservation, yield variability in effort for the stock.

An additional control variable affecting management performance is the time interval at which the stock is assessed. Stocks that are assessed more frequently may have more feedback control (Hilborn 1979; Parma 1990; Orensanz et al. 2004), so that the management system may be reacting to changes in the state of the stock cause by process variation in recruitment, growth and mortality (Hilborn and Walters 1992) as detected by the assessment. But any changes in the stock´s status detected by the assessment are determined through assessment methods that may be considerably in error due to estimation errors (Ludwig et al. 1981; Ludwig and Walters 1985; NRC 1998; Magnusson and Hilborn 2007). Accordingly, any evaluation of the effects of assessment frequency needs to be filtered through an analytical technique that captures the effect of these estimation errors.

One potential technique for addressing both the effect of estimation errors and the effect of assessment intervals on management performance, is to use a Management Strategy Evaluation (Smith 1993; De la Mare 1998; Punt et al. 2016; Kaplan et al. 2021) approach (MSE). It allows testing how MPs with different data input classes perform against performance metrics (PM) for conservation, yield, variability in yield and how different assessment intervals affect these statistics. In this way, the effects of data choices and estimation errors can be considered in determining the efficacy of any given management procedure.

Here we use MSE to show that a number of potential MPs could be used to ensure that the stock rebuilds above the $0.20B_0$ CITES criterion for a variety of different risk thresholds but that there is a large range of potential yield depending on the MP chosen.

2. Methods

We followed the seven basic steps outlined by (Punt et al. 2016) for our northeast porbeagle MSE simulations (outlined below).

2.1 Identification of the management objectives in concept and representation of these using quantitative performance statistics

The main management objectives in concept that we consider are those related to yield, stock status, and variability in effort. Performance statistics for yield, stock status, and variability are pretty standard for MSE (Hall et al. 1988; Punt et al. 2016; Forrest et al. 2018) and elsewhere in applied ecology (Mendelssohn 1980).

For yield, we measure MP performance relative against a so-called reference yield. The reference yield is calculated by projecting the population forward in the OM with a fixed fishing mortality and optimizing for the fishing mortality that results in the highest long-term yield. It is usually close to MSY but unlike MSY, it is not an equilibrium value, but rather, it accounts for the impact of recruitment deviations in the simulations and other non-equilibrium dynamics (see the OpenMSE webpage for more details). Values of the reference yield performance statistic are reported in relative terms i.e., in fractions of the reference yield so that values near one correspond to nearly achieving the maximum yield possible.

With respect to objectives, management objectives for the northeast porbeagle stock (NE POR) have not been defined formally in any specific management recommendations, but we can illustrate putative objectives that are based on the current CITES listing criteria and general objectives defined in the ICCAT convention. Rather than debate the definition of a Limit Reference Point for porbeagle shark, we use the CITES listing criterion for listing (see Appendix 5 Rev. CoP17, Appendix 5) that defines general guideline for a marked recent rate of decline is the rate of decline that would drive a population down within approximately a 10-year period from the current population level to the historical extent of decline guideline (i.e. 5-20 % of baseline for exploited fish species). By this definition we adopt the 20% of the unfished spawning stock biomass $0.2B_0$ as a minimum stock size objective P_{20}. The newly amended ICCAT Convention text defines one main objective for management as to “ensure maintenance of the populations of ICCAT species in the Convention area at or above levels capable of producing maximum sustainable yield”. Accordingly, we define the second stock status objective is that the stock should be above the spawning stock biomass that produces maximum sustainable yield, B_{MSY}.

CITES and the ICCAT Convention text are less clear about what probability limits apply to these objectives. Given the latter, it is not possible to define precise acceptable probability limits P_{lim} (or the probability that the stock is above level $0.2B_0$). Nor is it possible to define a criterion characterizing the acceptable probability that the stock is above B_{MSY} (P_{MSY}). But for both P_{lim} and P_{MSY} it is a reasonable assumption that it is desirable that it be more likely than not that MPs result the biomass being above the stock status objectives i.e., that P_{lim} and P_{MSY} should be greater than 50%. By this criterion it may be possible to eliminate MPs failing to meet this criterion from consideration. So our basic satisficing (Miller and Shelton 2010) criterion P_{crit} for MPs was:

$P_{crit}=P(B_t>0.2B_{MSY})>50\%$ & $P(B_t>B_{MSY})>50\%$ & $P(F_t/F_{MSY}) >50\%$

There is no basis for choosing a performance statistic for variance. For illustrative purposes we use the Average Annual Variability in Effort (AAVE), rather than variability in catch in order to capture how effort (including effort not targeting but catching NE porbeagle) might vary with the application of given MPs. In this way, the statistic captures how both fishing effort directed at the target species and also effort directed at non-target species might have to vary with the application of a given MP.

For variability in catch and in effort, lower variability of catch or effort are typically viewed as being better. To ease with reader’s comprehension when looking at graphics and tables, we adapt the performance metrics so that higher values reflect preferred outcome. To achieve this end, we transform some performance metrics so that positive is better. For a fishing mortality statistic (where a high probability of overfishing is viewed unfavorably), we use the probability of not overfishing (PNOF) and for AAVE, we express this quantity as the probability that AAVE is greater than 20%. In this way, a reader examining tables and graphics to represent MP performance sees quantities consistently represented high performance as being associated with high values.

The actual selection of MPs in practice will involve a much richer set of objectives and tradeoffs that we did not have the capacity to consider here. This might involve additional performance metrics, different time periods, and different probabilities associated with achieving the desired (or avoiding the undesired) outcomes.

2.2 Identification of a broad range of uncertainties to which the management strategy should be robust

For practical reasons, we limited our explorations of uncertainties to the input steepness, growth, selectivity and other parameters and index choices for OM conditioning (described below).

2.3 development of a set of operating models which provide a mathematical representation of the system to be managed, including biological characteristics of the stock, the fisheries which intercept the modelled stock, and how data are collected from the system

We define OMs using the R package OpenMSE. OMs are defined initially using a set of input parameters (Carruthers and Hordyk 2018); these parameters are described in POR-NE-MSE.html (supplemental materials). For steepness and growth parameters, we rely on Cortes et al. 2020´s simulated distributions of growth and steepness (Cortes 2020) as custom parameters in the OMs for steepness and the von Bertalanffy growth parameters; rather than the uniform distribution specified in POR-NE-MSE.html, where for the simulations a sample of 100 simulations is drawn from the Cortez et al 2020 distribution for steepness and the von Bertalanffy growth parameters. The mean and variance of steepness from these simulations was used a prior for the Rapid Conditioning Model (RCM), described below. The initial OMs were updated using the RCM function in the OpenMSE R package.

2.4) selection of starting values for parameters in the operating models and quantifying parameter uncertainty

As a preliminary analysis of MP performance, limits on time reduced the range of uncertainty that we could consider practically. We borrowed operating model parameters extensively from ICCAT 2020 (see OM reports). The initially defined OMs (see above) are modified using a function called Rapid Conditioning Model RCM in the OpenMSE package. RCM fits an age-structured stochastic stock reduction model (Walters et al. 2006) to the data (catch, indices of abundance, and length composition data). Then, RCM updates the original population dynamics parameters to reflect how the Stochastic Stock Reduction Analysis SRA (Walters et al. 2006) updated the prior parameters defined in POR-NE-MSE.html (see **Figure 1**). Parameters updated by RCM include:

Unfished recruitment OM@R0, only if catch is provided.

Depletion OM@D

Recruitment autocorrelation OM@AC which is estimated post-hoc from the recruitment deviation estimates.

Annual recruitment deviations OM@cpars$Perr_y. Historical recruitment are those estimated from the model, while future recruitment will be sampled with autocorrelation.

Because there is not much length composition data, selectivity parameters OM@L5, OM@LFS, and OM@Vmaxlen are not updated but rather sampled for a range of values specified as input. The projection period selectivity is assumed to be equal to that in the last historical year of the conditioning period, in this case the input selectivity values of the OM. While there were some length composition data available for NE porbeagle, the annual sample sizes were very small, and the RCM would not converge if it was included in the likelihood function.

For a preliminary analysis we built two OMs (the reference set) that are based on choices about different time series for fitting. Following Ortiz et al. 2022, we build two scenarios: S1 which was fitted to three indices reviewed by the ICES WKELASMO in 2022; and S2 that also included fits to a fourth historical index presented in 2009 stock assessment.

2.5 identification of candidate MPs which could realistically be implemented for the system

We classify MP according to different data input categories as follows:

- Catch-based (model-free MPs that only require catch data)
- Length-based (model-free MPs that require only length data)
- Index-based (model-free MPs that require only an index of abundance)
- Depletion-based (model, or model-free MPs that require an estimate of depletion)
- Abundance-based (typically model-based MPs that estimate the current abundance).

After removing any MPs that would not converge, there were 106 MPs available as part of the DLMtool and the SAMtools package in R (see Table 2). Given the uncertainty of the available data for eastern porbeagle stock assessment, the aim of testing this range of MPs was to determine what the consequences in terms of yield, conservation, and variance in effort of different MP choices to inform potential data-collection schemes into the future.

2.6 simulation of the application of each MP for each operating model

For the reference set, each MP was applied at the specified assessment interval (starting with 5 years for the reference set). 100 iterations of each MP were used for testing. TACs determined by a given MP at a specific interval were applied for all years between assessment intervals. Parameters describing the implementation of given MPs are described in POR-NE-MSE.html (supplemental materials). The overall schematic for the simulation framework is presented in **Figure 2**.

2.6.1 Effects of the interval of assessment on key performance metrics

Given our criterion for minimum MP performance, we did this testing using only the initial set of MPs that passed P_{crit}. This was largely for practical reasons in that it reduced the computation time to run the simulations for all MPs across different intervals. We explored the range of effects of increasing or decreasing the frequency at which MPs were applied. We tested MP intervals of 1, 4, 7, 10, 13, and 20 years.

2.7 summary and interpretation of the performance statistics

We limit our treatment of performance measures (PM) to four quantities capturing stock status, fishing exploitation rate, and variability in effort. As discussed above, we present performance statistic for the probability of being above the CITES threshold (P_{20}), the probability for being at B_{MSY} (P_{100}), the probability of not overfishing (PNOF) and the AAVE criterion. Performance measures are divided into short term (1-20 years), medium term (30-40 years), and long term (50-60 years). For defining average performance across the two OMs, the weighted average of performance statistic was taken assuming equal weighting between S1 and S2.

To provide a standard against which to compare MP performance we test a set of so-called reference MPs. These are MPs that receive perfect information from the operating model. They are used to provide an illustration of an MP without the confounding effects of process and observation error that the rest of the MPs suffer from in the simulations. Reference MPs are useful for provided the basis for comparing the performance of MPs simulated with errors in data input, process errors, and estimation errors that make difficult time series analysis (Clark and Bjornstad 2004).

While it not currently possible to define what manager´s tolerance for being above a certain objective, we could explore the number of MPs that would be admissible, i.e., meeting the criterion given different risk tolerance for being at or above these objectives. Accordingly, we loop across a range of tolerance options from 50% to 95% and select MPs that meet that criterion. We select MPs that pass P_{20}, P_{100} and PNOF for each tolerance level: for example, if tolerance for P_{20} is 90% we select all MPs for which $P_{20} \geq 90\%$; if the tolerance level is 80% for P_{20}, we select all MPs where $P_{20} \geq 80\%$, and so on. We apply the same procedure for the P_{100} and the PNOF criteria and summarize the number of admissible MPs by data input class and data input type.

For illustrative purposes, we extract the performance of surplus production and delay difference (Deriso 1980; Schnute 1985) models from the larger set of MPs explored above for the assessment interval simulations. These MPs most-closely resemble those that have been used for the stock assessment for the northeast porbeagle and therefore may have some bearing on the management consequences of the assessment interval for this stock. The MPs in this subset that use delay difference models DD, DD4010, state space delay difference models DDSS_4010, DDSS_75MSY, as well as surplus production models using a 40:10 harvest control rule SP_4010, a surplus production model that fishes at 75% of F_{MSY}, SP_75MSY, and a surplus production model that fishes at F_{MSY}, SP_MSY. SP_4010 probably represents the MP that is close to the ICES MP.

3. Results

3.1 Operating Models

Detailed description of S1 and S2 OMs can be found in the S1EPOM.html and S2EPOM.html OM reports, respectively. The S1 and S2 OMs are very similar (see compare_RCM.html). This is perhaps not all that surprising given that in fitting the models they differed only by the use of one index. There were differences in estimates of biomass, and fishing mortality and recruitment, but these differences were very minor (see compare_RCM.html RCM output tab). Similarly, differences in F_{MSY}, MSY, and the spawning depletion at MSY reference points differed only at the third significant figure (compare_RCM.html reference point tab): both OMs estimate that the mean SSB/SSB_{MSY} in eastern POR is approximately 0.5. The main differences between the two operating models were in the recruitment deviations see (compare_RCM.html/RCM output/Model predictions) but beyond these, the differences between the two were slight.

3.2 Broad Patterns of MP performance across data input class

Reference MPs, that are applied with perfect information show how the stock might be expected to build with when MPs are applied. Table 1 shows that fishing at F_{MSY} (the FMSYref procedure) will return the highest yield but the resulting P_{100} statistic is less than 50% for all time periods. As fishing mortality is reduced from F_{MSY} to values of 50% of F_{MSY} (FMSYref50) there is an approximately 10% reduction in catch relative to the refence yield but P_{20} and P_{100} values of essentially 100% (Table 1). Naturally, closing the fishery (the Nref MP) results in a high probability of the stock being above B_{MSY} and the $0.20B_0$ with nearly 100% probability across all time frames.

MPs varied considerably in their mean performance across operating models (see **Figure 3**). On average, the pattern of MP performance was similar for P_{20} and P_{100} statistics. However, MPs tend to have higher probabilities of being above P_{20} than of being above P_{100} which is not surprising given that the P_{20} performance metric is a lower biomass and that more rebuilding is needed to attain B_{MSY}. Model-free MPs (catch, length, and index based) tended to have good median P_{20} and P_{100} performance (see **Figure 3**, P20 and P100 mean columns) whereas MPs that have an estimate of depletion and abundance had higher median yield performance (**Figure 3**, yield.mean column). This pattern held across time frames. But while the model-based management procedures had higher yield performance than the model-free procedures, this came at the expense of both poorer conservation performance as well as higher variability showing the classic yield, conservation, and variance tradeoff (**Figure 3**). This is a pattern that matches other closed-loop simulations (Walters 1975; Hall 1981; Hall et al. 1988; Taylor et al. 2014; Hicks et al. 2016; Forrest et al. 2018).

The effect of different risk thresholds P_{LIM} , P_{MSY} and PNOF on the number of MPs that meet satisficing criteria

For all criteria, as the tolerance for being below a given objective decrease, the number of admissible MPs increases across all data classes (**Figure 4)**. MPs that required estimates of depletion or abundance were the most sensitive change in these tolerances with for example the number of admissible MPs against the CITES criterion declining from 16 to 7 as the tolerance for being below the CITES criterion goes from 50% to 95%. Model-free management procedures were slightly less sensitive to changes in these tolerances; they tended to be more conservative to start with, so that increasing the tolerance for being below a given objective made very little difference to if they were admissible.

Satisficed MPs

25 MPs passed the satisficing criteria P_{crit} (**Table 2**). Only one catch-based MP (SPMSY) met the satisficing criteria P_{crit}. When ordered in descending order on yield, P_{20} and P_{100}, there is not much to differentiate the top 5 MPs. The P_{crit} criterion removed many poor performing MPs with respect to conservation criteria so that the smallest probabilities of being below any of the CITES criterion, B_{MSY}, and F_{MSY} were 0.67, 0.55, 0.57 respectively (**Table 2**). But there was huge variation in the yield relative to the reference yield than ranged from 6% to 100%. The MPs with the poorest yield performance tended to be the model-free MPs with means across the satisficed MPs of 0.46, 0.12, 0.34 for the catch-based, index based, and length-based MPs, respectively. The simulations indicate that any combination of SP_75MSY, SP_4010, DD, SP_MSY, DD4010 MPs have nearly 100% probabilities being above P_{20} and at least an 87% change of being above B_{MSY}. These same MPs also scored very well in terms of yield, achieving nearly 100% of the reference yield. Common to the top 5 MPs is that they apply relatively simple stock assessment models that could continue to be applied with the existing CPUE series for the stock. However, if model-free MPs were to be required in the future then the consequence of using them in this case is lower overall catch performance.

Effect of management interval on MP performance

How MPs changed with increasing assessment frequency depended on the performance statistics being considered and the data input class of the MP (**Figure 5**). Even within a given MP data input class, the patterns were not always consistent. For abundance-based MPs, across most data input classes (columns), there was no consistent pattern showing how the intervals between the application of MPs affect performance: sometimes performance (i.e., higher values) increased with decreased time between the application of MPs; however sometimes it declined.

With the exception of a few MPs, in general AAVE (the probability that the average annual variability in effort is less than 20%) increased as the assessment interval increased (**Figure 5.**) This was the case across all data input classes. The results in consistent with expectations: if an MP determines a TAC and that TAC stays in place until the next assessment then, then it follows that variability in the catch decreases with decreased assessment frequency simply because the TAC is not being updated very frequently.

With respect PMs that measured stock size or status (P100, P40, and P10) and PNOF, performance varied by data input and by performance statistics. MPs that had estimate or measurement of abundance tended to have flat or decreasing trends as assessment interval increased. One exception to this pattern was the SPslope MP: in this case, performance increased as the assessment interval increased. For the catch-based SPMSY (Martell and Froese 2013) performance was flat. Depletion-based MPs tended to have flat (DCAC4010) or declining performance (for SPSRA and MCD4010) as the assessment interval increased. PMs for the model-free MPs (index based and length based), tended to have an asymptotic shape with increased assessment intervals. In these cases, MP performance improved as the assessment interval increased but after 7-10 years, there were on improvements in performance at all. This observation is important because it also means that for some MPs, performance declined with assessment interval increases from 1 to 7 years. In such cases the optimal choice might be to decrease the frequency of the assessment.

Like the other PMs, yield performance varied by data input class (**Figure 5**). For most of the abundance-based MPs, MP performance varied very little as the assessment interval increased except for state-space delay difference fishing at 75% of F_{MSY} (DDSS_75MSY) where yield performance declined as the assessment interval increased. Similarly, the yield performance of most of the depletion-based MPs was either flat or declined slightly as the assessment interval increased (**Figure 5**). The Yield performance metric had the most interesting pattern of performance vs. the assessment interval for the index-based MPs. In this instance, there were clear signs of a parabolic pattern suggesting an optimal assessment interval for the yield statistic for the index-based MPs. With the exception of the DCAC_ML whose performance appeared to have an optimal assessment interval at between 4 and 13 years, yield performance of length-based MPs increased as the assessment interval increased.

The performance of delay difference models vs assessment intervals different between the state space and their non-state space equivalents. Across all statistics the performance of the state-space variants was worse and at times, considerable worse (**Figure 6**). With the exception of AAVE, their performance deteriorated as the assessment interval increased. One important example is for the P20 statistic where DDSS_75MSY and DDSS_4010 where with an annual assessment interval the probabilities of being above 20%B_0 are 92% and 88% respectively but these probabilities decline monotonically as the assessment interval increases to be 84% and 70% if the assessment interval increases to 20 years.

In general MPs that used surplus production models (SP_4010, SP_75MSY, and SP_4010) had good performance. AAVE was typically above 90%, P100 was above 75%, P20 nearly 100%, and yield values very near to the reference yield. The good performance of these MPs is good news considering their prominent use for stock assessment for this stock.

4. Discussion

This paper describes the broad patterns of how variability in effort, yield, and conservation performance varies with MPs with different data requirements. In general, there were many MPs of varying data input classes that could meet a range of stock status objectives including CITES, B_{MSY} and F_{MSY} objectives depending on the degree of tolerance for failing to meet such objectives. These MPs include model-free and model-based rules. While the model-free MPs that we tested meet conservation and stock status objectives for the stock, the tradeoff in using them was typically loss of catch relative to the reference yield.

Naturally MPs need to be tested against a broader spectrum of OMs. Currently, the OMs are very close to the results of the assessment. This limits the scope of inference about the robustness of the MPs tested considerably: the absence of testing against more challenging OMs matters in that it has not been demonstrated that applying any of this MPs could maintain or restore the stock if the state of the stock where much different that is indicated from the current assessment. Several other elements of the OMs require additional refinement such as: refined life-history parameters closely resembling the eastern stock, as well as the fleet and the implementation elements of the OMs. The apparent good performance of some MPs reflects the fact that they were tested against OMs that very closely resembled the stock status that would have been determined using those MPs (MPs using surplus production models for example). Without testing against a greater number of MPs, it is not possible to know how robust these MPs are against a more challenging set of OMs.

This analysis emphasizes the value of an index of abundance for stock assessment. Other methods of estimating abundance or depletion could be entertained to estimate the stock´s abundance, like close-kin mark recapture (Bravington et al. 2016b, 2016a) or genetic mark-recapture for real-time harvest rate monitoring (GENETAG). But absent such tools, the simplest solution is to maintain one of the existing indices of abundance. In particular if one of the indices could be extended into the future as the stock rebuilds, then having the contrast in the time series would allow for more robust estimates of the stock´s productivity (Ludwig and Walters 1985). Mostly importantly however, MPs that use an index of abundance show good catch performance, good performance at avoiding CITES listing criteria, and high probabilities of being above B_{MSY} for the stock.

While a large number of MPs were tested in this analysis, they were not adjusted specifically to meet a set of MP criteria in the same way that custom tuning MPs have been designed for some MSE processes (Kurota et al. 2010). It is therefore entirely possible that better performing MPs that be designed using only catch data (Fischer et al. 2020) or other empirical MPs that still have good performance.

An important element to consider in evaluating the effect on the assessment interval on MP performance is the signal to noise ratio in the data. In the case of a stock like the northeastern porbeagle, the recruitment dynamics is such that increases in biomass are likely to occur slowly: this means that data that capture true increases in biomass (the signal) should also be expected to vary slowly over time. However, data in the simulations (and in reality) have observation error and so it is possible that short term (annual, biennial) apparent changes in the data are mainly observation error. This is the reason that except for the yield statistic, the performance statistics for the model-free MPs (index and length-based) improve as assessment intervals increase.

There was no general pattern showing that MP performance improved as assessment frequency increased. Rather, for some data classes MP performance improved as assessment frequency increased and for other others it decreased. While the magnitude of the effect of the assessment interval on MP performance varied in some cases it was potentially important: depending on what the satisficing criteria are, MPs could go from being admissible to being inadmissible by either changing the assessment interval. For those MPs whose performance varies with the assessment interval, then this quantity becomes a tuning parameter to be considered as part of broader MP performance evaluation. But given that the direction and magnitude of this performance is not predictable, the consequence is the effect of the assessment interval is something that will need to be examined on a case-by-case basis.

References

Bravington, M. V., Grewe, P.M., and Davies, C.R. 2016a. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. Nat. Commun. **7**: 1–8. Nature Publishing Group. doi:10.1038/ncomms13162.

Bravington, M. V., Skaug, H.J., and Anderson, E.C. 2016b. Close-kin mark-recapture. Stat. Sci. **31**(2): 259–274. doi:10.1214/16-STS552.

Carruthers, T.R., and Hordyk, A.R. 2018. The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. Methods Ecol. Evol. **9**(12): 2388–2395. doi:10.1111/2041-210X.13081.

Clark, J.S., and Bjornstad, C.N. 2004. Population time series: Process variability, observation errors, missing values, lags, and hidden states. Ecology **85**: 3140–3150.

Cortes, E. 2020. Preliminary Estimates of Vital Rates and Population Dynamics Parameters of Porbeagle Shark in the Northwest Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT **76**(10): 164–172.

Deriso, R.B. 1980. Harvesting strategies and parameter estimation for an age-structured model. Can. J. Fish. Aquat. Sci. **37**: 268–282.

Fischer, S.H., de Oliveira, J.A.A., and Kell, L.T. 2020. Linking the performance of a data-limited empirical catch rule to life-history traits. ICES J. Mar. Sci. **77**(5): 1914–1926. doi:10.1093/icesjms/fsaa054.

Forrest, R.E., Holt, K.R., and Kronlund, A.R. 2018. Performance of alternative harvest control rules for two Pacific groundfish stocks with uncertain natural mortality: Bias, robustness and trade-offs. Fish. Res. **206**(November 2017): 259–286. doi:10.1016/j.fishres.2018.04.007.

Hall, D.L. 1981. Alternative Harvest Strategies for Pacific Herring. University of British Columbia.

Hall, D.L., Hilborn, R., Stocker, M., and Walters, C.J. 1988. Alternative Harvest Strategies for Pacific Herring (*Clupea harengus pallasi*). Can. J. Fish. Aquat. Sci. **45**(5): 888–897. doi:10.1139/f88-107.

Hicks, A.C., Cox, S.P., Taylor, N., Taylor, I.G., Grandin, C., and Ianelli, J.N. 2016. Conservation and yield performance of harvest control rules for the transboundary Pacific hake fishery in US and Canadian waters.

In Management Science in Fisheries: An Introduction to Simulation-Based Methods. doi:10.4324/9781315751443.

Hilborn, R. 1979. - Comparison of fisheries control systems that utilize catch and effort data. J. Fish. Res. Bd Can. **36**(12): 1477–1489.

Hilborn, R., and Walters, C.J. 1992. Quantitative Fisheries Stock Assessment Choice, Dynamics and Uncertainty. Routledge, Chapman and Hall.

ICCAT. 2020. Report of the 2020 porbeagle shark stock assessment meeting. Madrid, Spain. Available from https://www.researchgate.net/publication/344218791_REPORT_OF_THE_2020_PORBEAGLE_SHARK_STOCK_ASSESSMENT_MEETING.

Kaplan, I.C., Gaichas, S.K., Stawitz, C.C., Lynch, P.D., Marshall, K.N., Deroba, J.J., Masi, M., Brodziak, J.K.T., Aydin, K.Y., Holsman, K., Townsend, H., Tommasi, D., Smith, J.A., Koenigstein, S., Weijerman, M., and Link, J. 2021. Management Strategy Evaluation: Allowing the Light on the Hill to Illuminate More Than One Species. Front. Mar. Sci. **8**(June): 1–22. doi:10.3389/fmars.2021.624355.

Kurota, H., Hiramatsu, K., Takahashi, N., Shono, H., Itoh, T., and Tsuji, S. 2010. Developing a management procedure robust to uncertainty for southern bluefin tuna: a somewhat frustrating struggle to bridge the gap between ideals and reality. Popul. Ecol. **52**(3): 359–372. doi:10.1007/s10144-010-0201-1.

De la Mare, W.K. 1998. Tidier fisheries management requires a new MOP (management-oriented paradigm). Rev. Fish Biol. Fish. **8**(3): 349–356. doi:10.1023/A:1008819416162.

Ludwig, D., and Walters, C.I.J. 1985. Are Age-structured Models Appropriate for catch and Effort Data? Can. J. Fish. Aquat. Sci. **42**(6): 1066–1072.

Ludwig, D., Walters, C.J., and Pulliam, H.R. 1981. Measurement errors and uncertainty in parameter estimates for stock and recruitment. Can. J. Fish. Aquat. Sci. **38**: 711–720.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish Fish. **8**(4): 337–358. doi:10.1111/j.1467-2979.2007.00258.x.

Martell, S., and Froese, R. 2013. A simple method for estimating MSY from catch and resilience. Fish Fish. **14**(4): 504–514. doi:10.1111/j.1467-2979.2012.00485.x.

Mendelssohn, R. 1980. A systematic approach to determining mean-variance tradeoffs when managing randomly varying populations. Math. Biosci. **50**(1–2): 75–84. Elsevier. doi:10.1016/0025-5564(80)90122-4.

Miller, D.C.M.M., and Shelton, P.A. 2010. “Satisficing” and trade-offs: evaluating rebuilding strategies for Greenland halibut off the east coast of Canada. ICES J. Mar. Sci. J. du Cons. **67**(9): 1896–1902. doi:10.1093/icesjms/fsq083.

NRC. 1998. Improving Fish Stock Assessment. National Academies Press, Washington D.C. doi:10.17226/5951.

Orensanz, J. (Lobo), Hand, C.M., Parma, A.M., Valero, J., and Hilborn, R. 2004. Precaution in the harvest of Methuselah ’ s clams — the difficulty of getting timely feedback from slow- paced dynamics 1. Can. J. Fish. Aquat. Sci. **61**: 1355–1372. doi:10.1139/f04-136.

Parma, A.M. 1990. Optimal Harvesting Of Fish Populations With Non-Stationary Stock-Recruitment Relationships. Nat. Resour. Model. **4**(1): 39–76. John Wiley & Sons, Ltd. doi:https://doi.org/10.1111/j.1939-7445.1990.tb00131.x.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., and Haddon, M. 2016. Management strategy evaluation: Best practices. Fish Fish. **17**(2): 303–334. doi:10.1111/faf.12104.

Schnute, J.T. 1985. A general theory for the analysis of catch and effort data. Can. J. Fish. Aquat. Sci **42**(3): 414–429.

Smith, A. 1993. Management Strategy Evaluation - the Light on the Hill. *In* Population Dynamics for Fisheries Management, Australian Society for Fish Biology Workshop Proceedings, Australian. Perth, WA, Australia. pp. 249–253. doi:10.2307/j.ctvrnfpkk.10.

Taylor, N.G., Hicks, A.C., Taylor, I.G., Grandin, C., and Cox, S. 2014. Status of the Pacific Hake (whiting) stock in U . S . and Canadian waters in 2014 with a management strategy evaluation. Int. Jt. Tech. Comm. Pacific Hake: 156.

Walters, C.J. 1975. Optimal Harvest Strategies for Salmon in Relation to Environmental Variabilitv and Uncertain Production Parameters. Can. J. Fish. Aquat. Sci. (32): 1777–1784.

Walters, C.J., Martell, S.J.D., and Korman, J. 2006. A stochastic approach to stock reduction analysis. Can. J. Fish. Aquat. Sci. **63**: 212–223.

Table 1. Summary of Reference MP Performance. F_{MSY}ref fishing the stock at exactly F_{MSY}, F_{MSY}50 fishes the stock at 50%F_{MSY}, F_{MSY}ref75 fishing the stock at 75% of F_{MSY}, and NFref closes the fishery for the entire simulation period.

	Long			Medium			Short		
	P_{20}	P_{100}	**Yield**	P_{20}	P_{100}	**Yield**	P_{20}	P_{100}	**Yield**
F_{MSY}**ref**	0.963	0.430	0.994	0.976	0.420	0.994	0.981	0.456	0.994
F_{MSY}**ref50**	1.000	1.000	0.919	1.000	1.000	0.919	0.998	0.861	0.919
F_{MSY}**ref75**	1.000	0.961	0.987	1.000	0.969	0.987	0.991	0.711	0.987
NFref	1.000	1.000	0.000	1.000	1.000	0.000	0.999	0.933	0.000

Table 2. Summary of MPs meeting the P_{crit} satisficing criteria. MP is the Management Procedure (hyperlinked for details), MPInputClass is the category of data required for the MP, P_{20} is the probability of being above 20% SSB0, P_{100} is the probability of being at SSB_{MSY}, Yield is the mean relative to the reference yield, PNOF is the probability of not overfishing, and AAVE is the average annual variability in effort. All values at the equally weighted means across the operating models. Columns are in descending order on yield, P_{20}, and P_{100}.

MP	**MPInputClass**	P_{20}	P_{100}	**Yield**	**PNOF**	**AAVE**
SP_75MSY	Abundance-based	1.00	0.99	1.00	1.00	0.70
SP_4010	Abundance-based	1.00	0.91	1.00	0.94	0.70
DD	Abundance-based	1.00	0.94	1.00	0.95	0.70
SP_MSY	Abundance-based	0.99	0.87	1.00	0.91	0.70
DD4010	Abundance-based	1.00	0.96	0.99	0.95	0.70
DCAC_ML	Length based	0.87	0.75	0.86	0.78	0.70
SSS_75MSY	Abundance-based	0.95	0.83	0.86	0.90	0.70
DCACs	Depletion-based	0.76	0.60	0.85	0.64	0.69
DCAC_40	Depletion-based	0.75	0.56	0.83	0.57	0.69
DCAC4010	Depletion-based	0.79	0.62	0.79	0.61	0.50
DDSS_4010	Abundance-based	0.85	0.61	0.79	0.70	0.55
SPSRA	Depletion-based	0.75	0.57	0.77	0.67	0.52
DDSS_75MSY	Abundance-based	0.86	0.72	0.77	0.78	0.65
SSS_4010	Abundance-based	0.92	0.87	0.57	0.90	0.68
MCD4010	Depletion-based	0.67	0.55	0.49	0.74	0.39
SPMSY	Catch based	0.72	0.68	0.46	0.69	0.59
ICI2	Index based	1.00	1.00	0.22	0.98	0.53
YPR_ML	Length based	0.68	0.67	0.20	0.58	0.61
BK_ML	Length based	0.75	0.73	0.16	0.69	0.66
SBT1	Index based	1.00	1.00	0.13	1.00	0.61
Fratio_ML	Length based	0.78	0.77	0.13	0.73	0.69
Iratio	Index based	1.00	1.00	0.12	1.00	0.59
ICI	Index based	1.00	1.00	0.08	1.00	0.64
SPslope	Abundance-based	0.78	0.77	0.07	0.73	0.81
Gcontrol	Index based	1.00	1.00	0.06	1.00	0.58

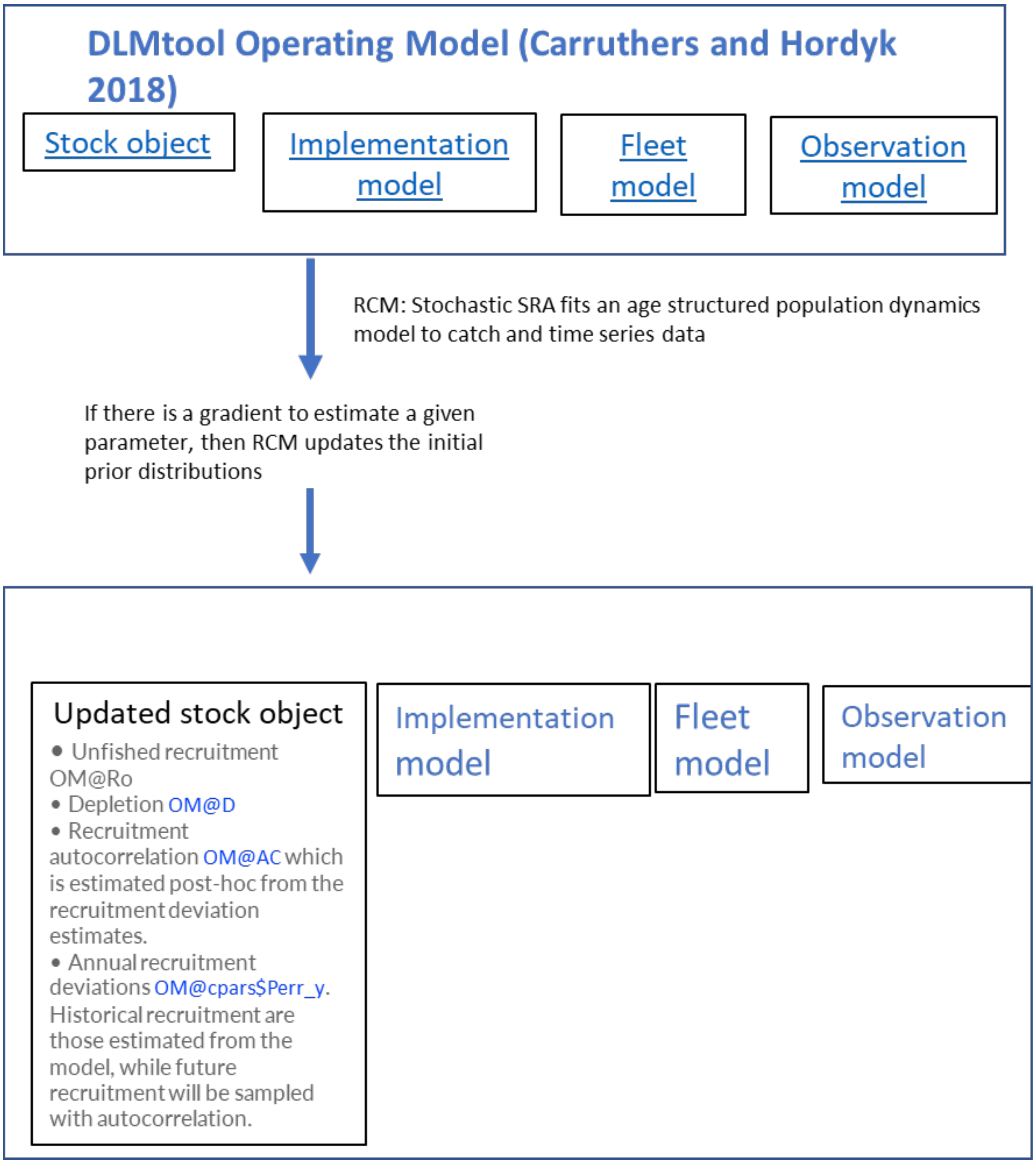

Figure 1. Schematic of the RCM fitting to update Operating Models.

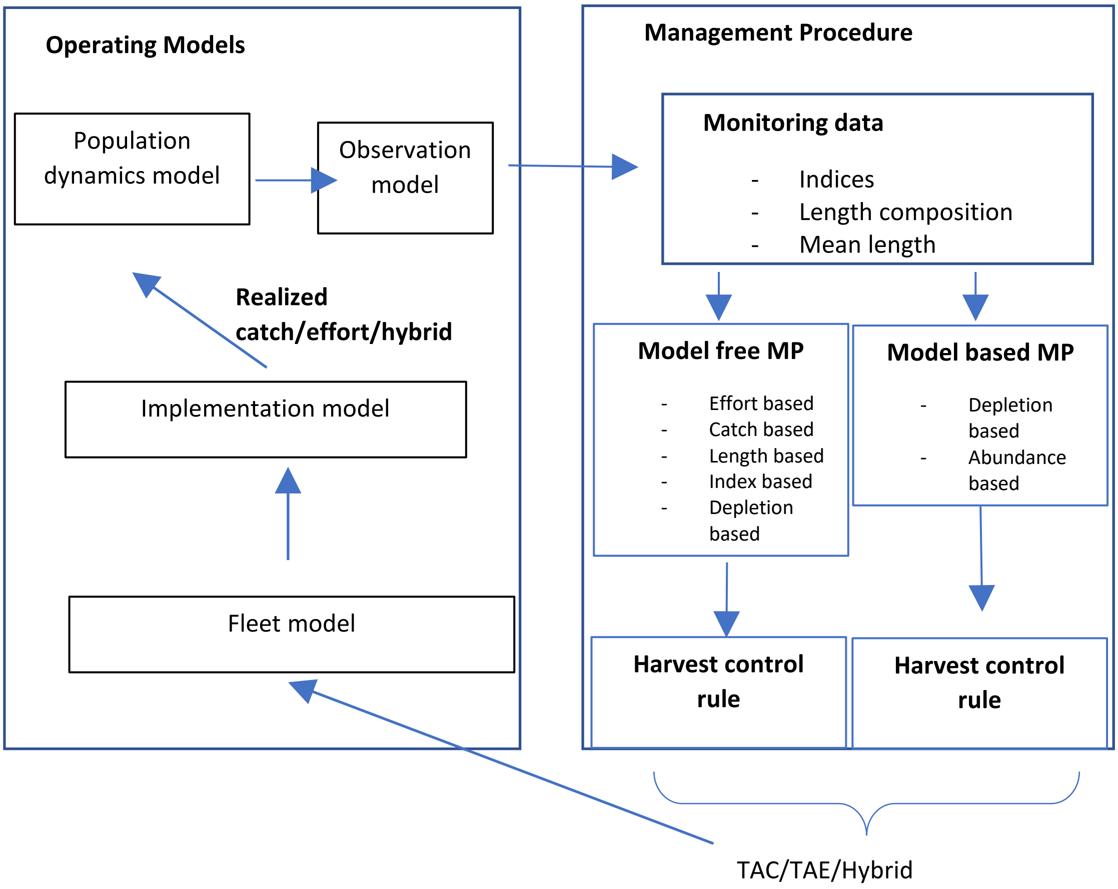

Figure 2. Schematic of closed-loop simulation scheme

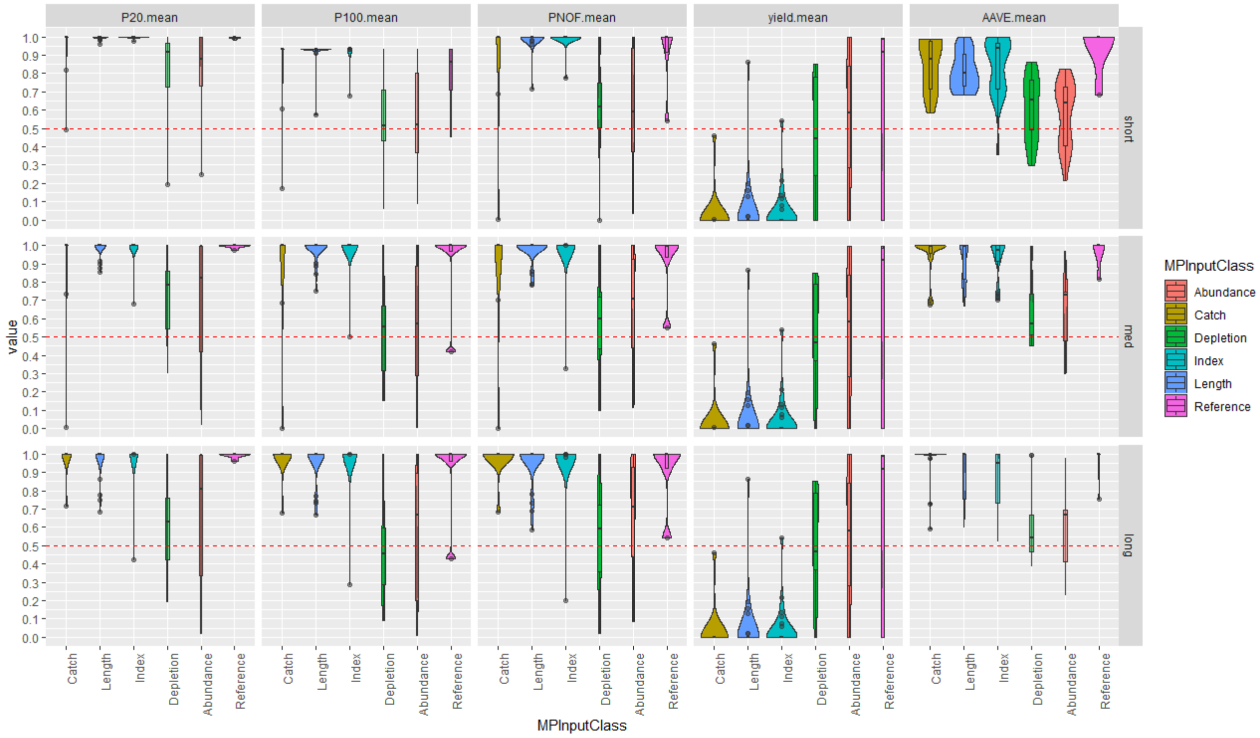

Figure 3. Distributions of MP performance by data input class and time frame.

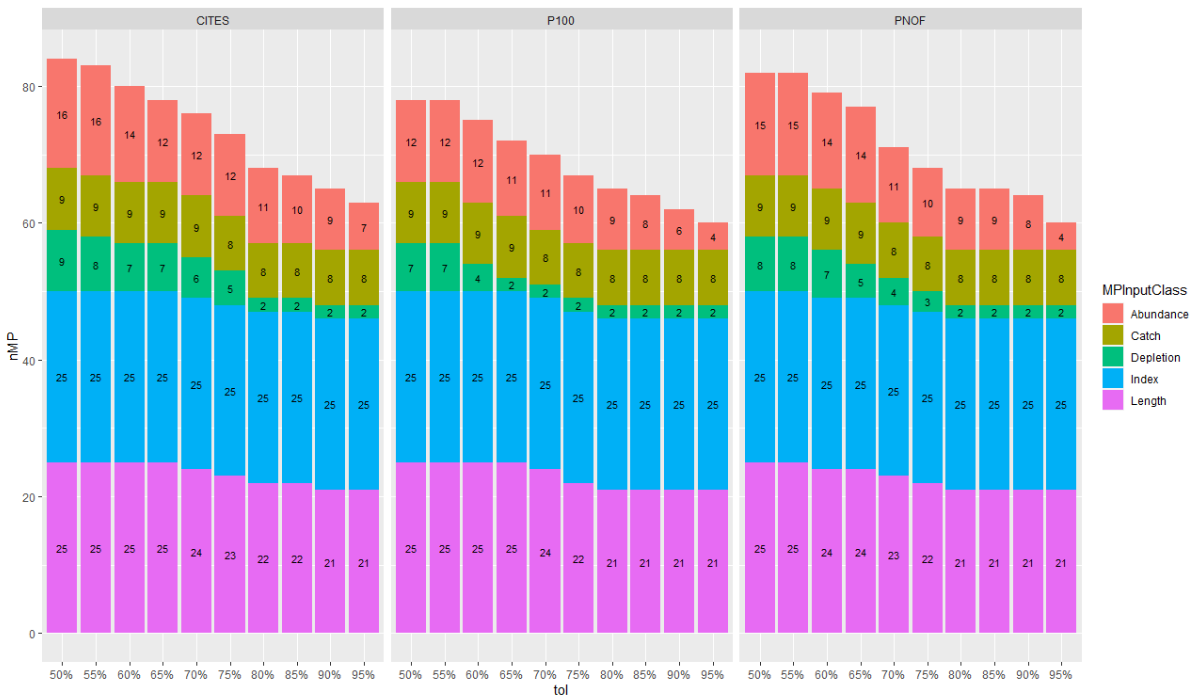

Figure 4. Number of admissible MPs (y), given the tolerance (X) for a given objective (panel header). CITES is the probability of being over 20%SSB0, P_{100} is the probability of being above SSB_{MSY}, and PNOF is the probability of not overfishing.

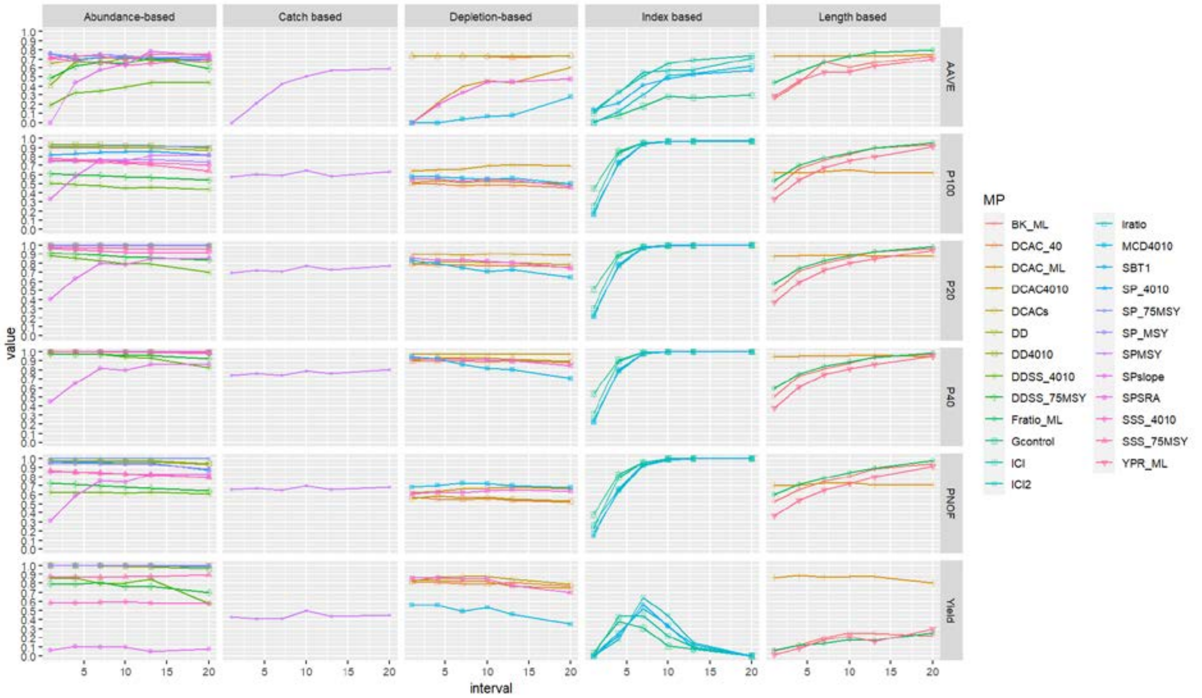

Figure 5. Mean value (Y) across operating models of performance metrics (rows) by data type (columns) for the average annual variability on effort is less than 20% (AAVE), the probability of being at B_{MSY} (P100), the probability of being above the putative CITES appendix 2 threshold (P20), the probability of being above 40% of the unfished biomass (P40), the probability of not overfishing (PNOF), and the yield, relative to reference yield (Yield).

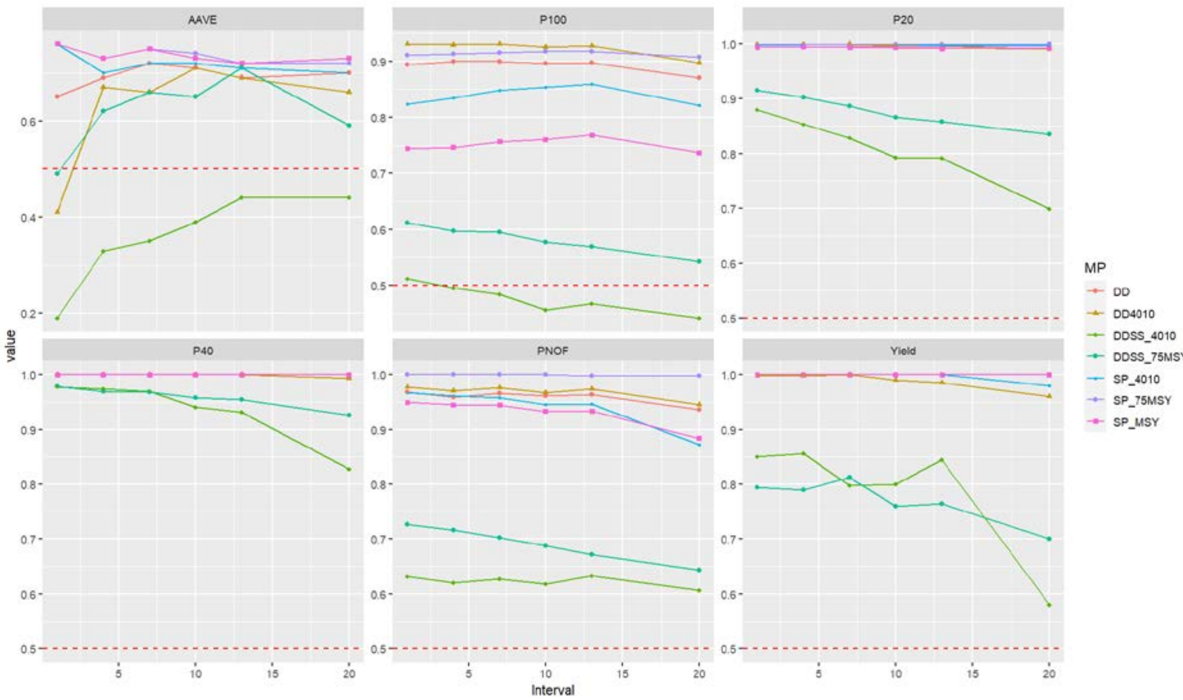

Figure 6. Effect of the assessment interval for MPs using surplus production and delay difference models on the probability for the average annual variability on effort is less than 20% (AAVE), the probability of being at B_{MSY} (P100), the probability of being above the putative CITES appendix 2 threshold (P20), the probability of being above 40% of the unfished biomass (P40), the probability of not overfishing (PNOF), and the yield, relative to reference yield (Yield).