FISH and FISHERIES

The interim management procedure approach for assessed stocks: responsive management advice and lower assessment frequency

Journal:	*Fish and Fisheries*
Manuscript ID	FaF-19-Aug-OA-222
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	29-Aug-2019
Complete List of Authors:	Huynh, Quang; University of British Columbia, Institute for the Oceans and Fisheries Hordyk, Adrian; University of British Columbia, Institute for the Oceans and Fisheries Forrest, Robyn; Fisheries and Ocean Canada, Pacific Biological Station Porch, Clay; Southeast Fisheries Science Center Anderson, Sean; Fisheries and Oceans Canada, Pacific Biological Station Carruthers, Thomas; University of British Columbia, Institute for the Oceans and Fisheries
Key terms:	fisheries management, management strategy evaluation, simulation, stock assessment, total allowable catch
Abstract:	Stock assessments are often used to provide management advice, such as a total allowable catch (TAC), to fishery managers. Many stocks are not assessed annually, and the TAC from the previous assessment is often maintained in the interim years between assessments. We developed two interim management procedures (MPs) that update the estimate of current vulnerable biomass from a surveyed index of abundance to adjust the TAC from a previous assessment. These MPs differ in how they handle uncertainty in observed indices. Using closed-loop simulation, we evaluated the two interim MPs (with 10- and 5-year assessment intervals) against two "status quo" approaches: (1) an annual assessment, and (2) a stock assessment every 5 or 10 years with a fixed TAC between assessments. We evaluated performance across three life-history types and three operating model scenarios (Base, Hyperstable, and Depleted). The interim MPs performed similarly to annual assessments in terms of trends in biomass and yield, regardless of the assessment interval of the interim MPs. The interim MPs produced more yield than the Fixed TAC MP with 10-year assessment intervals, particularly in depleted scenarios. The Fixed TAC MP performed more similarly to interim MPs when the assessment interval for the Fixed TAC MP was decreased to five years. Our results show that interim MPs should be considered for infrequently assessed stocks or rebuilding stocks and highlight potential cost savings of interim MPs over annual assessments.

SCHOLARONE™
Manuscripts

The interim management procedure approach for assessed stocks: responsive management advice and lower assessment frequency

Quang C. Huynh[1], Adrian R. Hordyk[1], Robyn E. Forrest[2], Clay E. Porch[3], Sean C. Anderson[2], Thomas R. Carruthers[1]

[1] Institute for the Oceans and Fisheries, The University of British Columbia, 2202 Main Mall, Vancouver BC V6T 1Z4

[2] Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Rd, Nanaimo BC V9T 6N7

[3] National Marine Fisheries Service, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, FL, 33149

Corresponding author's email: q.huynh@oceans.ubc.ca

Keywords: fisheries management, management strategy evaluation, simulation, stock assessment, total allowable catch

For Review Only

Abstract

Stock assessments are often used to provide management advice, such as a total allowable catch (TAC), to fishery managers. Many stocks are not assessed annually, and the TAC from the previous assessment is often maintained in the interim years between assessments. We developed two interim management procedures (MPs) that update the estimate of current vulnerable biomass from a surveyed index of abundance to adjust the TAC from a previous assessment. These MPs differ in how they handle uncertainty in observed indices. Using closed-loop simulation, we evaluated the two interim MPs (with 10- and 5-year assessment intervals) against two "status quo" approaches: (1) an annual assessment, and (2) a stock assessment every 5 or 10 years with a fixed TAC between assessments. We evaluated performance across three life-history types and three operating model scenarios (Base, Hyperstable, and Depleted). The interim MPs performed similarly to annual assessments in terms of trends in biomass and yield, regardless of the assessment interval of the interim MPs. The interim MPs produced more yield than the Fixed TAC MP with 10-year assessment intervals, particularly in depleted scenarios. The Fixed TAC MP performed more similarly to interim MPs when the assessment interval for the Fixed TAC MP was decreased to five years. Our results show that interim MPs should be considered for infrequently assessed stocks or rebuilding stocks, and highlight potential cost savings of interim MPs over annual assessments.

For Review Only

Table of Contents

Abstract 2
Table of Contents 3
Introduction 5
Methods 9
Management procedures (MPs) 9
Uncertainty in the index 10
Closed-loop simulation study 12
MP evaluation 16
Results 18
How well do interim MPs perform compared to annual assessments? 19
How well do interim MPs perform compared to the Fixed TAC MP? 21
Assessment frequency 22
Discussion 23
Interim MP implementation 24
Strategic management 28
Acknowledgements 29
Conflict of Interest Statement 29
Data Sharing Policy 29
References 30

For Review Only

Tables.....38

Figure Captions.....41

For Review Only

Introduction

For many fished stocks, management recommendations for harvest levels are prescribed in terms of a catch limit, commonly referred to as a total allowable catch (TAC). In output-controlled fisheries where explicit catch limits must be provided, a stock assessment model, such as a statistical catch-at-age model (e.g., Methot & Wetzel, 2012) or a surplus production model (e.g., Prager, 1994; Pedersen & Berg, 2017), is used to provide TAC recommendations to fishery managers. Important estimates from assessments include current vulnerable biomass and reference points, such as the fishing mortality (F) that produces the maximum sustainable yield (F_{MSY}).

If stock assessment outputs were the only tools for setting annual catch limits, then annual assessments would be required to ensure regular, up-to-date management advice. Management bodies and fishing industry organizations often prioritize frequent assessments for high-value and high-volume fisheries. In the U.S. and Canada, annual assessments are conducted for important stocks such as Pacific halibut (*Hippoglossus stenolepis*; Stewart & Hicks, 2018), Pacific hake (*Merluccius productus*; Edwards, Taylor, Grandin, & Berger, 2018), and Pacific herring (*Clupea pallasii*; DFO, 2018). Many stocks under the purview of the International Council for the Exploration of the Sea (ICES), such as Norwegian spring-spawning herring, are also assessed on an annual basis (Zimmermann & Enberg, 2017).

However, the practice of annual assessments tends to be the exception rather than the norm. In many cases, annual assessments are not feasible due to cost limitations, time constraints, data limitations, or insufficient scientific capacity. Instead, assessments are typically conducted periodically, with varying frequency. For example, socio-economically important stocks such as Gulf of Mexico Red Snapper may be assessed every 3–4 years (SEDAR, 2018),

For Review Only

while other species such as Gulf of Mexico Vermilion Snapper (SEDAR, 2016), Gulf of Mexico Cobia (SEDAR, 2013), and many Canadian Pacific stocks managed under the Integrated Fisheries Management Plan for Groundfish (DFO, 2019) are assessed less frequently, with 5–10 year assessment intervals or longer.

Several approaches have been used to provide TAC recommendations in the interim period between assessments. One approach is to project the vulnerable biomass from the most recent assessment to generate a TAC recommendation for each interim year (e.g., Gulf of Mexico Red Snapper, Cass-Calay et al., 2015). However, this approach requires specific assumptions about future levels of recruitment, natural mortality, and fishing behavior that can be difficult to forecast. A common alternative is to fix the TAC at a constant level until the next assessment is adopted. For example, TACs for most assessed Canadian Pacific groundfish stocks are based on the target fishing mortality rate applied to the biomass at the end of the latest assessment and often stay static, with some exceptions, until advice from new assessments is received (e.g., DFO, 2019; Figure 1).

Ideally, TACs should be updated in response to changing fishing pressure and fish population size. For example, if the fish population is decreasing, the catch prescribed by a fixed TAC will correspond to increasingly higher exploitation rates as the population gets smaller (Harford & Carruthers, 2017). On the other hand, foregone yield can occur where a fixed TAC corresponds to a lower exploitation rate for an increasing fish population. Low TACs for an increasing fish population may also lead to higher discard rates due to high encounter rates and lack of available quota.

Assessment intervals have generally not been considered in a strategic context for managing fisheries, even though they can be crucial for successful fisheries management

For Review Only

(Hutniczak, Lipton, Wiedenmann, & Wilberg, 2019; Wiedenmann, Wilberg, Sylvia, & Miller, 2017). Fixed TACs over long intervals amount to static management, which has been found to have poorer performance than adaptive management procedures (Harford & Carruthers, 2017; Sagarese et al., 2018). Static management also has implications for other exploited stocks in multispecies fisheries. For example, low TACs of "choke" species in fully monitored multispecies fisheries, such as British Columbia's groundfish fishery, can limit fishing of other commercially desirable species (DFO, 2012).

Assessment frequency is often constrained by scientific capacity, but data-collection programs may contain valuable information to support decision-making during the interim period between assessments. For example, increases in biomass could be inferred from fishery-independent surveys, or by the appearance of a strong cohort in age-composition data. Such information could provide a basis for updated management advice, but in cases where there are long intervals between assessments, these data generally remain unused until the next scheduled assessment.

Management procedures (MPs) for setting annual TACs based on surveyed indices of abundance have been proposed and applied (Geromont, De Oliveira, Johnston, & Cunningham, 1999; Geromont and Butterworth, 2015; ICES, 2014). Such MPs follow a general formula: after setting an initial TAC, the catch advice in subsequent years either increases or decreases in accordance with the population biomass, as indicated by the index. Control parameters can constrain the rate of change in TAC from one year to the next and scale the overall magnitude of the catch. For example, the magnitude of catch shortly after implementation of such MPs is strongly influenced by the magnitude of the initial TAC. For un-assessed stocks, initial TACs may be set empirically, e.g., using historical catch (Geromont and Butterworth, 2015), or be

For Review Only

based on consensus by managers and stakeholders (Geromont, De Oliveira, Johnston, & Cunningham, 1999). In some cases, such MPs have been found to perform as well as those using conventional assessment models (Carruthers et al., 2016).

In this study, we describe a similar approach, whereby the TAC is set by an assessment model at periodic intervals and by an index-based rule in the interim period between assessments. During an assessment year, we set TACs based on maximum sustainable yield (MSY) targets. In years without an assessment, the interim MPs update the TAC every year based on an updated estimate of current vulnerable biomass inferred from survey indices (calibrated from the most recent assessment), while keeping the target fishing mortality rate of F_{MSY} that was estimated in the most recent assessment. In effect, the assessment model informs some of the control parameters, e.g., the initial TAC, for the index-based rule.

This approach has two potential benefits for assessed stocks. First, for species currently assessed on an annual basis, there is the potential to increase cost efficiency if advice can be robustly updated with less frequent stock assessments. Second, for species with infrequent stock assessments that are currently managed with constant or near-constant TACs, a robust approach to setting interim TACs may reduce risk of over- or under-exploitation and, as a more responsive form of advice, improve management performance with respect to both conservation and economic objectives. These interim MPs could also represent an improvement over cases where multi-year catch advice is based on multi-year model projections, which assume average conditions going forward and do not incorporate current information from the system that would be provided by survey data.

In this study, we proposed two interim MPs that differ in how they handle uncertainty in index values. We then compared the two interim MPs to: (i) an annual assessment MP; and (ii)

For Review Only

an MP that keeps the TAC fixed between assessments. We also evaluated the relative performance of the fixed TAC MP and the two interim MPs with varying frequencies of assessments.

We evaluated the performance of the alternative MPs using closed-loop simulation, in which a simulated population is updated over time, based on successive applications of MPs that use simulated data generated from an operating model (OM), subject to observation error and biases (Butterworth & Punt, 1999; Cochrane, Butterworth, De Oliveira, & Roel, 1998). The study addresses three questions:

(1) How well do the interim MPs perform compared to annual assessments?

(2) How well do the interim MPs perform compared to the fixed TAC MPs?

(3) How does assessment frequency affect the relative performance of the interim MPs and fixed TAC MPs?

Methods

Management procedures (MPs)

In year *y*, a typical procedure for setting the TAC of an assessed stock in the following year *y+1* is

$$TAC_{y+1} = \begin{cases} F_{\text{target}} B_{\text{ref}}, & \text{if } y \text{ is an assessment year} \\ TAC_y, & \text{otherwise} \end{cases}, \quad (1)$$

where F_{target} is the target fishing mortality rate and B_{ref} is the reference biomass used as the basis for setting the TAC. We use a simple policy based on MSY, i.e., $F_{\text{target}} = \hat{F}_{\text{MSY}}$ or an appropriate proxy (the circumflex ^ denotes estimates from the assessment model). Here, B_{ref} is the vulnerable biomass in year *y+1* estimated from the assessment, i.e., $B_{\text{ref}} = \hat{B}^V_{y+1}$ (where

For Review Only

superscript V denotes vulnerable biomass). In the interim, the TAC is held constant at the value from the assessment year.

We propose an alternative approach for empirically updating the TAC in the interim years between assessments. First, assume that there is a survey I_y used to index vulnerable biomass in year y,

$$I_y = q\, B_y^V, \tag{2}$$

where q is the catchability coefficient. Equation (2) shows that, with a catchability coefficient, the vulnerable biomass can be updated in the interim from observed index values to provide a catch recommendation,

$$TAC_{y+1} = \begin{cases} F_{\text{target}}\, B_{\text{ref}}, & y \text{ is an assessment year} \\ F_{\text{target}}\, I_y / \hat{q}, & \text{otherwise} \end{cases}, \tag{3}$$

where $\hat{q}$ is the estimated coefficient from the most recent assessment to scale the index values to absolute vulnerable biomass. Equation (3) assumes that the survey provides an index of the fishery vulnerable biomass, i.e., it has similar vulnerability properties. Equation (3) forms the basis for an interim MP.

Uncertainty in the index

Annual estimates of I_y are subject to multiple sources of error, e.g., insufficient sample sizes, hyperstability, and insufficient coverage of stock range by sampling gear. Ideally, adjustments to the TAC should only be made to the degree that the index reflects true changes in the underlying biomass.

For setting the interim TAC, we propose two ways to modify Equation (3) to accommodate error in the index. First, a simple method of smoothing, such as a moving average, dampens the effects of high inter-annual variability in the index and penalizes fluctuations in the

For Review Only

index over time ("Averaged Index"). We use a three-year moving average of the index in the interim,

$$TAC_{y+1} = \begin{cases} F_{\text{target}} B_{\text{ref}}, & \text{if } y \text{ is an assessment year} \\ F_{\text{target}} \tilde{I}_y / \hat{q}, & \text{otherwise} \end{cases}, \tag{4}$$

where

$$\tilde{I}_y = \tfrac{1}{3} \sum_{k=y-2}^{y} I_k. \tag{5}$$

A moving average with a larger time window, e.g., 5 years, could be used if one had a lower tolerance for changing the catch advice (see Supplementary Material A).

Second, the change in catch advice from an assessment can be buffered (reduced) proportional to the uncertainty in the index ("Buffered Index"). To begin with, Equation (3) can be re-written as

$$TAC_{y+1} = \begin{cases} C_{\text{ref}}, & \text{if } y \text{ is an assessment year} \\ C_{\text{ref}} \frac{I_y}{I_{\text{ref}}}, & \text{otherwise} \end{cases}, \tag{6}$$

where $C_{\text{ref}} = F_{\text{target}} B_{\text{ref}}$ is the reference catch and $I_{\text{ref}} = \hat{q}\, B_{\text{ref}}$ is the reference index. Equation (6) shows that the interim approach described by Equation (3) is equivalent to adjusting the reference catch by the ratio of the observed index and the reference index. The TAC increases if the observed index is larger than the reference index and vice versa. Equation (6) follows a similar format to MPs that have been proposed for data-limited fisheries where the reference quantities are determined empirically, e.g., using historical catch, rather than through an assessment model (Geromont & Butterworth, 2015; ICES, 2014).

Changes in the catch advice can then be buffered by adjusting Equation (6) as follows,

$$TAC_{y+1} = \begin{cases} C_{\text{ref}}, & \text{if } y \text{ is an assessment year} \\ C_{\text{ref}} \frac{I_y + b\sigma}{I_{\text{ref}} + b\sigma}, & \text{otherwise} \end{cases}, \tag{7}$$

For Review Only

where σ is the standard deviation of the index and b is an additional non-negative scaling factor that is inversely proportional to one's risk tolerance for changing the catch advice under uncertainty. Large values of either σ or b penalize changes in the catch advice. Equation (7) has the desirable properties of converging to Equations (3) and (6) as either b or σ approaches zero. As b or σ approaches infinity, the TAC becomes fixed until an assessment year.

We calculated σ as the standard deviation of index values from a de-trended time series, for example, the standard deviation of the residuals from the assessment model's fit to the index values. In doing so, one quantifies the uncertainty of the index, incorporating observation and process errors, based on the fit in the assessment. For selecting the value of b, some guidance could be taken from statistical hypothesis testing. With a standard normal distribution for a null hypothesis, critical values of 1 and 2.58 correspond to significance levels of 0.31 and 0.01, respectively, and represent relatively moderate and low tolerance, respectively, for rejecting a null hypothesis. Therefore, values of $b = 1$ and $b = 2.58$ imply moderate and low tolerance, respectively, for changing the catch advice with Equation (7).

Closed-loop simulation study

We evaluated the performance of the MPs using closed-loop simulation, in which alternative methods of setting TACs are applied over time in a simulated fishery system. We used the R package DLMtool (version 5.3.1). Pertinent features of the software are described here, with the full description available in Carruthers and Hordyk (2018). The operating model (OM) is an age-structured model in which the user specifies parameters that define the fish stock, fishing fleet, observation model, and management implementation dynamics of the system. After the historical dynamics of the stock are generated ("historical period"), candidate MPs are

For Review Only

applied successively over time ("projection period"). During the projection period, annual catches are removed from the OM population according to the TAC prescribed from the selected MP, and data are generated by the OM and used to generate the TAC at the next time step. A 50-year projection period was used in the study.

We developed operating models for three species to evaluate the MPs:

(1) Capelin (*Mallotus villosus*) — a short-lifespan forage species. The OM was developed from an assessment using a stochastic stock-reduction analysis (Walters, Martell, & Korman, 2006) based on data available for the Gulf of St. Lawrence stock in Canada (MPO, 2013; Carruthers, 2017);

(2) Vermilion Snapper (*Rhomboplites aurorubens*) — a moderate-lifespan reef species, with dome-shaped vulnerability to the fishery and low recruitment variability. The OM was based on the Gulf of Mexico stock in the U.S. (SEDAR, 2016; Carruthers, 2018a);

(3) Pacific Ocean Perch (POP) (*Sebastes alutus*) — a long-lifespan rockfish species. The OM was based on the Queen Charlotte Sound stock in Canada (DFO, 2017; Carruthers, 2018b).

For all three species, the spawning biomass in the operating model was set to B^{S}_{MSY} where superscript S denotes spawning biomass) at the beginning of the projection period in the Base scenario. Process error was incorporated in the OMs with lognormal deviations in annual recruitment among simulation replicates. For the historical period, the fishing mortality was taken from the most recent assessment and re-scaled by the multiplier that produced the specified biomass at the beginning of the projection period (Table 1, Supplementary Material B). The OMs

For Review Only

were designed to provide insight into the performance of different MPs among contrasting life histories and fleet dynamics and were not intended to fully describe the assessed stocks.

The observation model incorporated variables that control random and persistent error in the generated data, including: (1) hyperstability/hyperdepletion in the index of abundance (i.e., whether trends in the index over time are linearly proportional to changes in biomass); (2) bias (i.e., persistent under-estimates or over-estimates within a simulation) in catch and life history parameters (maturity, growth, natural mortality, and steepness of the stock-recruit relationship); and (3) annual, random sampling error in the index, catch, and age compositions. A full description of the observation model is provided in Table 2.

In the Base scenario, the observed index $\acute{I}_{i,y}$ and catch (in weight) $\acute{C}_{i,y}$ in simulation *i* and year *y* is

$$\acute{I}_{i,y} = q_i\,(B_{i,y})^{\beta_i}\exp\,(\varepsilon^{I}_{i,y}) \tag{8a}$$

and

$$\acute{C}_{i,y} = C_{i,y}\exp\,(\gamma^{C}_{i})\exp\,(\varepsilon^{C}_{i,y}), \tag{8b}$$

respectively, where the acute accent (´) denotes the simulated observation of the corresponding variable in the operating model, q_i represents the scaling parameter for the index, $\beta_i \sim U(0.67, 1.5)$ represents the hyperstability exponent of the simulated index, $\varepsilon^{C}_{i,y} \sim N(-0.5[\sigma^{C}_{i}]^2, [\sigma^{C}_{i}]^2)$ and $\varepsilon^{I}_{i,y} \sim N(-0.5[\sigma^{I}_{i}]^2, [\sigma^{I}_{i}]^2)$ represent independent, normally-distributed deviates in the simulated data "observed" from the operating model, the superscript references the data type (*I* is index and *C* is catch), and $\gamma^{C}_{i} \sim N(-0.001, 0.05^2)$ represents independent, normally-distributed deviates among simulations that parameterize bias in observed catches (among simulations, observed catches are unbiased). Sampled deviates are bias-corrected for the lognormal

For Review Only

distribution. Age compositions were sampled from the catch-at-age matrix from a multinomial distribution with an assumed sample size (Table 2).

For life history parameters such as the Beverton-Holt steepness of the stock-recruit function (h, Mace & Doonan, 1988), the i-th simulated observation $\acute{h}$ is similarly parameterized as

$$\acute{h}_i = h_i \exp\left(\gamma_i^h\right), \tag{9}$$

where $\gamma_i^h \sim N\left(-0.005, 0.1^2\right)$.

The generated data were then passed to the assessment model. The assessment model was a single-fleet statistical catch-at-age (SCA) model from the MSEtool R package version 1.2.1 (Huynh, Carruthers, & Hordyk, 2019; Supplementary Material C). From the assessment, the TAC was generated through the following MPs:

(1) "Annual assessment" — the TAC was set via Equation (1) with an assessment performed every year.

(2) Interim MP with index averaging ("Averaged Index") — the TAC was set by Equation (4) with a 3-year moving average of the observed index.

(3) Interim MP with buffering ("Buffered Index") — the TAC was set by Equation (7) with b = 1 and σ calculated from the standard deviation of the index residuals in the most recent assessment.

(4) "Fixed TAC" — the TAC was set via Equation (1) and remained static until the next assessment.

The estimated values of F_{MSY} and B_{y+1}^{V} from the most recent available assessment were used as F_{target} and B_{ref}, respectively. For Averaged Index, Buffered Index, and Fixed TAC MPs,

For Review Only

assessments were conducted every 10 years, with an additional run that evaluated these MPs with a shorter assessment frequency of 5 years. The TACs were implemented without error in the simulation. Pope's approximation (Pope, 1972) was used to remove the TAC from the population at the midpoint of the year, with a maximum apical fishing mortality set to 3 to prevent unrealistically high levels of fishing effort.

For each species, we ran three scenarios: Base (described above), Hyperstable, and Depleted (Table 3). The Hyperstable scenario evaluated the MPs when the indices used for the assessments and interim TAC calculations were hyperstable, i.e., $\beta << 1$. In our simulation runs, the hyperstable index had a disproportionately smaller change in index values over time compared to an index that is directly proportional to the biomass of the simulated population, as used in the Base and Depleted scenarios. Hyperstability masks depletion trends and arises from factors such as spatial aggregation behavior of fish (Walters, 2003; Walters & Martell, 2004). This scenario evaluates the cost, relative to the Base scenario, of misidentifying an index as proportionate to population size. The Depleted scenario evaluates performance of the MPs when the stock is at a lower biomass level at the start of the projection period (30% B^{S}_{MSY}) than in the Base scenario. This evaluates how well each of the MPs performs for rebuilding stocks. These two sensitivity scenarios were chosen for this study because they exhibit characteristics that tend to generate unfavorable outcomes when not accounted for in the MPs, i.e., lower stock size and reduced fishery yields.

MP evaluation

For Review Only

We evaluated the performance of the MPs in three ways. First, we calculated median trends in OM spawning biomass, fishing mortality, and yield to highlight their behaviors under the different MPs during the projection period.

The second metric evaluated how the MPs changed the biomass over time by calculating the range in relative biomass (RRB). The RRB in simulation *i* over all projection years *y* was calculated as

$$RRB_i = 100 \times \frac{\left[\max_y (B^S_{y,i}) - \min_y (B^S_{y,i})\right]}{B^S_{\mathrm{MSY}(i)}}. \quad (10)$$

RRB was calculated as a percentage which can be used to compare the stability of the population biomass among MPs, where 0% indicates no change and higher values indicate greater change. It is expected that RRB will vary among MPs and assessment intervals. For example, if a fish population were at B^S_{MSY}, a fixed TAC that is lower than MSY would, on average, increase the size of the population over time, whereas an annually adaptive TAC would be more likely to keep the population stable. In many management settings, it is also desirable to avoid high inter-annual variability in catch and the risk of stock collapse, both of which could occur due to high variance in biomass.

Third, we calculated the geometric mean of the relative yield (GMRY) over the entire projection period to evaluate the magnitude of catches achieved after implementation of the MPs. The GMRY in simulation *i* over n_y projection years is

$$GMRY_i = \left(\prod_y C_{i,y}/RefY_i\right)^{1/n_y}, \quad (11)$$

where $RefY$ is the reference yield—the mean annual catch that maximizes the total catch in the projection period with a constant fishing mortality. By accounting for recruitment deviations, $RefY$ does not assume that the population has come to equilibrium by the end of the projection

For Review Only

period. The value of $RefY$ varies among simulations, but approaches MSY as the length of the projection period increases. A GMRY value greater than one implies that the geometric mean of the catches is greater than $RefY$. The GMRY has similar properties to the sum of the logarithm of yield utility function for comparing time series of catch among MPs (Walters & Martell, 2004). The geometric mean penalizes MPs that produce any years with very low catches and, compared to the arithmetic mean, reduces the influence of individual years with very high catches.

We carried out a total of 250 replicate simulations for each OM-MP combination. We checked the stability of the median B^S/B^S_{MSY} in the last year of the projection period, median RRB, and median GMRY to evaluate whether enough simulations had been carried out. Median values of all three were stable with 250 simulations (Supplementary Material D).

Results

This section is organized into three sub-sections for each question posed in the Introduction:

(1) How well do the interim MPs perform compared to annual assessments?

(2) How well do the interim MPs perform compared to the fixed TAC MPs?

(3) How does assessment frequency affect the relative performance of the interim MPs and fixed TAC MPs?

For the first two sub-sections, we present results for the interim MPs and Fixed TAC MPs with a 10-year assessment interval. The third subsection presents results comparing these MPs with 5-year assessment intervals.

For Review Only

How well do interim MPs perform compared to annual assessments?

In the Base scenario, the interim MPs (Averaged Index and Buffered Index MPs) with 10-year assessment intervals and the annual assessment MP performed similarly to each other (Figures 3–5).

Inter-annual variability in median biomass was correlated with life history, with highest variability in a short-lived Capelin and lowest variability in longer-lived POP (Figure 3). The median biomass was around B^S_{MSY} for Vermilion Snapper and POP, and the median fishing mortality near F_{MSY} for all three species (Figures 3-4; first rows). For Capelin, the median biomass was closer to 60% B^S_{MSY}, although this appeared to occur as Pope's approximation degrades with high F (nevertheless, the relative dynamics for Capelin can still be compared). Yield was maintained close to $RefY$ for all three species and for all MPs except the Fixed TAC MP, which showed cyclic behavior corresponding to the intervals between assessments (Figure 5). RRB was low for Vermilion Snapper and POP (<100% in most cases), but high for Capelin (>100% in all cases; Figure 6). GMRY was lowest for Capelin (but with more variability among replicates) and highest for Vermilion Snapper (but with less variability among replicates (Figure 7, top row). GMRY performance was similar among MPs for all species, except for the Fixed TAC MP, which had lower GMRY for Vermilion Snapper.

In the Hyperstable scenario, the median biomass of Capelin declined to just below 50% B^S_{MSY} in the first ten years and remain stabilized at this level (Figure 3, second row). Median biomass for Vermilion Snapper declined at a slower rate to around 50% B^S_{MSY} before beginning to increase in the last ten years of the projection period (Figure 3, second row). The biomass of low-productivity species POP was persistently reduced to very low levels (<< 50% B^S_{MSY}) for all MPs (Figure 3, second row). The Capelin and Vermilion Snapper stocks appeared to be more

For Review Only

resilient to hyperstable indices, with comparatively higher biomass at the end of the projection period (Figure 3).

For all three species, median F was much greater than F_{MSY} (Figure 4, second row). The pattern in yield was characterized by initially large increases that could not be sustained over time (Figure 5, second row), and median GMRY was often lower compared to the Base scenario (Figure 7). This pattern was quite notable for POP with collapses in yield by the end of the projection period under all MPs, while the pattern was less extreme for Capelin and Vermilion Snapper (Figure 5, second row).

In the Depleted scenario, the biomass of all three species increased to higher stock levels under all MPs (Figure 3, third row). Recovery was fastest for Capelin and slowest for POP, where the median trajectory had not reached an asymptote by the end of the projection period. Fishing mortality was immediately reduced for all species under all MPs, with medians near F_{MSY} (Figure 4; third row). However, the fixed TAC MP produced cycles in F for all species, corresponding to the intervals between assessments. Corresponding reductions in yield were observed at the beginning of the projection period, with subsequent increases as the population recovered (Figures 5; third rows). Similar performance among MPs was also observed in the RRB and GMRY metrics, with all MPs having similar median values. This reflects the similar changes in biomass and yields among MPs (Figures 3 and 5).

Values of RRB and GMRY differed among species and scenarios. The median RRB for Capelin was highest, followed by Vermilion Snapper and POP (Figure 6). Compared to the Base scenario, median GMRY was generally lower in the Hyperstable and Depleted scenarios, due to the unsustainable windfall in catch and low stock sizes, respectively, at the beginning of the

For Review Only

projection period (Figure 7). The greatest overall reductions in median GMRY were for POP (Figure 7, third column) and smallest reductions for Capelin (Figure 7, first column).

How well do interim MPs perform compared to the Fixed TAC MP?

Compared to the interim MPs, the Fixed TAC MP produced larger oscillations and occasionally extreme median values in biomass, *F*, and yield for all three species and all three scenarios (Figures 3–5, red lines). The oscillation periods coincided with the 10-year intervals between assessments for the Fixed TAC MP. The interim MPs resulted in less extreme values in biomass and fishing mortality for Vermilion Snapper and POP, respectively, compared to the Fixed TAC MP (Figures 5–6, second and third columns).

In the Base scenario, the oscillations in biomass were most notable for Vermilion Snapper, and more muted for POP. The biomass trajectories for Capelin with the Fixed TAC MP did not appear to be more variable than with the interim MPs.

In the Hyperstable scenario, the oscillations in biomass were still notable for Vermilion Snapper with the Fixed TAC MP, with corresponding oscillations in *F* and yield. Cyclic oscillations in biomass were not, however, produced for the other two species with the Fixed TAC MP (Figures 3-5, second row).

In the Depleted scenario, the interim MPs and the Fixed TAC MP all recovered the biomass to higher stock levels (Figure 3, third row). Median yield initially stayed low under the Fixed TAC MP until the next assessment and was followed by either oscillations for Capelin and Vermilion Snapper or stepwise increases for POP (Figure 5; third row). Under the interim MPs, the yield gradually increased in the first decade and subsequently stabilized for all three species (Figure 5; third row).

For Review Only

The oscillations in biomass associated with the Fixed TAC MP were reflected in the higher median RRB compared to both the Averaged Index and the Buffered Index MPs (Figure 6). The differences between the Fixed TAC MP and the two interim MPs were related to longevity, with the smallest differences for POP and largest differences for Capelin. Differences in RRB were smallest in the Hyperstable scenario compared to the Base and Depleted scenarios due to the tendency of the biomass to decrease in the Hyperstable scenario.

The Fixed TAC MP also generated smaller median GMRY compared to the interim MPs (with 10-year assessment intervals; Figure 7). For Capelin and Vermilion Snapper, the decreases in median GMRY from the interim MPs to the Fixed TAC MP were most notable in the Base and Depleted Scenarios, while for POP, the decrease was most notable in the Hyperstable Scenario.

Assessment frequency

When the assessment frequency was increased from 10-year to 5-year intervals, the performance of the interim MPs did not appreciably change in terms of median trajectories, RRB, and GMRY (Supplementary Material E; Figures 6-7). On the other hand, for the Fixed TAC MP, the trajectories of biomass, fishing mortality, and yield more closely approximated those of the Annual Assessment, Averaged Index, and Buffered Index MPs (Supplementary Material E). This was most apparent for Vermilion Snapper (Figures 6 and 7, second column). With an increased assessment frequency, the extremes in the Fixed TAC trajectories were greatly reduced, with the trajectories more closely matching those of the other three MPs (Supplementary Material E). For the Fixed TAC MP, the median RRB also decreased and the

For Review Only

462 median GMRY increased with the shorter assessment interval (Figures 6–7). Larger reductions
463 in median RRB corresponded to larger increases in median GMRY.
464
465 **Discussion**
466 In this study, we presented interim MPs for updating catch advice between stock
467 assessments and highlighted their performance relative to annual assessments and to fixed TACs
468 with three life history archetypes and three alternative OM scenarios. While the interim MPs
469 generally kept the biomass stable in the Base scenario, trends in biomass varied widely in the
470 Hyperstable scenario depending on life history. For Capelin, the short generation time and high
471 productivity of the stock appeared to minimize biological risk and sustain relatively high yields.
472 For Vermilion Snapper, vulnerability to the fishery was limited to mature animals, which in
473 combination with dome-shaped fishery vulnerability, appeared to reduce the depletion risk
474 compared to the other two species. The biomass of POP was lowest among the three species due
475 to its low productivity and vulnerability to fishing of some immature age classes.
476 The assessment interval and the ability to update the catch advice more frequently were
477 particularly important in the Hyperstable and Depleted scenarios. Under the Hyperstable
478 scenario, fewer extreme values of median *F* were observed when either: (1) the assessment
479 frequency with the Fixed TAC MP was reduced; or (2) an interim MP was used instead of the
480 Fixed TAC MP. Under the Depleted scenario, foregone yield occurred when the population
481 recovered during periods of low TAC under the Fixed TAC MP with 10-year assessment
482 intervals. Interim MPs, would therefore be more cost-effective means of providing robust TAC
483 advice than either annual assessments or more frequent assessments with fixed TACs, due to the
484 lower required assessment frequency.

For Review Only

From this study, we recommend routine evaluation of the assessment interval when testing MPs for assessed stocks (e.g., Hutniczak, Lipton, Wiedenmann, & Wilberg, 2019; Wiedenmann, Wilberg, Sylvia, & Miller, 2017). While Zimmerman and Enberg (2017) have reported only minor loss in precision in the performance of stock assessment models with less frequent assessments, they only evaluated annual versus biennial assessments and did not evaluate situations where the assumptions of the estimation model were violated, such as the proportionality of the index to the underlying population. Their study also did not account for feedback between the system dynamics and management advice using closed-loop simulation.

Additional scenarios could be evaluated in the future for other stocks of interest. For example, one could evaluate how the interim MPs would perform for lightly exploited stocks where effort is increasing (e.g., a new or developing fishery). Another example would be a situation where an index has hyperdepletion properties to evaluate the potential loss of yield due to an overly pessimistic survey. The Hyperstable scenario in this study showed that a non-proportional index of abundance could lead to poor outcomes under all tested MPs, especially for a long-lived stock such as POP. In a situation where it is recognized that no proportional index exists, non-index based approaches could be evaluated for providing management advice (e.g., Geromont and Butterworth, 2015; Klaer, Wayte, and Fay, 2012; Thorson and Cope, 2015). Closed-loop simulation comparison of such methods with assessment-based MPs and other index-based MPs would quantify the value of developing a better survey for the stock.

Interim MP implementation

Interim MPs could be tuned to meet specific management goals. For example, we tested the interim MPs here with a simple MSY-based control rule ($F_{\text{target}} = F_{\text{MSY}}$), but if management

For Review Only

desires to avoid overfishing, it would be prudent to choose a more precautionary F_{target} that is less than F_{MSY}. A ramped harvest control rule, such as the 40-10 control rule, could be implemented such that F_{target} is linearly reduced from its maximum value if the reference biomass, e.g. spawning biomass, is below the upper reference point (Deroba & Bence, 2008; Restrepo & Powers, 1999). In an interim MP with such a harvest control rule, an increasing index could be used to update the biomass estimate and increase F_{target} towards the upper threshold reference point.

Additionally, reference points based on MSY can be imprecise, largely due to uncertainty in the steepness of the stock-recruit relationship (Forrest, Holt, & Kronlund, 2018; Haltuch, Punt, & Dorn, 2009). Alternative reference points based on quantities such as the spawning potential ratio (SPR; Clark 1991, 2002), or average estimated historical biomass (Forrest, Holt, & Kronlund, 2018) may be substituted if they can be shown to improve performance. Benchmark performance metrics could be calculated to identify whether the configuration and implementation of the interim MPs meet pre-specified management thresholds.

Many management systems require rebuilding plans in order to increase depleted population levels within a specified timeframe (e.g., DFO, 2009; PFMC, 2016). Assuming a fixed fishing mortality rate, recovery time will be driven by life history (Carruthers and Agnew, 2016). In our study, we found that the median spawning biomass of POP in the Depleted scenario still did not reach B^{S}_{MSY} during the 50-year projection period, due to its long rebuilding time. In the context of the interim MPs, the duration of rebuilding could be decreased by decreasing F_{target}, which necessitates further decreases in short-term yield for faster recovery. Specific performance metrics, such as the probability of rebuilding within a certain timeframe,

For Review Only

can be developed in order to evaluate potential F_{target} values and the tradeoff between rebuilding time and short-term yield (Wetzel & Punt, 2016).

There are differences in how perceived uncertainty is addressed in deriving the catch advice between our Buffered Index and Averaged Index MPs. The moving average in the Averaged Index MP is intended to reduce annual changes in the catch advice despite sometimes large inter-annual variability in the index. In contrast, annual changes in the TAC from the Buffered Index MP was primarily determined by the fit of the assessment to the simulated data, i.e., when the assessment fit the data well, the catch advice predominantly followed the change in the index. In this study, we tuned the Averaged Index and Buffered Index MPs to allow for moderate changes in the catch advice. Although both performed similarly in this study, there is no straightforward way to tune the two interim MPs such that they handle uncertainty equivalently. Empirical testing via simulation could be used to evaluate outcomes under alternative tuning options, e.g., length of moving average and magnitude of the risk buffer for the Averaged Index and Buffered Index MPs, respectively. Catches would more closely track variability in surveyed abundance with shorter moving averages, and lower risk buffers, respectively.

Data availability can influence the success of an assessment-based MP. Here, we assumed that for an assessment, data were available from the current year (end of season), and assessment and TAC advice were generated for the next year. Time lags between data availability and assessment can lead to an improper match between the information from an assessment and conditions at the time of implementation of the catch advice (Kell, et al. 2005; Hutniczak, Lipton, Wiedenmann, & Wilberg, 2019). This can be important for short-lived stocks or for fisheries that target young cohorts not accounted for from the most recent assessment.

For Review Only

Further evaluation of the interim MPs will need to account for the time lags in the management system. Additionally, greater efficiencies in personnel time may be possible if there are systematic analytical and reporting mechanisms for survey data (e.g., Anderson, Keppel, & Edwards, in press) for the interim MPs as part of a strategy to streamline assessment processes.

The interim MPs presented here are conditional on the state of knowledge assumed in the most recent assessment. In management settings such as the Federal fisheries in the United States, there are often two types of stock assessments: benchmark and update assessments. The benchmark assessment considers all available information and potentially revises the structure and assumptions of the assessment model (e.g., SEDAR, 2018). Update assessments, on the other hand, use the same structural assessment model as the benchmark, but can add any new time series data produced since the last benchmark assessment (e.g., Cass-Calay et al. 2015). This study did not consider cases in which a benchmark assessment significantly revises the structure and information from the previous assessment. New assessments are occasionally required in cases where new life-history information becomes available—for example when a new growth study informs changes in the asymptotic length or size of maturity, or systemic changes in the fishery have occurred with a change in sizes of vulnerability or retention due to management regulations or market demands. Such changes can affect the appropriate fishing mortality rate, and new assessments should be conducted in response to these changes. Sensitivity explorations can evaluate how misspecification of assessment models affect performance of interim MPs.

For the interim MPs, we also assumed that the survey vulnerability matches that of the fishery. This occurs in cases where fishery catch-per-unit-effort is used to index the population

For Review Only

575 or if the survey uses the same gear as the fishery. Future case studies can evaluate how interim
576 MPs perform if survey and fishery vulnerability are substantially different.
577
578 ***Strategic management***
579 The development and application of quantitative stock assessments and their subsequent
580 updating to produce catch advice over time require considerable resources which are not
581 available for all stocks (Newman, Berkson, & Suatoni, 2015). This study demonstrated the
582 strategic advantages of using an interim MP over Annual Assessment and Fixed TAC MPs in the
583 assessment and management of exploited stocks. First, our interim MPs performed similarly to
584 an Annual Assessment MP. A reduction of the assessment frequency from an annual interval to
585 as much as a 10-year interval gave similar performance, in terms of the stock dynamics and
586 achieved yield. This would substantially reduce the scientific capacity needed to support
587 management. It is also especially important for rebuilding stocks, for which the TAC advice can
588 quickly become outdated as the population size increases.
589 Second, if the assessment interval cannot be decreased, then it would be preferable to
590 adopt an interim MP rather than simply fixing the TAC. Performance of the Fixed TAC MP was
591 improved by reducing the assessment interval from 10 to 5 years. As more time passes between
592 assessments, the catch advice from the Fixed TAC became further detached from current
593 biomass levels. In both cases, an older assessment that would be perceived as outdated under a
594 fixed TAC management approach remains relevant in an interim approach so long as major
595 departures from the assumptions in the assessment have not occurred.
596 Third, the interim MP approach can alleviate the time and resources needed to produce
597 assessments as frequently, and the time can be focused on assessing other managed stocks. For

For Review Only

example, despite periodic stock assessments for most stocks in the U.S., the number of updated assessments tends to be low relative to the number of managed stocks. Reducing the frequency of stock assessment by using the interim approach could allow additional resources to focus on the approximately 70% of U.S. federal fisheries that are managed by data-limited approaches (Neubauer et al. 2018; Newman, Berkson, & Suatoni, 2015). Similar benefits could exist for other regions around the world where frequent stock assessments are costly and time consuming.

Acknowledgements

This work was funded by Grants and Contribution Funds (Ocean and Freshwater Science Contribution Program) from Fisheries and Oceans Canada. AH acknowledges funding from the Natural Resources Defense Council. Adam Keizer (Fisheries and Oceans Canada) kindly provided time series of TACs for BC groundfish stocks. Comments on an earlier draft from an anonymous reviewer, José de Oliveira, and Jaclyn Cleary improved this manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest in this manuscript.

Data Sharing Policy

The R code that supports this study is available at: https://doi.org/10.5281/zenodo.3380607

For Review Only

References

Anderson, S.C., Keppel, E.A., & Edwards, A.M. (in press). A reproducible data synopsis for over 100 species of British Columbia groundfish. DFO Canadian Science Advisory Secretariat (CSAS) Research Document 2019/041.

Butterworth, D.S., & Punt, A.E. (1999). Experiences in the evaluation and implementation of management procedures. ICES Journal of Marine Science, *56*, 985–998. https://doi.org/10.1006/jmsc.1999.0532

Carruthers, T.R. (2017). Capelin in the Gulf of St Lawrence: a DFO Case Study Operating model. Retrieved from http://www.datalimitedtoolkit.org/Case_Studies_Table/Capelin_GSL_DFO/Capelin_GSL.html [accessed April 8, 2018]

Carruthers, T.R. (2018a). Operating model for Gulf of Mexico Vermilion Snapper (*Rhomboplites aurorubens*). Retrieved from http://www.datalimitedtoolkit.org/Case_Studies_Table/Vermillion_Snapper_GOM_NOAA/Vermillion_Snapper_GOM_NOAA.html [accessed October 2018]

Carruthers, T.R. (2018b). Pacific Ocean Perch QC BC: a DFO Case Study Operating model. Retrieved from: http://www.datalimitedtoolkit.org/Case_Studies_Table/Pacific_Ocean_Perch_QC_BC_DFO/Pacific_Ocean_Perch_QC_BC_DFO.html [accessed October 2018]

Carruthers, T.R. & Agnew, D.J. (2016). Using simulation to determine standard requirements for recovery rates of fish stocks. Marine Policy, *73*, 146-153. https://doi.org/10.1016/j.marpol.2016.07.026

For Review Only

Carruthers, T.R. & Hordyk A.R. (2018). The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. Methods in Ecology and Evolution, *9*, 2388-2395. https://doi.org/10.1111/2041-210X.13081

Carruthers, T.R., Kell, L.T., Butterworth, D.D.S, Maunder, M.N., Geromont, H.F., Walters, C., …, Davies, C.R. (2016). Performance review of simple management procedures. ICES Journal of Marine Science, 73, 464-482. https://doi.org/10.1093/icesjms/fsv212

Cass-Calay, S.L., Porch, C.E., Goethel, D.R., Smith, M.W., Matter, V., & McCarthy, K.J. (2015). Stock Assessment of Red Snapper in the Gulf of Mexico 1872-2013 – with provisional 2014 landings. SEDAR Red Snapper 2014 Update Assessment. Prepared for the Science and Statistical Committee of the Gulf Fishery Management Council. Retrieved from http://sedarweb.org/docs/suar/SEDARUpdateRedSnapper2014_FINAL_9.15.2015.pdf [Accessed April 8, 2019]

Clark, W.G. (1991). Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries and Aquatic Science, *48*, 734–750. https://doi.org/10.1139/f91-088

Clark, W.G. (2002). F35% revisited ten years later. North American Journal of Fisheries Management, *22*, 251–257. https://doi.org/10.1577/1548-8675(2002)022%3C0251:FRTYL%3E2.0.CO;2

Cochrane, K L., Butterworth, D.S., De Oliveira, J.A.A., & Roel, B.A. (1998). Management procedures in a fishery based on highly variable stocks and with conflicting objectives: experiences in the South African pelagic fishery. Reviews in Fisheries Biology and Fisheries, *8*, 177-214. https://doi.org/10.1023/A:1008894011847

For Review Only

Deroba, J.J., & Bence, J.R. (2008). A review of harvest policies: Understanding relative performance of control rules. Fisheries Research, *94*, 210-223. https://doi.org/10.1016/j.fishres.2008.01.003

DFO. (2009). A fishery decision-making framework incorporating the precautionary approach. Retrieved from http://www.dfo-mpo.gc.ca/reports-rapports/regs/sff-cpd/precaution-eng.htm [Accessed April 8, 2019]

DFO. (2012). Stock Assessment for the inside population of Yelloweye Rockfish (*Sebastes ruberrimus*) In British Columbia, Canada for 2010. DFO Canadian Science Advisory Secretariat (CSAS) Science Advisory Report 2011/084 13p.

DFO. (2017). Pacific Ocean Perch (*Sebastes alutus*) stock assessments for Queen Charlotte Sound, British Columbia in 2017. DFO Canadian Science Advisory Secretariat (CSAS) Science Advisory Report 2017/043. Available from http://waves-vagues.dfo-mpo.gc.ca/Library/40646737.pdf [accessed April 8, 2019]

DFO. (2018). Stock assessment for Pacific Herring (*Clupea pallasii*) in British Columbia in 2017 and forecast for 2018. DFO Canadian Science Advisory Secretariat (CSAS) Science Advisory Report 2018/002.

DFO. (2019). Pacific Region Integrated Fisheries Management Plan. Groundfish, Effective February 21, 2019, Version 1.1. Retrieved from http://www.pac.dfo-mpo.gc.ca/fm-gp/mplans/ground-fond-ifmp-pgip-sm-eng.pdf [Accessed August 7, 2019].

Edwards, A.M., Taylor, I.G., Grandin, C.J., and Berger, A.M. (2018). Status of the Pacific Hake (whiting) stock in U.S. and Canadian waters in 2018. Prepared by the Joint Technical Committee of the U.S. and Canada Pacific Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. 222 p.

For Review Only

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

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

Geromont, H.F., & Butterworth, D.S. (2015). Generic management procedures for data-poor fisheries: forecasting with few data. ICES Journal of Marine Science, *72*, 251-261. https://doi.org/10.1093/icesjms/fst232

Haltuch, M.A., Punt, A.E., & Dorn, M.W. (2009). Evaluating the estimation of fishery management reference points in a variable environment. Fisheries Research, *100*, 42-56.

Harford, W.J., & Carruthers, T.R. (2017). Interim and long-term performance of static and adaptive management procedures. Fisheries Research, *190*, 84-94. https://doi.org/10.1016/j.fishres.2009.03.001

Hutniczak, B., Lipton, B., Wiedenmann, J., & Wilberg, M. (2019). Valuing changes in frequency of fish stock assessments. Canadian Journal of Fisheries and Aquatic Science, *76*, 1640-1652. https://doi.org/10.1139/cjfas-2018-0130

Huynh, Q.C., Hordyk, A. R., & Carruthers, T. R. (2019). MSEtool: Management Strategy Evaluation Toolkit. R package version 1.2.1. Available at: https://cran.r-project.org/package=MSEtool

Kell, L.T., Pilling, G.M., Kirkwood, G.P., Pastoors, M., Mesnil, B., …, Ulrich-Rescan, C. (2005). An evaluation of the implicit management procedure used for some ICES roundfish

For Review Only

stocks. ICES Journal of Marine Science, *62*, 750-759. https://doi.org/10.1016/j.icesjms.2005.01.001

Klaer, N.L., Wayte, S.E., & Fay, G. (2012). An evaluation of the performance of a harvest strategy that uses an average-length-based assessment method. Fisheries Research, *134-136*, 42-51. https://doi.org/10.1016/j.fishres.2012.08.010

ICES. (2014). Evaluation of a multi-annual plan including an index based HCR for North Sea horse mackerel, 17-18 June 2014, IJmuiden, the Netherlands. ICES CM 2014/ACOM:56. 102 pp.

Mace, P.M., & Doonan, I.J. (1988). A generalized bioeconomic simulation model for fish population dynamics. New Zealand Fisheries Assessment Research Document 88/4.

Methot, R.D., & Wetzel, C. (2012). 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

MPO. (2013). Assessment of the Estuary and Gulf of St. Lawrence (Divisions 4RST) Capelin Stock in 2012. DFO Canadian Science Advisory Secretariat (CSAS) Science Advisory Report 2013/021.

Neubauer, P., Thorson, J.T., Melnychuk, M.C., Methot, R., & Blackhart, K. (2018). Drivers and rates of stock assessments in the United States. PLoS ONE 13(5), e0196483. https://doi.org/10.1371/journal.pone.0196483

Newman, D., Berkson, J., & Suatoni, L. (2015). Current methods for setting catch limits for data-limited fish stocks in the United States. Fisheries Research, *164*, 86-93. https://doi.org/10.1016/j.fishres.2014.10.018

For Review Only

Pedersen, M.W., & Berg, C.W. (2017). A stochastic surplus production model in continuous time. Fish and Fisheries, *18*, 226-243. https://doi.org/10.1111/faf.12174

PFMC (Pacific Fishery Management Council). (2016). Pacific Coast Groundfish Fishery Management Plan for the California, Oregon and Washington Groundfish Fishery, Appendix F: Overfished Species Rebuilding Plans, August 2016. 20 pages.

Pope, J.G. (1972). An Investigation of the Accuracy of Virtual Population Analysis Using Cohort Analysis. ICNAF Research Bulletin, *9*, 65-74.

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

Restrepo, V.R., & Powers, J.E. (1999). Precautionary control rules in US fisheries management: specification and performance. ICES Journal of Marine Science, *56*, 846-852. https://doi.org/10.1006/jmsc.1999.0546

Sagarese, S.R., Rios, A.B., Cass-Calay, S.L., Cummings, N.J., Bryan, M.D., …, Matter, V.M. (2018). Working Towards a Framework for Stock Evaluations in Data-Limited Fisheries. North American Journal of Fisheries Management, *38*, 507-537. https://doi.org/10.1002/nafm.10047

SEDAR (Southeast Data, Assessment and Review). (2013). SEDAR 28: Stock Assessment Report for Gulf of Mexico Cobia. Available from https://sedarweb.org/sedar-28-stock-assessment-report-gulf-mexico-cobia [accessed April 8, 2019]

SEDAR (Southeast Data, Assessment and Review). (2016). SEDAR 45: Stock Assessment Report for Gulf of Mexico Vermilion Snapper. Available from https://sedarweb.org/sedar-45-final-stock-assessment-report-gulf-mexico-vermilion-snapper [accessed April 8, 2019]

For Review Only

SEDAR (Southeast Data, Assessment and Review). (2018). SEDAR 52: Stock Assessment Report for Gulf of Mexico Red Snapper. Available from https://sedarweb.org/sedar-52-gulf-mexico-red-snapper-final-stock-assessment-report [accessed April 8, 2019]

Shertzer, K.W., & Prager, M.H. (2007). Delay in fishery management: diminished yield, longer rebuilding, and increased probability of stock collapse. ICES Journal of Marine Science, *64*, 149–159. https://doi.org/10.1093/icesjms/fsl005

Stewart, I., & Hicks, A. (2018). Assessment of the Pacific halibut (*Hippoglossus stenolepis*) stock at the end of 2017. IPHC-2018-AM094-10. Retrieved from: https://iphc.int/uploads/pdf/am/2018am/iphc-2018-am094-10.pdf [accessed April 8, 2019]

Thorson, J.T., & Cope, J.M. (2015). Catch curve stock-reduction analysis: An alternative solution to the catch equations. Fisheries Research, *171*, 33-41. https://doi.org/10.1016/j.fishres.2014.03.024

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. Canadian Journal of Fisheries and Aquatic Science, *60*, 1433-1436. https://doi.org/10.1139/f03-152

Walters, C.J., Martell, S.J.D., & Korman, J. (2006). A stochastic approach to stock reduction analysis. Canadian Journal of Fisheries and Aquatic Science, *63*, 212-223. https://doi.org/10.1139/f05-213

Walters, C.J., & Martell, S.J.D. (2004). Fisheries Ecology and Management. Princeton University Press, Princeton, NJ, 338 pp.

Wetzel, C.R., & Punt, A.E. (2016). The impact of alternative rebuilding strategies to rebuild overfished stocks. ICES Journal of Marine Science, *73*, 2190-2207. https://doi.org/10.1093/icesjms/fsw073

For Review Only

Wiedenmann, J., Wilberg, M., Sylvia, A., & Miller, T. (2017). An evaluation of acceptable biological catch (ABC) harvest control rules designed to limit overfishing. Canadian Journal of Fisheries and Aquatic Science, *74*, 1028-1040. https://doi.org/10.1139/cjfas-2016-0381

Zimmermann, F., & Enberg. K. (2017). Can less be more? Effects of reduced frequency of surveys and stock assessments. ICES Journal of Marine Science, *74*, 56-68. https://doi.org/10.1093/icesjms/fsw134

For Review Only

Tables

Table 1. Range in operating model parameters for the four stocks. Additional life history (growth and maturity) and vulnerability schedules are shown in Figure 2. Parameters do not vary among simulations except for the standard deviation of recruitment, which is drawn from a uniform distribution for a given simulation iteration. Parameters were based on the most recently available stock assessment for each species (see Methods).

Parameter	**Capelin**	**Vermilion Snapper**	**Pacific Ocean Perch (POP)**
Maximum age (years)	8	15	60
Historical years of fishing	53	65	77
Spawning depletion at MSY (B^S_{MSY}/B^S_0)	0.32	0.35	0.29
Fishing mortality at MSY (F_{MSY}) (year^{-1})	1.00	0.34	0.08
Natural mortality (M) (year^{-1})	0.80	0.25	0.06
von Bertalanffy asymptotic growth coefficient (K) (year^{-1})	0.55	0.33	0.16
von Bertalanffy asymptotic length (L_∞) (cm)	19.0	34.4	44.2
von Bertalanffy location parameter (t_0) (year)	-0.50	-0.80	-0.65
Beverton-Holt stock-recruit steepness (h)	0.60	0.60	0.70
Recruitment variability standard deviation (lognormal distribution)	0.48 – 1.00	0.20 - 0.39	0.65 – 0.96

For Review Only

Table 2. Observation model parameters common to the operating models for the three species. Subscripts indicate whether values are vary among simulations *i*, years *y*, or both.

Parameter	**Symbol**	**Distribution**
Index hyperstability exponent[1]	β_i	Base and Depleted: $U(0.67, 1.5)$ Hyperstable: $U(0.33, 0.67)$
Annual observation error in the index[1]	$\varepsilon_{i,y}^{I}$	$\varepsilon_{i,y}^{I} \sim N(-0.5[\sigma_i^C]^2, [\sigma_i^C]^2)$, where $\sigma_i^I \sim U(0.10,\ 0.25)$
Annual observation error in the catch[2]	$\varepsilon_{i,y}^{C}$	$\varepsilon_{i,y}^{C} \sim N(-0.5[\sigma_i^I]^2, [\sigma_i^I]^2)$, where $\sigma_i^C \sim U(0.1, 0.2)$
Age composition effective sample size [multinomial sampling]	O_i	$O_i \sim U(50, 100)$
Bias in observed catch [among simulations][2]	γ_i^C	$\gamma_i^C \sim N(-0.001, 0.05^2)$
Bias in observed steepness (h)[3]	γ_i^h	$\gamma_i^h \sim N(-0.005, 0.10^2)$
Bias in observed natural mortality (M), ages of 50% and 9 % maturity (m_{50} and m_{95}, respectively), and von Bertalanffy growth parameters (L_∞, K, and t_0)[4]	$\gamma_i^M, \gamma_i^{m_{50}}, \gamma_i^{m_{95}}, \gamma_i^{L_\infty}, \gamma_i^K, \gamma_i^{t_0}$	$N(-0.001, 0.05^2)$

[1] See Equation (8a); [2] See Equation (8b); [3] See Equation (9); [4] Similar to Equation (9)

For Review Only

Table 3. Summary of operating model scenarios.

Name	**Description**
Base	Parameters in Tables 1–2 and Figure 2, spawning biomass is B^S_{MSY} at the beginning of the projection period
Hyperstable	Same as Base, but $\beta_i \sim U(0.33, 0.67)$
Depleted	Same as Base, but spawning biomass is 30% B^S_{MSY} at the beginning of the projection period

For Review Only

Figure Captions

Figure 1. TAC (total allowable catches) for a subset of Canadian Pacific stocks managed under the Integrated Groundfish Management Plan (DFO, 2019). Numbers (and colors) indicate the species in alphabetical order: (1) Big Skate, (2) Dover Sole, (3) Halibut, (4) Lingcod, (5) Longnose Skate, (6) Pacific Cod, (7) Pacific Ocean Perch, (8) Petrale Sole, (9) Quillback Rockfish, (10) Redstripe Rockfish, (11) Rougheye Rockfish, (12) Sablefish, (13) Walleye Pollock, (14) Widow Rockfish, (15) Yelloweye Rockfish, (16) Yellowtail Rockfish. Values were provided by Fisheries and Oceans Canada.

Figure 2. Life history (growth and maturity) and vulnerability schedules at age used in the operating models for Capelin, Vermilion Snapper and POP. Growth is expressed as mean length-at-age relative to that at the maximum age.

Figure 3. Annual median B^S/B^S_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of the assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are shown in the Supplementary Material.

Figure 4. Annual median F/F_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are shown in the Supplementary Material.

Figure 5. Annual medians of relative yield from 250 simulations for each species (columns) and scenario (rows). The relative yield is the ratio of the catch to the reference yield. Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are in the Supplementary Material.

Figure 6. Dot-and-whisker plots of *RRB* (range in relative biomass) for each species (columns) and scenario (rows). For each MP, dots and numbers indicate the median from 250 simulations, and whiskers span the interquartile range.

Figure 7. Dot-and-whisker plots of *GMRY* (geometric mean of relative yield) for each species (columns) and scenario (rows). For each MP, dots and numbers indicate the median from 250 simulations, and whiskers span the interquartile range. Dotted, horizontal lines indicate a value of 1.

For Review Only

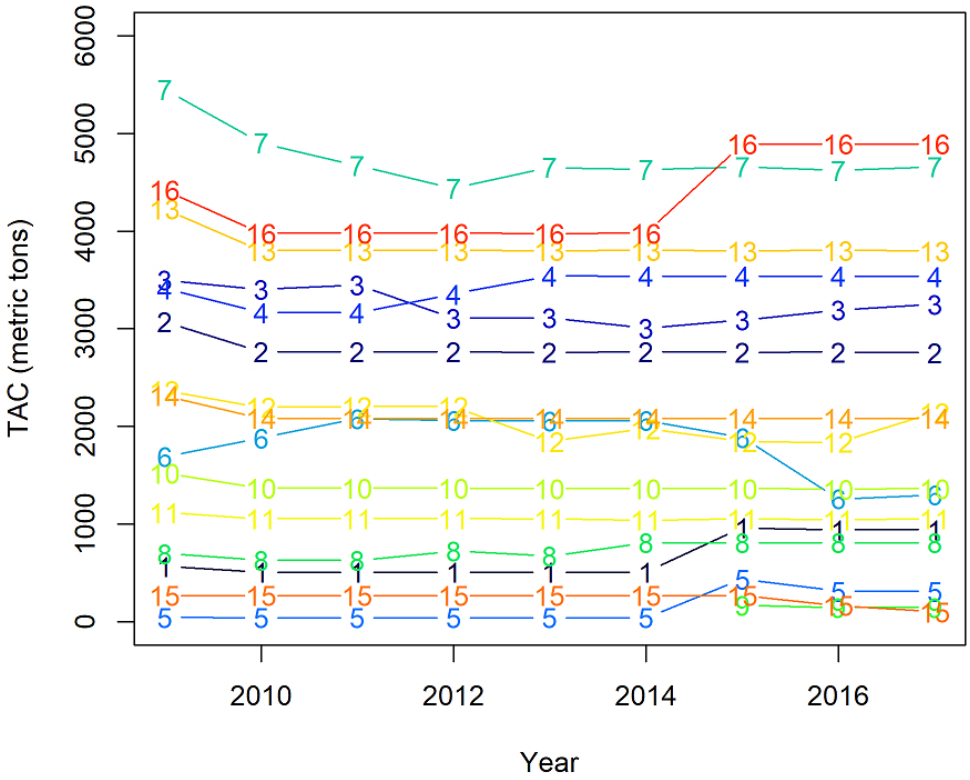

Figure 1. TAC (total allowable catches) for a subset of Canadian Pacific stocks managed under the Integrated Groundfish Management Plan (DFO, 2019). Numbers (and colors) indicate the species in alphabetical order: (1) Big Skate, (2) Dover Sole, (3) Halibut, (4) Lingcod, (5) Longnose Skate, (6) Pacific Cod, (7) Pacific Ocean Perch, (8) Petrale Sole, (9) Quillback Rockfish, (10) Redstripe Rockfish, (11) Rougheye Rockfish, (12) Sablefish, (13) Walleye Pollock, (14) Widow Rockfish, (15) Yelloweye Rockfish, (16) Yellowtail Rockfish. Values were provided by Fisheries and Oceans Canada.

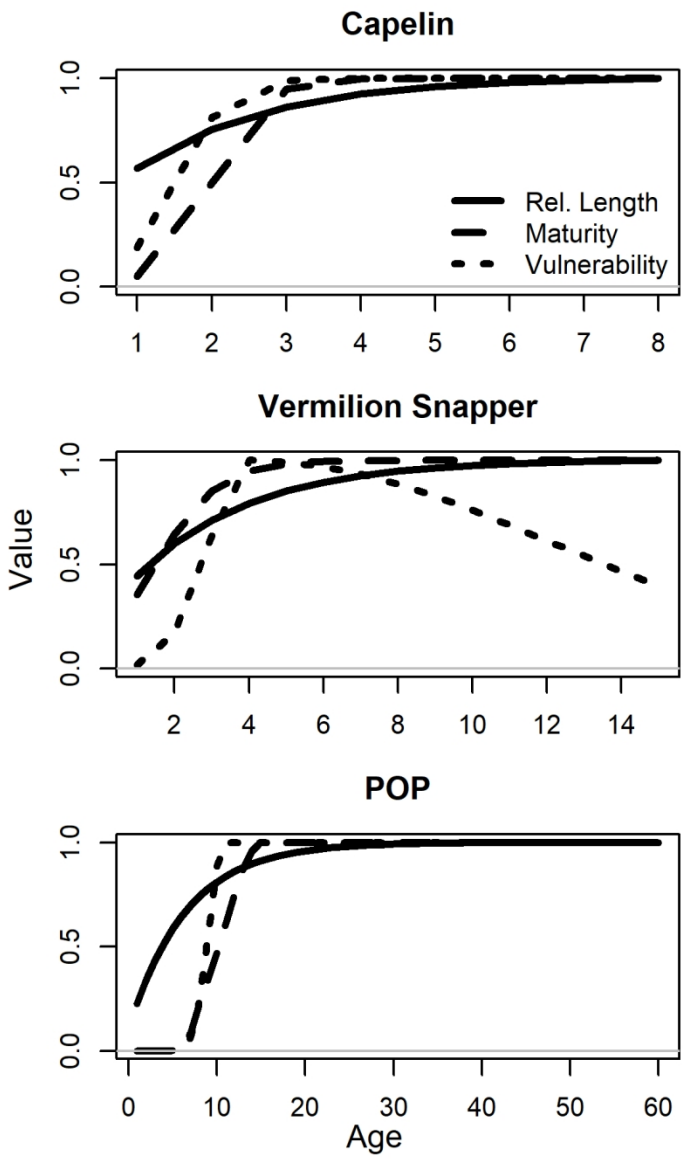

Figure 2. Life history (growth and maturity) and vulnerability schedules at age used in the operating models for Capelin, Vermilion Snapper and POP. Growth is expressed as mean length-at-age relative to that at the maximum age.

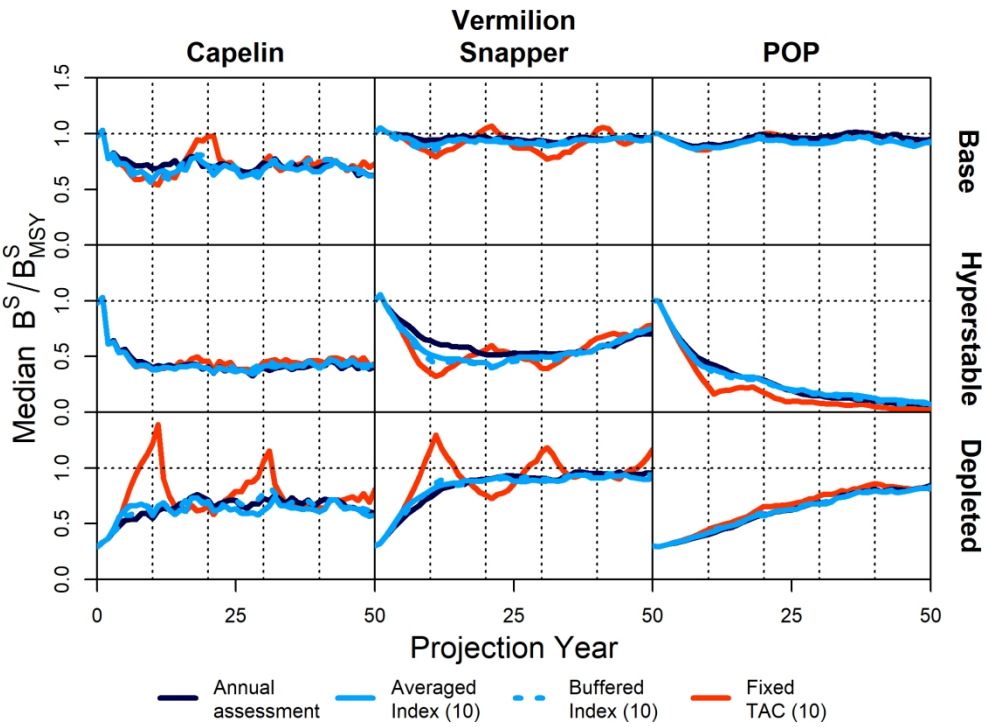

Figure 3. Annual median B^S/B^S_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of the assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are shown in the Supplementary Material.

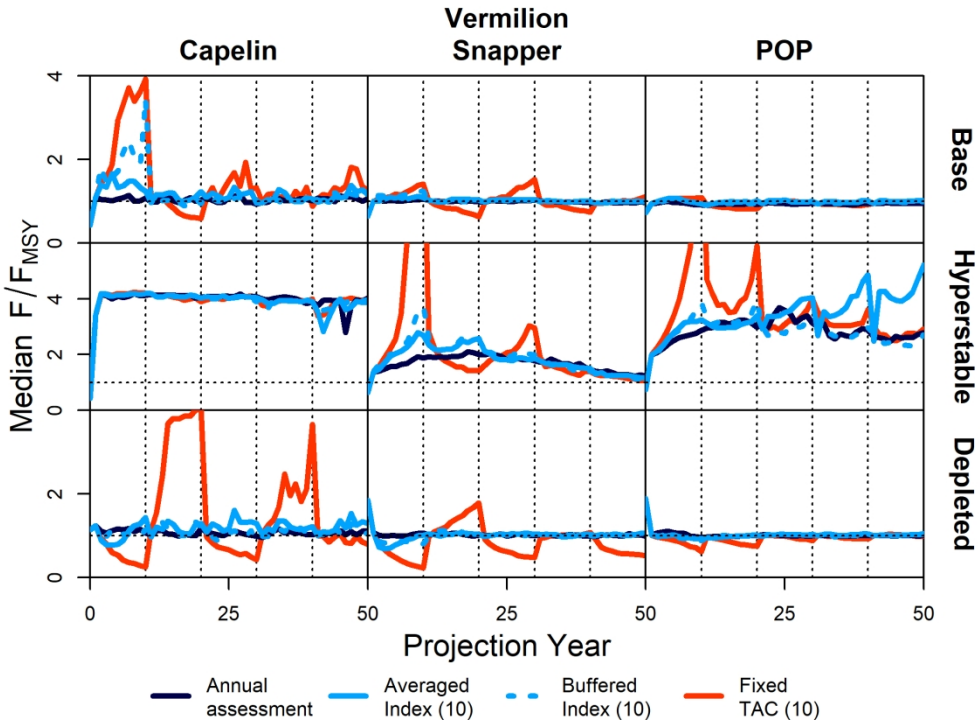

Figure 4. Annual median F/F_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are shown in the Supplementary Material.

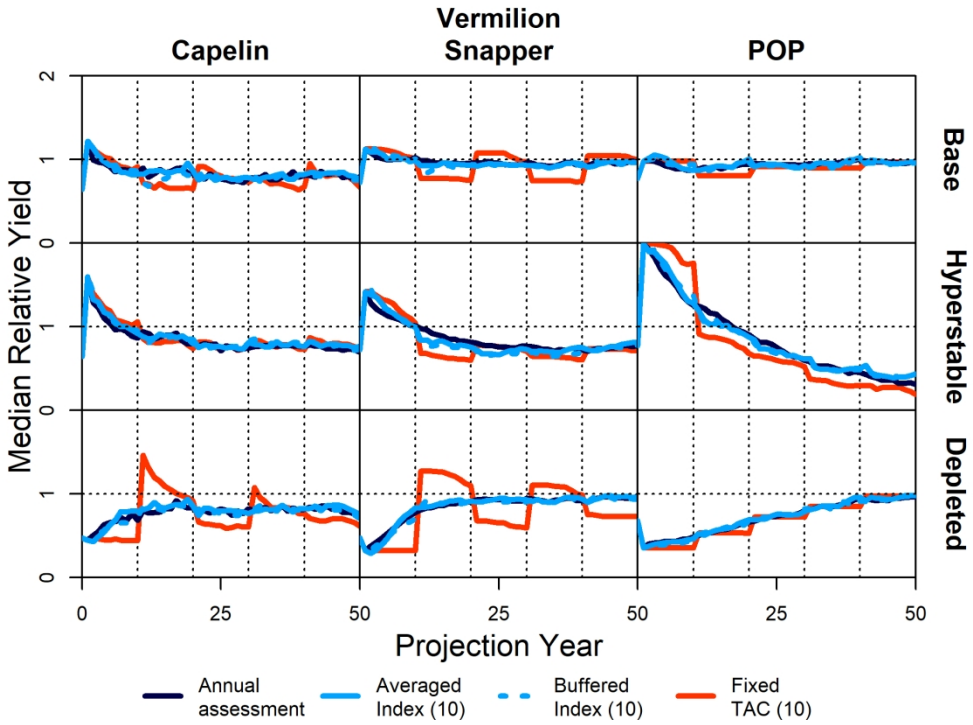

Figure 5. Annual medians of relative yield from 250 simulations for each species (columns) and scenario (rows). The relative yield is the ratio of the catch to the reference yield (*RefY*). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for the Annual assessment MP). Parentheses in the legend indicate the assessment interval. Medians for MPs with 5-year assessment intervals are in the Supplementary Material.

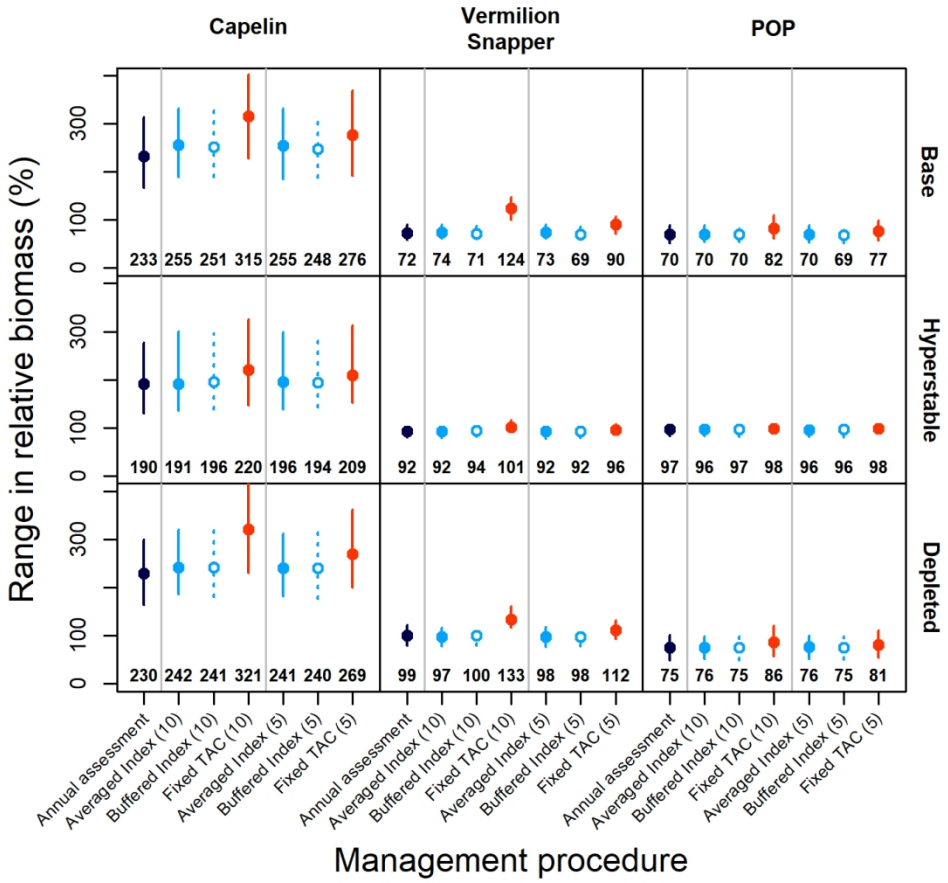

Figure 6. Dot-and-whisker plots of *RRB* (range in relative biomass) for each species (columns) and scenario (rows). For each MP, dots and numbers indicate the median from 250 simulations, and whiskers span the interquartile range.

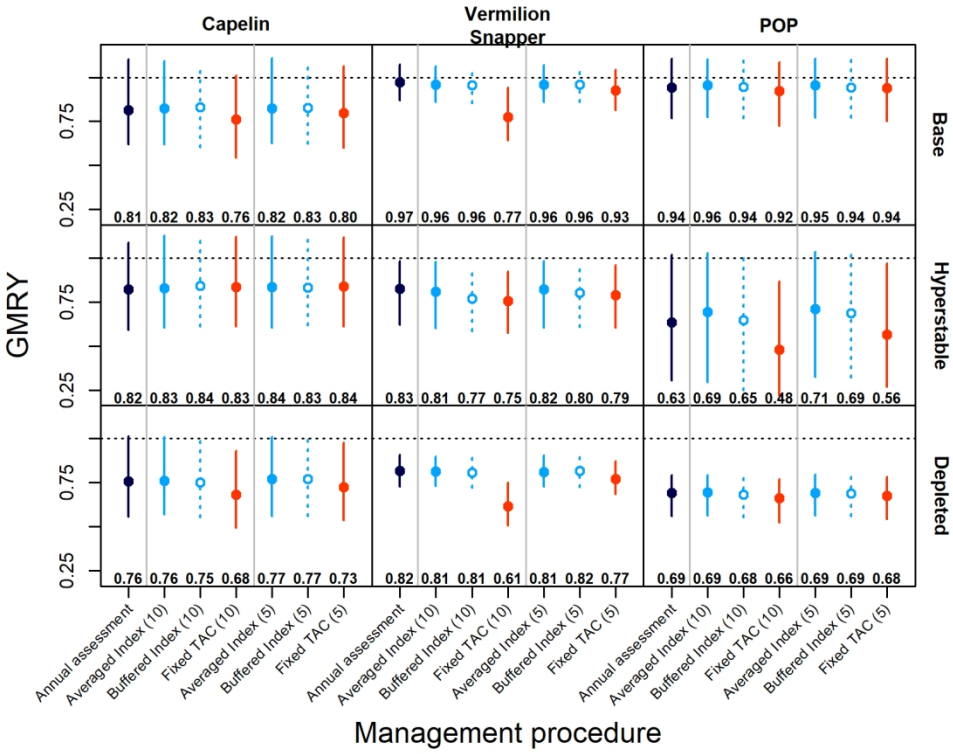

Figure 7. Dot-and-whisker plots of *GMRY* (geometric mean of relative yield) for each species (columns) and scenario (rows). For each MP, dots and numbers indicate the median from 250 simulations, and whiskers span the interquartile range. Dotted, horizontal lines indicate a value of 1.

Supplementary Materials for:

The interim management procedure approach for assessed stocks: responsive management advice and lower assessment frequency

Quang C. Huynh[1], Adrian R. Hordyk[1], Robyn Forrest[2], Clay E. Porch[3], Sean C. Anderson[2], Thomas R. Carruthers[1]

[1] Institute for the Oceans and Fisheries, The University of British Columbia, 2202 Main Mall, Vancouver BC V6T 1Z4

[2] Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Rd, Nanaimo BC V9T 6N7

[3] National Marine Fisheries Service, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, FL, 33149

For Review Only

Table of Contents

A. Illustration of the interim management procedure.....2

B. Description of the operating models.....3

C. Description of the Statistical Catch-at-Age (SCA) assessment model.....4

D. Evaluation on the number of simulations.....7

E. Interim MPs and Fixed TAC MPs with 5-year assessment intervals.....10

A. Illustration of the interim management procedure

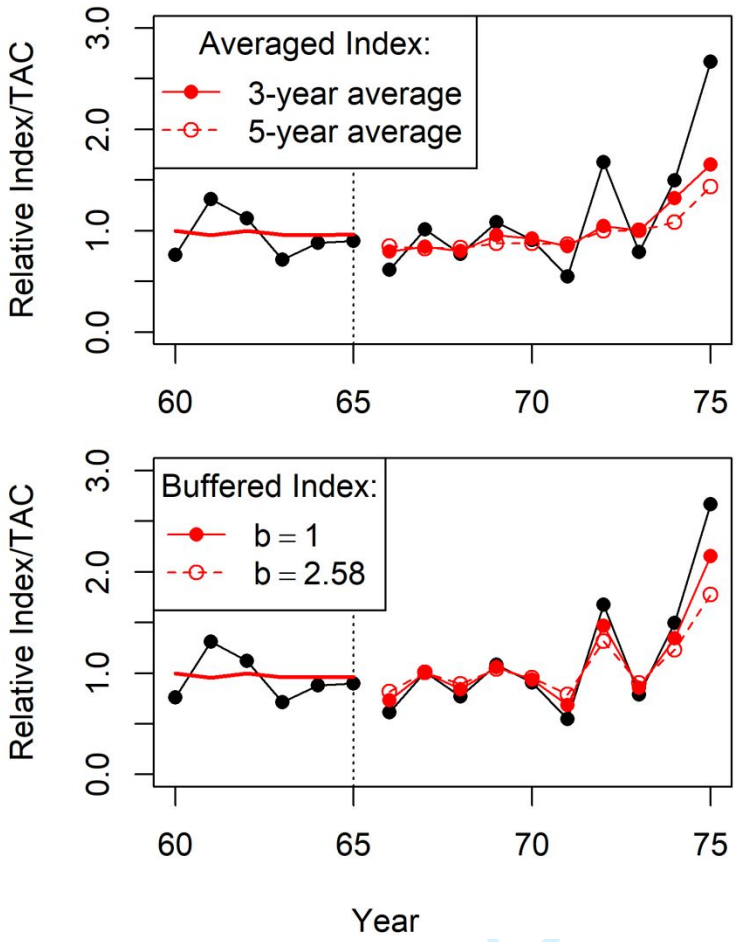

Figure A.1. Demonstration of the interim MPs. Black points and lines indicate index values from a survey. The red lines prior to year 65 are predicted values from an assessment performed in year 65 and a TAC subsequently proposed. The TAC and index are scaled to equal to one in Year 65. After year 65, either future observed index values (black) or moving averages of the index can be used to update the TAC (connected red circles). Top: In the Averaged Index MP, the relative change in TAC is proportional to either a 3-year or 5-year moving average of the index. Bottom: In the Buffered Index MP, the magnitude of the proposed change in TAC is first calculated to be proportional to the observed index but reduced by the buffer in Equation 7 of the main text. Either MP can be tuned so that the relative change in TAC is smaller than that in the index by different magnitudes based on one's risk tolerance. For example, in Year 75, the observed index has tripled since Year 65, but the change in catch advice between those years is smaller than that in the index.

B. Description of the operating models

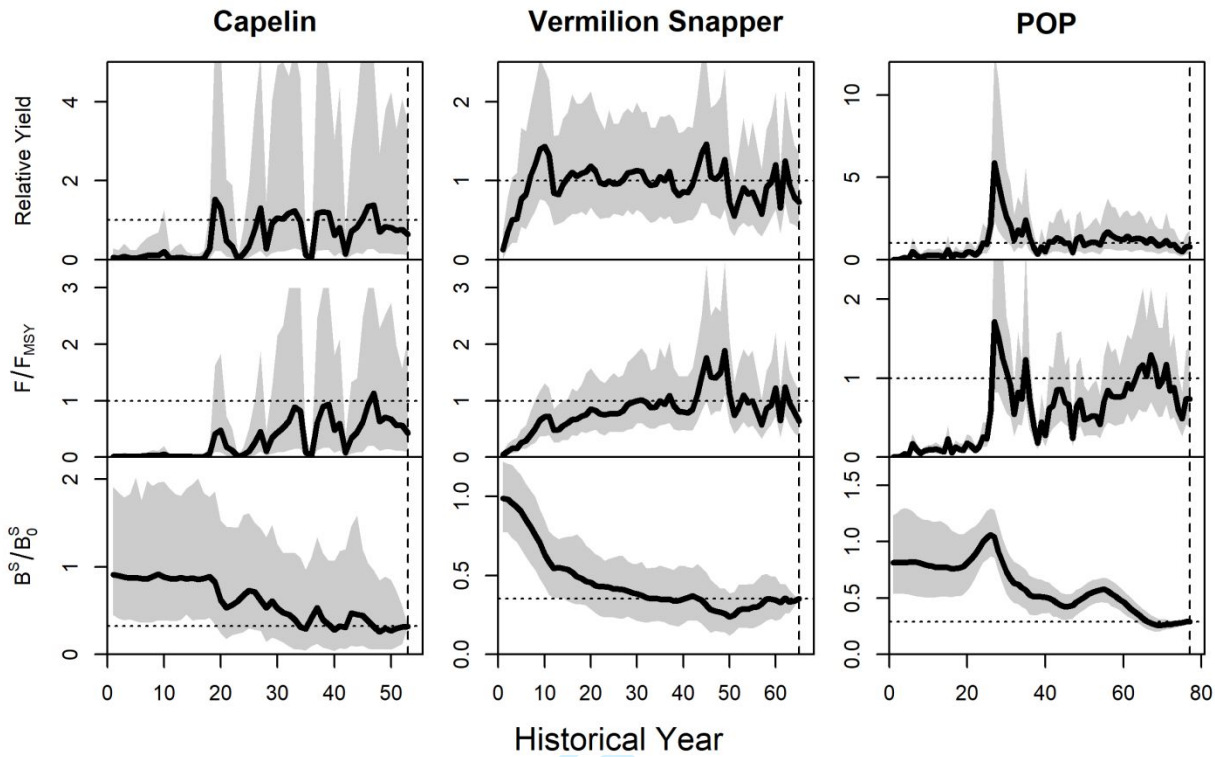

Figure B.1. Historical fishing and stock dynamics (relative yield, relative fishing mortality F/F_{MSY}, and spawning depletion B/B_0) of the three operating models. The relative yield is the ratio of the catch to the reference yield, defined in the Methods section of the main text. Black lines indicate the median trajectory and grey shades cover the 95% range of the respective values. The dashed, vertical line indicates the end of the historical period and the dotted horizontal line indicates B^S_{MSY}/B^S_0 in the bottom row and a value of 1 for all other rows.

For Review Only

C. Description of the Statistical Catch-at-Age (SCA) assessment model

The population abundance $N_{y,a}$ of age a at the beginning of year y is

$$N_{y,a} = \begin{cases} R_y & a = 1 \\ N_{y-1,a-1}(1 - v_{a-1}U_{y-1})\exp(-M_{a-1}) & a = 2,..,A-1 \\ N_{y-1,A-1}(1 - v_{A-1}U_{y-1})\exp(-M_{A-1}) + N_{y-1,A}(1 - v_A U_{y-1})\exp(-M_A) & a = A \end{cases}, \quad \text{(C.1)}$$

where R_y is the recruitment of age 1, v_a is the vulnerability at age, U_y is the harvest rate, and M_a is the instantaneous natural mortality rate, and A is the maximum age (as a plus-group).

The vulnerability, assuming a logistic function, is

$$v_a = \left\{1 + \exp\left[-\ln(19) \times \left(\frac{a - a_{50}}{a_{95} - a_{50}}\right)\right]\right\}^{-1}, \quad \text{(C.2)}$$

where a_{50} and a_{95} are the ages of 50% and 95% vulnerability, respectively.

The vulnerability, assuming a double-Gaussian function, is

$$v_a = \begin{cases} \exp\left[-0.5\left(\frac{a-\mu}{\omega^{asc}}\right)^2\right], & a \le \mu \\ \exp\left[-0.5\left(\frac{a-\mu}{\omega^{des}}\right)^2\right], & a > \mu \end{cases}, \quad \text{(C.3)}$$

where μ is the age of full vulnerability, and ω^{asc} and ω^{des} are parameters that control the width of the ascending and descending limbs (referenced by superscript asc and des, respectively) of the vulnerability function.

From a time series of observed catch C_y (in weight), the harvest rate U_y is

$$U_y = C_y / B_y^V, \quad \text{(C.4)}$$

with the vulnerable biomass B_y^V calculated as

$$B_y^V = \sum_{a=1}^{A} v_a w_a N_{y,a} e^{-0.5M_a}, \quad \text{(C.5)}$$

For Review Only

where w_a is the weight-at-age schedule. Pope's approximation is used so that the harvest rate is conditional on the catch taken at the midpoint of the year.

With a Beverton-Holt stock-recruit function, the recruitment (at age-1) in year y is

$$R_y = \frac{4hR_0B^S_{y-1}}{(1-h)B^S_0 + (5h-1)B^S_{y-1}} \exp\left(\delta_y - 0.5\tau^2\right), \tag{C.6}$$

where R_0 is the unfished recruitment, h is steepness, δ_y are normally-distributed deviates with standard deviation τ, and B^S_0 is the unfished spawning biomass derived from R_0 and other life history parameters. The spawning biomass B^S_y is calculated as

$$B^S_y = \sum_{a=1}^{A} m_a w_a N_{y,a}, \tag{C.7}$$

where m_a is the maturity-at-age schedule.

The model estimates R_0, vulnerability parameters, and recruitment deviates, with steepness fixed to the assumed value from the observation model.

The log-likelihood L^I of the observed index I_y is

$$L^I = -n^I \ln \hat{\sigma} - \frac{1}{2\hat{\sigma}^2} \sum_y \left(\ln\left[I_y\right] - \ln\left[\hat{q}\, \hat{B}^V_y\right]\right)^2, \tag{C.8}$$

where n^I is the number of years with index values, $\hat{q}$ is an estimated coefficient that scales the vulnerable biomass to index values (with the circumflex [^] denoting an estimate) and $\hat{\sigma}$ is the estimated standard deviation of the index.

The log-likelihood L^X of the catch-at-age matrix $X_{y,a}$ is

$$L^X = \sum_y \sum_a O_{y,a} \ln\left[\hat{p}_{y,a}\right], \tag{C.9}$$

where $O_{y,a}$ is the assumed sample size of the age composition, and $\hat{p}_{y,a}$ is the estimated catch-at-age composition, calculated as

$$\hat{p}_{y,a} = \frac{\hat{v}_a \hat{U}_y \hat{N}_{y,a} \exp\left(-0.5M_a\right)}{\sum_a \hat{v}_a \hat{U}_y \hat{N}_{y,a} \exp\left(-0.5M_a\right)}. \tag{C.10}$$

For Review Only

The log-likelihood L^{δ} of annual recruitment deviates $\hat{\delta}_y$, estimated as penalized parameters, is

$$L^{\delta} = -\, n^{\delta} \ln \tau - \frac{1}{2\tau^2} \sum_y \hat{\delta}_y{}^2. \tag{C.11}$$

where n^{δ} is the number of estimated recruitment deviates and τ is the standard deviation of recruitment deviates.

The total log-likelihood Λ of the model is

$$\Lambda = L^I + L^X + L^{\delta}. \tag{C.12}$$

From the fitted model, estimates of terminal-year vulnerable biomass and MSY reference points can be derived for the management procedures.

For Review Only

D. Evaluation on the number of simulations

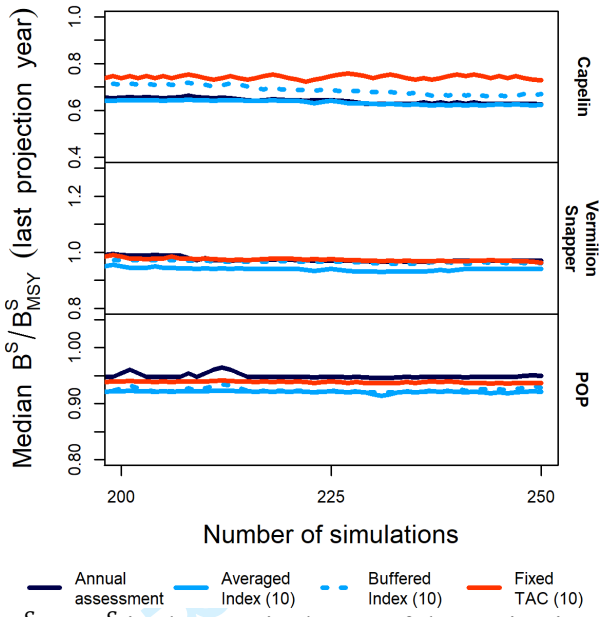

Figure D.1. The median B^S_{MSY}/B^S_0 in the terminal year of the projection period for annual assessment MP and other MPs with a 10-year assessment interval. Stability in the median is used to evaluate whether enough simulations have been run.

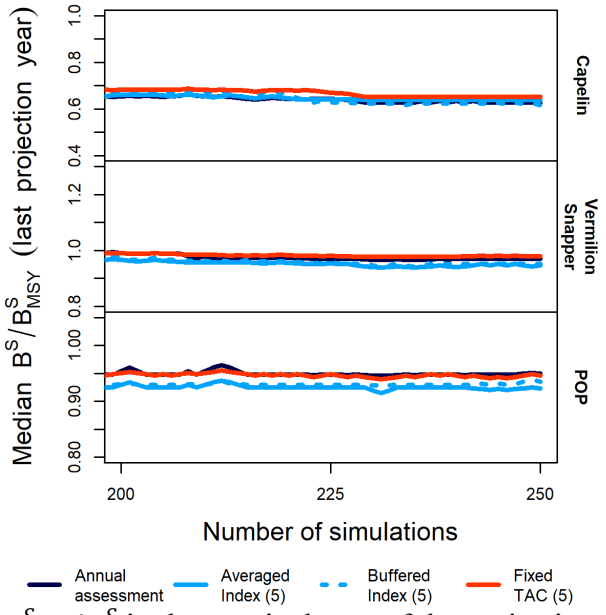

Figure D.2. The median B^S_{MSY}/B^S_0 in the terminal year of the projection period for annual assessment MP and other MPs with a 5-year assessment interval. Stability in the median is used to evaluate whether enough simulations have been run.

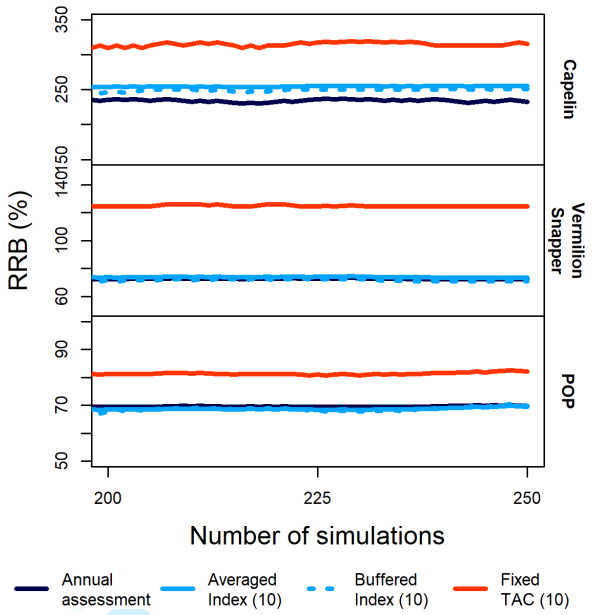

Figure D.3. The median *RRB* (relative range in biomass) during the projection period for annual assessment MP and other MPs with a 10-year assessment interval. Stability in the median is used to evaluate whether enough simulations have been run.

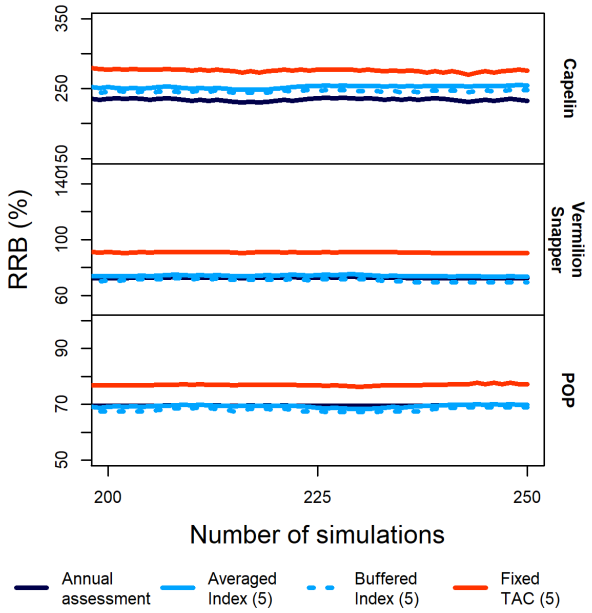

Figure D.4. The median *RRB* (relative range in biomass) during the projection period for annual assessment MP and other MPs with a 5-year assessment interval.

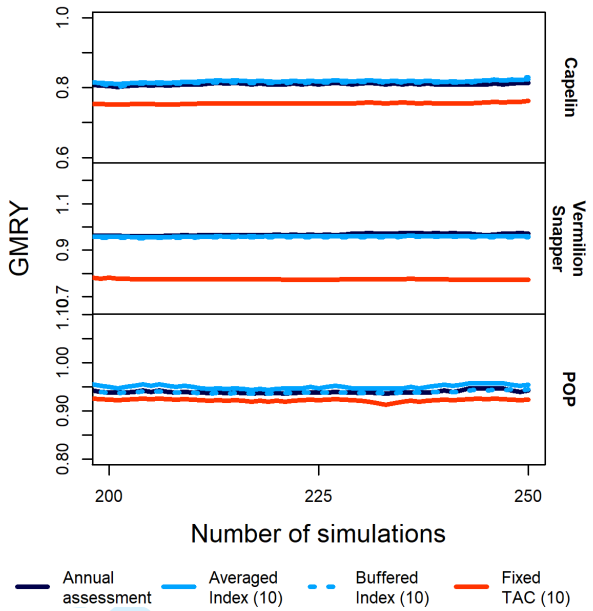

Figure D.5. The median $GMRY$ (geometric mean of the relative yield) during the projection period for annual assessment MP and other MPs with a 10-year assessment interval. Stability in the median is used to evaluate whether enough simulations have been run.

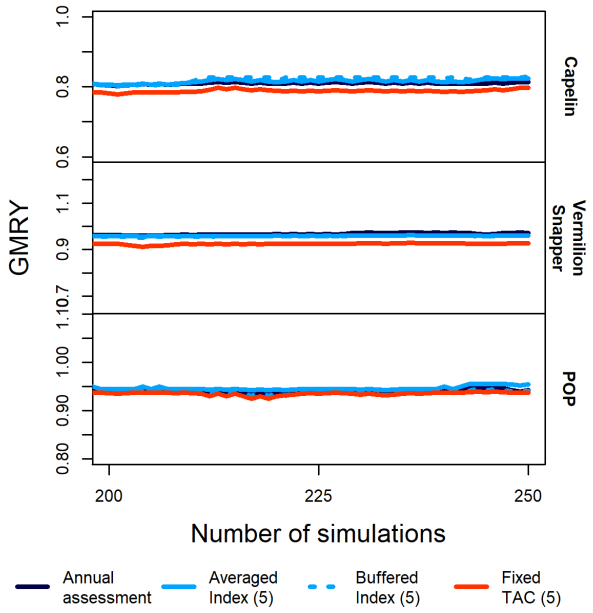

Figure D.6. The median $GMRY$ (geometric mean of the relative yield) during the projection period for annual assessment MP and other MPs with a 10-year assessment interval.

E. Interim MPs and Fixed TAC MPs with 5-year assessment intervals

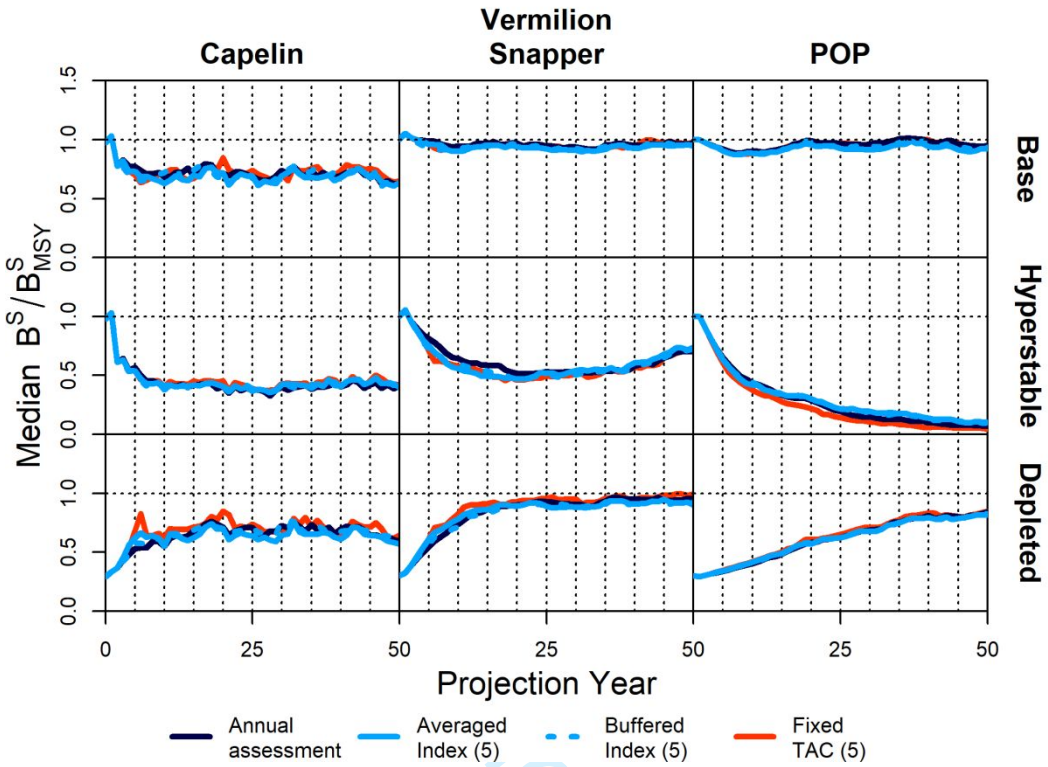

Figure E.1. Annual median B^S/B^S_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of the assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for Annual assessments). Parentheses in the legend indicate the assessment interval.

For Review Only

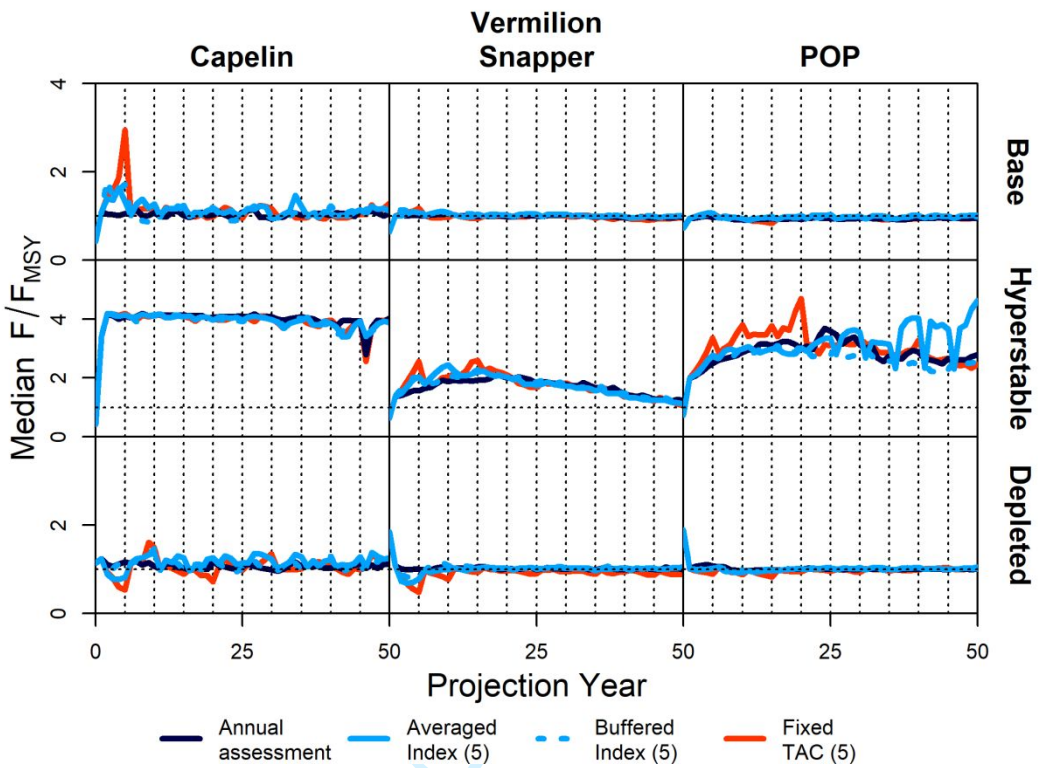

Figure E.2. Annual median F/F_{MSY} from 250 simulations for each species (columns) and scenario (rows). Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for Annual assessments). Parentheses in the legend indicate the assessment interval.

For Review Only

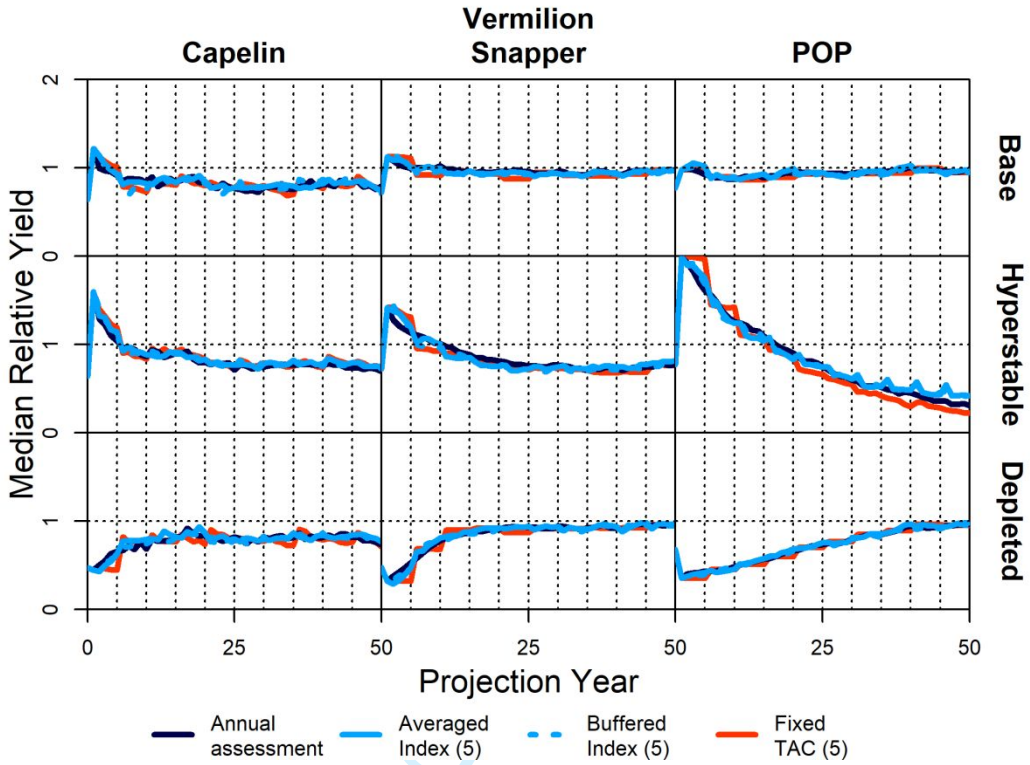

Figure E.3. Annual medians of relative yield from 250 simulations for each species (columns) and scenario (rows). The relative yield is the ratio of the catch to the reference yield. Colored lines correspond to the four MPs. Dotted horizontal lines represent a value of 1. Dotted vertical lines indicate timing of assessment of the Averaged Index, Buffered Index, and Fixed TAC MPs (not shown for Annual assessments). Parentheses in the legend indicate the assessment interval.

For Review Only

Cover Letter to the Editor (and Reviewers)

To whom it may concern:

Submitted is a manuscript titled "The interim management approach for assessed stocks: responsive management advice and lower assessment frequency", which is a re-submission of FaF-19-Apr-OA-096. We thank the editor for the opportunity for re-consideration in Fish and Fisheries. We respond to the Reviewer comments broadly in this cover letter, and in more detail in the Response document.

Upon review of the comments by the two reviewers and a re-reading of the original submission, it was clear that the work was relevant for publication, but (1) there were technical issues with the analyses, and (2) the presentation of results were burdensome to read and interpret.

For (1), we reviewed the software and the operating models for some of the issues brought up by Reviewer #2. For example, that FMSY fishing did not lead to BMSY for the POP species. We also found that FMSY = 9 for Capelin was too high to be credible even for a short-lived species. We found that the best way to proceed was to revise the simulations using a newer version of the software that was released while the original manuscript was in review. The new manuscript addresses these issues as well as others found during software development. For example, we reduced the steepness of Capelin to 0.6 so that FMSY = 1; we didn't find any estimates of steepness for the Osmeridae family (in which capelin belongs) but studies such as Siple, Essington, & Plaganyi (2018, https://doi.org/10.1111/faf.12326) used steepness = 0.6 for forage fish, and we do so here.

For (2), we agree that the presentation of multiple performance metrics and comparisons of management approaches bogged down the paper such that the larger picture was easily missed. Detailed performance metrics may be relevant to one management body, but not necessarily another. For example, NMFS can evaluate whether interim MPs avoid overfishing (F > FMSY) for U.S. stocks to meet National Standard 1 legal requirements, while Canadian law does not explicitly define "overfishing".

In the interest of remaining relevant to the widest audience, we pared back all the performance metrics and used two simple ones: geometric mean of the relative yield (GMRY) and range in relative biomass (RRB) to describe differences between the management procedures. In future evaluation of interim MPs, performance metrics can be developed that are of interest to specific management bodies. We still provide information on trends in biomass, F, and yield that remain relevant to a wide audience, but we do not define and calculate specific performance thresholds.

As a result, the Results and Discussion sections have been re-structured and are much simpler to read. The main take-away messages from this paper are:

For Review Only

1. **Interim management procedures are much more cost-effective in terms of assessment analyst time than annual assessments, and return approximately the same performance in terms of yield and stock dynamics.**
2. **Interim management procedures outperform fixed TACs unless the assessment frequency for the fixed TACs is reduced. Interim procedures should be strongly considered when increasing assessment frequency is not possible.**
3. **There are strategic implications associated with assessment frequency that have not been frequently noted in the literature. Assessment frequency can affect management performance (see Fixed TACs in this paper), and reduction in assessment frequency (, i.e., if interim MPs are adopted) would allow more time to work with un-assessed stocks.**

We believe these conclusions are supported by the analysis and are much more transparent in the re-submission.

Sincerely,

Quang Huynh
Corresponding author

For Review Only

Reviewer: 1

Comments to the Author

Hyunh et al. Fish and Fisheries

The authors explore the use of interim MPs, where the catch is adjusted annually based on trends in an index of abundance (with full assessments every 5-10 years), and compare performance to annual assessments or using a fixed TAC over the interval. This is an interesting and important topic, and I am happy to see it being explored more here. I think the finding that interim MPs may be as good or better than annual assessments is potentially very useful recommendation. However, I have a number of issues that need addressing before it is suitable for publication.

Major issues

First, the authors do a good job in the Intro of talking about the reasons we might need interim MPs, but do not do a good job placing their work in the larger context of studies that have looked into this. As it stands the intro makes it seem like the idea of interim MPs is new and this is the first testing of it. For example Geromont and Butterworth (2015; ICES Journal of Marine Science (2015), 72(1), 262–274. doi:10.1093/icesjms/fsu017) look at the performance of what they call "simple MPs" and compare it to the assessment-based estimates. Although they are not MSE testing interim approaches between assessments, they argue in favor of the simple MPs but with updated assessments. Also, simple MPs are used in S. Africa for some data rich fisheries (see Geromont, H. F., de Oliveira, J. A. A., Johnston, S. J., and Cunningham, C. L. 1999.

Development and application of management procedures for fisheries in southern Africa. ICES Journal of Marine Science, 56: 952–966.) and possibly elsewhere, so it is likely that these simple approaches have been simulation tested at some point.

We thank the reviewer for pointing out these references. The MPs proposed by Geromont and Butterworth and the Namibian hake case study in Geromont et al. 1999 illustrate past examples of survey-based MPs. The difference between those MPs and ours is that ours are periodically updated by an assessment (for parameters such as FMSY and q in the assessment-based MP), while the former are presented as model-free empirical MPs. Nevertheless, the model-free MPs have been found for some life histories to perform as well or better than assessment-based approaches, e.g. Carruthers et al. 2016. This past work is important to discuss in the Introduction, and we do so in the new submission (lines 119-130).

I was surprised to not see a hyper depletion scenario given the hyperstability one. I was also surprised to only see exploration of overfished population but not one with biomass close to B0. I could envision opposite effects to the scenarios explored here, and justification is needed for why they weren't included given that hyperdepletion and having high biomass are both common in fisheries.

We could have run many more scenarios, but at the risk of overwhelming the reader. With the current three scenarios (base, hyperstable, depleted), we saw a variety of trajectories in median biomass, F, yield to understand how interim MPs compare to other approaches.

We present the hyperstable case as "what is the cost of being wrong in our assumption of a proportionate index?" Such scenarios should be evaluated first because it is high-risk (initial stock assessments will be much more optimistic than in reality) and high-consequence (likely to crash the stock). In the reverse case, we can hypothesize what will happen (an overly pessimistic survey could lead to lower catches, and we can also quantify any loss in yield relative to the base scenario).

For our base scenario, we thought setting depletion = BMSY was a high starting point for biomass, relatively speaking. For many management systems, threshold reference points for severe depletion, e.g. "overfished" status, typically occur at less than BMSY, e.g. 50% BMSY. If we started at higher BMSY, the likely behavior of the annual assessments and in the interim MP would have been to gradually bring the stock down to BMSY. Fixed TACs with long assessment intervals will result in some increases and decreases in biomass over time before approaching BMSY.

In the Discussion, we do add text suggesting additional scenarios that (1) may be of interest in new case studies, and (2) were not considered here (lines 493-503).

While it is reasonable to just look at index-based approaches, I think discussion of why other methods in DLMtool were not explored. Methods that do not rely on an index, like catch curve methods, would potentially be more robust to hyperdepletion, but might be more sensitive other factors like recruitment variability.

For Review Only

The focus of the paper is on MPs for assessed stocks, for which there generally will be indices of abundance since most conventional assessment methods require them. We don't want to test every MP in DLMtool in a single paper. In the Discussion, we point out that non-index (catch-based, composition-based) approaches need to be explored if there does not exist a reliable survey. In the context of this paper, comparisons of such methods could illustrate the value (or not) of developing a good survey. We think that evaluation and comparisons of non-index approaches would be highly informative as a separate paper.

The Results section is currently challenging to get through. At first read I liked the idea of having the discrete sections tailored to the 3 questions, but after a couple more reads it became clear that it would be better to combine discussion of all of the MPs (annual, interim, fixed). I like the idea of the pairwise comparison of methods, but there is no need to have separate sections for the performance measures – currently split into the benchmark and pairwise. For example, during the benchmark discussion comparisons are made between the interim and annual MPs, but you could easily also add some quantitative differences here from the pairwise comparisons. Also, there are many very short paragraphs throughout the Results, which makes for a choppy read. Consider combining ideas into more coherent, longer paragraphs. The first few paragraphs read almost like Results paragraphs, and this type of summarizing could be utilized in the Results section.

Upon further reflection, we agree that the benchmark and pairwise performance metrics were overwhelming. We felt the best way to proceed was to choose new and simple metrics that can highlight the differences with fewer figures. We think the biomass range and yield metrics, along with median trajectories, highlight the important differences among the management approaches and different assessment intervals. With fewer results to report, the Results section can still be sub-headed for the three questions without overwhelming the reader.

In the Results, the phrase "performed similarly" is used frequently throughout, but is a bit vague and subjective. For example, one might think a difference in STY or OY of 2-4% between MPs would classify as performing similarly, but the stakeholders might disagree. Consider more specific details here, and for other vague comparison phrases – you've done the pairwise comparisons, why not report the values when feasible?

We found it difficult to report numbers in the Results section given the number of operating model species, scenarios, and MP combinations in the study. For example, if we report values of GMRY for some combinations, the reader might concluded those combinations were more important than others. Overall, we would prefer to focus on broad conclusions. We present all median values of RRB and GMRY in the Figures and the readers can examine specific combinations of interest.

First 4 Figures are about the method and operating model results, and not about the MPs. Figures 1, 3, and 4 could probably be put in an appendix or supplemental – they don't really add much to the paper. Also, Figure 3 is nice to see, but as it stands I look at it and see not much difference since the index isn't that noisy. Consider making an example with a more variable

For Review Only

index to better emphasize the differences. If you cut these Figures, you could either add additional figures, such as something about the assessment frequency (right now all assessment frequency Figures are in the appendices, which seems unusual given that this was one of the 3 main questions being explored here.
), or you could separate out the really dense and hard to interpret pairwise violin plots. I find these Figures overwhelming in the amount of information, but also using the same scale for all makes many of the plots challenging to interpret. Consider a new way to present these Figures.

We understand the concern regarding the number of figures in the paper. We decided to keep Figures 1 (illustration of TACs) and 3 (species growth, selectivity, and maturity) in the paper. We think Figure 1 is important to include in the main text since it supports the Introduction and illustrates the status quo for many fisheries. Fixed TACs represent static management, which is less preferable compared to more adaptive MPs.

Figure 2 (referred to above as Figure 3, demonstration of the interim MP) and Figure 4 have been moved to the Supplement. We experimented with different trends and variability in the index in an attempt to find a good example graphically of the interim MP. High noise results in a relatively flat TAC over time, which we think is less compelling than the current figure, which shows a period of a flat index in which the TAC remains flat as well, and a period of a sharp increase, where the increase in TAC is less than that of the index due to the hedging strategy to accommodate uncertainty. Since the Methods provide a good description of the interim MP, we decided to keep the figure as is (subject to changes suggested by Reviewer #2) and move it to the Supplement.

We have removed the pairwise performance metrics since we agree they were burdensome to interpret. The new RRB and GMRY metrics are simpler to interpret and require fewer figures to present for the full range of species, scenarios, and MPs evaluated in the paper.

I found the performance measures a bit confusing. MSEs usually quantify risk and rewards, where reward is high yield (short or long term) and risk is overfishing, low biomass, becoming collapsed, high yield variability, etc. Here, the risks are actually framed in the positive, so it's the probability of low F, high biomass, etc. Is this the standard output in DLMtool? I just worry that that this will lead to unnecessary confusion. Also, you look at interannual variability in yield and conclude that yield is less variable for the fixed TAC, but another metric could be the maximum change in yield between years. Certainly a fixed TAC would be less variable overall, but there could be huge jumps between assessments, particularly for a longer interval.

As previously mentioned, we have substantially revised the performance metrics for this paper in order to enhance readability. We now use more general metrics to remain relevant to the widest audience. The Discussion section discusses how performance metrics can be tailored to meet the goals and legal requirements of specific management agencies.

Why were only 250 iterations done for each scenario? This seems low given the way I think the uncertainty is included in the operating model. My understanding is that many of parameters are lognormally-distributed, but some of the uncertainties are drawn from a uniform distribution? Correct? This needs to be explained better either in Table 2 or the Appendix. If so, that would

For Review Only

make it seem like 250 is quite low – have you done robustness tests for different number of simulations.

In the Supplement, we present three figures to evaluate whether enough simulation iterations have been run. We evaluate how the median terminal year B/BMSY, median GMRY (over projection period), and median RRB (over projection period) change with increases in the number of simulations. We find that the values of stabilized by 250 simulations.

Why would biomass for vermillion snapper stabilize around 50% BMSY but not POP in the hyper stability runs? Is this the dome-shaped selectivity?

That appears to be the case. Also, the selectivity to the fishery is limited to only mature animals (see Figure 2 in the revised manuscript), which generally makes stocks highly resilient to fishing pressure. We now note these factors in the Discussion. See lines 460-469.

Similarly, why does biomass for snapper stabilize below BMSY for the depleted scenario across MPs?

As noted in the cover letter, we found an issue in the operating model where the mean annual recruitment deviation in the projection period was less than one. This has been resolved in the new manuscript, where the median B/BMSY is close to 1 for the depleted scenario.

For POP it looks like it is plateauing but at a much lower threshold, despite the median F ~ FMSY for both of these stocks for the depleted stock. These sorts of results might make sense with some explanation, but as it stands there is no discussion and I can't tell what in the operating model could be driving this.

As noted in the cover letter, we found an issue in the operating model where the mean annual recruitment deviation in the projection period was less than one. This has been resolved in the new manuscript. Still, B/BMSY is less than one by the end of the projection period due to the slow rebuilding time. The mean generation time of POP is 24 years and 2 times the mean generation time does not appear to be a sufficient time frame to rebuild to BMSY while fishing at FMSY (see Carruthers and Agnew 2016; https://doi.org/10.1016/j.marpol.2016.07.026). We discuss this situation in the Discussion in the paragraph on rebuilding plans and the likely need to set Ftarget < FMSY in interim MPs in order to achieve rebuilding targets (see lines 522-531).

The sample of age composition data seems low (25-50). Especially given an 8 year-lived species but with an M = 0.8 and a 60 year-lived species with so many possible ages to sample from. In both cases I could see very few observations of older fish, which could bias assessment estimates. What was the assessment error across life histories

In the revised simulations, we have increased the sample size to address this concern. We expect high observation error arising from low sample sizes is expected to decrease

For Review Only

precision, but the multinomial sampling process and the selectivity function (dome vs. logistic) from the operating model are properly specified in the assessment model, so these alone shouldn't be expected to generate a bias.

From our initial observations, the bias of assessment models at the end of the historical period was generally close to zero for quantities such as FMSY and terminal biomass. This makes sense because we specified the observation model such that the mean values of observed parameters across simulations should equal the true value in the operating model. Also, the structure of the assessment model, e.g., the selectivity function, were correctly specified, and we mention in the Discussion that this was the case in our study. As such, in order for our results to hold, future assessments would need to be revised in response to changes in the stock dynamics, e.g., a change in selectivity from logistic to dome.

Of course, it is entirely possible that a flawed, i.e., misspecified, assessment model could still perform favorably in meeting management goals. For example, the assumption of logistic selectivity when it is in fact dome-shaped could result in precautionary catch advice, which reduces risk of stock collapse at the expense of foregone yield. A harvest control rule could also limit the risky effects of a flawed assessment model. Such topics are beyond the scope of this paper, but this point highlights the fact that assessment error is not necessarily the end-all-be-all when evaluating management procedures. We note in the Discussion that misspecification of the assessment model could be evaluated in the future (see lines 557-572).

Discussion: it would be really helpful / useful for careful for detailed discussion of when and where these types of approaches may or may not be suitable, particularly beyond the areas explored here. If you could expand this analysis to really expand this work to find out where interim MPs might break down and perform poorly, what would you do?

In the revised version of the paper, we have evaluated interim procedures broadly instead of using detailed performance metrics. The revised Discussion is more compact and covers topics relevant to our study, such as (1) the need to find a good index for interim MPs (based on the hyperstable scenario results), (2) the need to consider re-building plans and lower Ftargets (based on the POP depleted scenario results), (3) the need to evaluate assessment frequency. Additional considerations for the interim MPs (slightly beyond what our paper covered) include the necessity to use MSY proxies, tune the buffers and moving averages, and explore performance when assessment models are mis-specified.

The concept of "performing poorly" also depends on the values and desires of management, which varies depending on location, country, stakeholders, etc. For example, management may desire stability in yield (as an economic objective). Stability in yield can be defined in several ways, e.g. average inter-annual variability versus the frequency of large step-wise changes, and is left as a future exercise.

Minor issues:

Line 96: add "biomass" after vulnerable. **Done**

For Review Only

Line 101: Clarify that you mean Canadian Pacific groundfish stocks. **We've added the DFO (2019) citation, which states the stocks managed by Pacific DFO groundfish management plan.**

Paragraph in Results (starting on line 434) has a bit too much detail. Make sure that the Figure caption is sufficient for the reader to figure out what is going on. **The pairwise performance metrics have been removed, so this paragraph has been deleted.**

Sentence starting line 545 is confusing and needs reworking. **The benchmark performance metrics have been removed, so this sentence has been deleted.**

Line 573 tradeoffs between… **This sentence has been deleted.**

Line 573 What is F creep? **In the previous draft, we were trying to explain why there were overfishing years without there being many depleted years. If F is slowing increasing (creeping) above FMSY (overfishing), that still may not lead to depleted stocks (where B < 0.5 BMSY). We are no longer concerned with reference points in comparing the management procedures, so this sentence has been deleted.**

Table 2 says there are four stocks **This sentence has been fixed.**

Table 2 also say parameters "are in lognormally distributed" **This sentence has been deleted.**

Figure 8. Remind the reader what the PMs are in the caption. **Performance metrics (PMs) have been removed.**

Geromont and Butterworth 2015 cited but not in refs list **This citation has been added.**

For Review Only

Reviewer: 2

Comments to the Author
Review of: "The interim management procedure approach for assessed stocks: responsive management advice and lower assessment frequency"
Authors: Hyunh, Hordyk, Forrest, Porch, Anderson and Carruthers.
Reviewer: José De Oliveira

The manuscript evaluates two interim management procedures (MPs) (using a running average index and a buffered index, respectively, between assessments) against doing annual assessments or fixing the TAC between assessments. The authors demonstrate similar performance between annual assessments and the two interim MPs, but with potentially substantial cost advantages for the latter, while more frequent assessment would be needed when coupled with fixed TACs in order to match the performance of the two interim MPs. The work is therefore of high interest to managers and should be considered for publication with moderate to minor modifications. I have some reservations about the results for pacific ocean perch (POP), and suggest that the

Appendices could be moved to supplementary materials (apart from one figure). Comments are listed below, categorised by major, moderate and minor comments. Separate comments are provided for Appendix A.

1. Major comments
1.1. L157-70. Perhaps it is different in North America, but usually, when the assessment is actually conducted (y), it is the year before the year when the TAC is needed (y+1), and the year after catch data are available for input into the assessment (y-1), so if a biomass is estimated at the beginning of y+1, then usually an assumption is needed about fishing mortality during year y. If this is the case, then the description of Bref does not make this clear, and taking such lags into account is extremely important (see e.g. Kell et al. 2005). The authors should clarify whether they have done this, because ignoring the lags will impact the performance of the MPs, and also comparisons amongst them.

The study tries to be independent of management systems, although we choose examples among U.S., Canadian, and ICES stocks. Currently, the Methods section is correct in describing the simulation (Bref was the biomass in year y+1, the year for which the TAC is calculated). We agree that assessment lags can be quite important, especially for a short-lived stock, and we discuss the need to incorporate lags from the management system for future studies.

1.2. Figures 4 and 5 (also L393, L543-4). I have concerns about some aspects of Figure 4, but especially the B/Bmsy trends for POP in Figure 5. In Figure 4, it is strange that for Capelin, fishing up to 50% above FMSY for a large proportion of recent years still results in B/BMSY~1. Similar for Vermilion Snapper (VS); in fact, for VS, F's mostly well above FMSY (from year 40) lead to an increase in stock size (from year 50) to BMSY. This seems a strange result, and something the authors don't seem to be particularly bothered with. Can they explain what is going on?

In the historical period, recruitment deviates are re-sampled when the stock cannot be depleted to the specified depletion level (B=BMSY). We don't see an issue with Vermilion Snapper historical dynamics. Upon closer inspection, the figure shows that F>FMSY in years 40-50 depletes the stock below BMSY. In years 50-63, F>FMSY for a few years, but mostly F<FMSY, which in total allows the stock to recover to BMSY.

Of greater concern to me though is what is shown for POP in Figure 5: the Ftarget in all MPs is FMSY (L289), yet the stock seems to drop below BMSY for all MPs in the Base scenario, and only increases to well below BMSY for the depleted scenario (seemingly reaching plateau). It looks like FMSY and BMSY are not compatible for this stock. The authors should explain this as it may indicate something wrong with the POP results.

We found an issue for POP where the mean annual recruitment deviates in the projection period were less than one which would result in B < BMSY when F = FMSY. Simulations were re-run and this issue has been fixed in the revised manuscript.

2. Moderate comments

For Review Only

2.1. L222-3. I think it is a stretch to continue the analogy, since this is not clearly demonstrated or theoretically proven, so I suggest removing the sentence beginning "To continue the analogy..." **Done.**
2.2. L231-2. "After the historical dynamics of the stock are generated..." - it is not clear from the text or Figure 4 what the historical dynamics are based on? See also 2.3 and 2.24 below. **This sentence should address the comment: "In the historical period, the fishing mortality was taken from the assessment and re-scaled by a multiplier that produced the specified depletion."**
2.3. L250-3. "The historical fishing effort was scaled..." Again, it is not clear what the history was based on. Is it made up or based on some assessments? Was the trend from an assessment, and the entire trajectory scaled to deplete the stock to Bmsy? Please make this clear. **See 2.2.**
2.4. L254-9. It would be much easier to read if these were turned into bullet points. **Done, we have made the sentence into a numbered list.**
2.5. L261-3. For a clearer linkage to Table 2, and clearer description, please include the distribution of the deviates. For example, for equation (8a), something like eps~N(0;sig^2) with appropriate superscripts and subscripts. Same applies to equation (8b), but this is needed for both the gamma and the epsilon. The text describes the gamma as normally distributed, but does not make clear that it has a mean of zero (unless this is considered a deviation parameters, as described in L269?). The additional equations will make clear the distributions, and the text could perhaps be shortened as a result. The variance parameters of the deviate distributions could then be referenced in Table 2 to make a clear link between the text description and Table 2 (see 2.21 below). **Table 2 is now more comprehensive in describing the observation model. The mathematical notation should be precise enough to identify the distributions of stochastic deviates and whether they vary only by simulation or by year and simulation.**
2.6. L274. The second gamma should be defined (e.g. with a distribution equation as for the deviates described in 2.5 above). The variance parameter associated with this could then be included in Table 2 (currently missing). **Done, see 2.5.**
2.7. L277. The details of the catch-at-age assessment model (Appendix A) are not needed, and I suggest moving this to supplementary materials. **Done.**
2.8. L289-92. The description here is incomplete because you have not said what Bref is. Is this B_y+1? **The text has been identify Bref = B_y+1.**
2.9. L335, L352. What was the basis for selecting 0.9 (GMRY) and 0.25 (AAVY) for calculating probabilities? A short sentence needed in each case. **Performance metrics were removed, usually they will be defined by management. In the interest of reaching a broad audience, we have used more general metrics in the revised manuscript.**
2.10. L337-40. The description of RefY is obscure and confusing. How exactly was this calculated? Was it done on a simulation-by-simulation basis by looking for the constant F that maximised yield in the projection period, and then selecting that yield as the reference? Please make this clearer. **The revised description should now be clearer: "the reference yield is the mean annual catch that maximizes the total catch in the projection period with a constant fishing mortality. The reference yield varied among simulations due to stochasticity in recruitment deviations" RefY approaches MSY as the projection period increases.**
2.11. L360-6. It is not entirely clear from (1) and (2) over which period of years these metrics were considered. Was it the entire projection period of 50 years? Please make this clear. **The description in the revised manuscript is clear that all metrics are calculated over all projection years.**

For Review Only

2.12. L368. "A positive value indicates..." - it is strange to talk about a "positive" value, implying there are negatives. This is surely not the case for ratios. Do you mean "a value larger than 1"? **That is what was meant, but we no longer use this metric.**
2.13. L379-80. "Supporting figures...Appendix B. I suggest that Appendix B be moved to supplementary materials and only Figure B1 is kept. The other figures don't really add much to the paper, and a simple description of the results of these is sufficient. See also 2.16 below. **These figures are no longer used, and we use a supplementary materials for extra figures.**
2.14. L434-43. "For example,...fewer overfishing years." The captions to the figures already contain a lot of this information, so I suggest the text be shortened substantially (see e.g. L452-4 for a brief description for Figure 10). **These metrics are no longer used, which substantially reduces the text and makes the Results section clearer.**
2.15. L494, L571-3. What do the authors mean by "may generate creep in fishing mortality", "due to creep over time", and "F creep"? This is not clear and can be confused with "technological creep" in e.g. in the CPUE standardisation context. Please explain what you mean. Is it because the assessment parameters change between assessments, but the interim MPs will keep them fixed until the next assessment? ? **In the previous draft, we were trying to explain why there were overfishing years without there being many depleted years. If F is slowing increasing (creeping) above FMSY (overfishing), that still may not lead to depleted stocks (where B < 0.5 BMSY). We are no longer concerned with reference points in comparing the management procedures, so this sentence has been deleted.**
2.16. L505-20. I suggest relegating Figures B2 and B3 to supplementary materials and integrating Figure B1 into the main paper. If there are too many figures, then perhaps Figures 1 and 3 can also be relegated to the supplementary materials. **These figures are no longer used, and we use a supplementary materials for extra figures.**
2.17. L536-7. "...compared to the other two species". Do you mean compared to POP? Looking at Figure 8, it appears that depletion risk is reduced (i.e. higher PB50) for VS compared to POP for the Hyperstable scenario. For Capelin, there is little difference between the three scenarios when considering PB50. **These metrics are no longer used.**
2.18. L565-6. "This resulted in a slightly high risk of stock depletion in some cases..." - what is this referring to with "slightly high"? Compared to what? **These metrics are no longer used.**
2.19. Table 1. Include references for the parameters listed (e.g. in the title row of the table). Also, what is the range given in the final row? Is this also to be sampled from a uniform distribution from the range (as in Table 2)? Please clarify.
Only the recruitment SD varies among simulations, and this is noted in the Table caption.
Also, for such a slow growing, long-lived species, I was surprised to see such high recruitment variability parameters for POP (higher than for Capelin!). Is this correct? **Yes, some large year-classes were observed in the POP stock historically, which would explain the relatively large sigmaR. Nevertheless, the dynamics of POP (e.g., slow rebuilding time) in the projection period are what we expected from a long-lived species.**
2.20. Table 2. L851-3. "Among simulations, the bias for the observed catch or life history parameters are lognormally distributed with mean 1 and standard deviation set by the CV provided in the table below." In the case of the bias in catch, this contradicts the main text, which explains (L267-8) "...and gamma_i are independent, normally distributed deviates among simulations that parameterize bias in observed values." Please check which is correct and be consistent. **Table 2 has been re-formatted and is much clearer with the mathematical notation.**

For Review Only

2.21. Table 2. I think it would be very helpful to the reader if the parameters/variables in the equations were shown here to provide a clear link between the equations and the values used. For example, instead of saying "exponent", include the beta_i parameter, etc. If the suggestion in 2.5 above is followed, then the appropriate symbols could also be included for the CVs. **Table 2 has been re-formatted and is much clearer with the mathematical notation, with citation to equation numbers.**
Furthermore, the "CV of the bias in the observed steepness among simulations" (linked to equation (9)) should also be included. **Table 2 has been re-formatted and is much clearer with the mathematical notation with citation to equation numbers.**
2.22. Table 3. Explain how the range for the hyperstable scenario is used. Also sampled from a uniform distribution?
Table 2 is updated to show the distribution of beta in the hyperstable scenario.
2.23. Figure 2. the middle plot seems to be superfluous and could be deleted. Instead the bottom plot could have the same form as the top plot with the average indices replaced by the buffered index. Furthermore, the final two sentences of the caption should not be in the caption - they could be moved to the main text (e.g. around L214).
This Figure has been simplified as suggested.
2.24. Figure 4. Please explain what the relative yield is relative to? Also explain the source of the historical period (from assessments?).
The relative yield is relative to the reference yield. While the reference yield pertains only to future yields in the projection period, it provides a consistent comparison of yield between the historical and projection periods.

3. Minor comments
3.1. L29. "COMPARED TO fixed TACs from assessments with a 10-year interval,..." **Abstract has been revised & sentence re-phrased.**
3.2. L96. "...project vulnerable BIOMASS from the most recent..." **Fixed.**
3.3. L99. Use "forecast" or "predict" instead of "verify"? **We've used "forecast" as suggested.**
3.4. L159. Remove strange apostrophe after "otherwise" in equation (1) 3.5. L161. "In the interim, the TAC IS held constant." This is the case for the way equation (1) is currently written.
The apostrophe is a comma; the equation is part of a longer sentence. L161 is changed as suggested.
3.6. L166. Replace "reference biomass" with "Bref" for clearer linkage to equation (1) - Bref has already been defined higher up, so okay to use the acronym. **Fixed.**
3.7. L172. "...vulnerable biomass in year y, WHICH IS MODELLED AS". Including this will then mean that circumflexes can be added to all the variables in equation (2), to be consistent throughout and with later use of q in equation (3). **See next point.**
3.8. L173-4. Add circumflexes to variables in equation (2) and to q in the next line. **In the interim procedure, we use the observed index values. Equation 2 describes the assumption of the index to be proportional to the true vulnerable biomass via (the true) q. Thus no circumflexes should be added here.**
3.9. L206. Geromont & Butterworth 2015 is not in your reference list. **Added.**
3.10. L211. "...or b, WHEN BOTH ARE POSITIVE, penalize..." Of course, if either is set to zero then it does not matter what you do! **Sigma is always positive. We've described b as non-negative, which should address this comment.**

For Review Only

3.11. L229. There is no "Carruthers and Hordyk 2018a" in the reference list - drop the "a"? **Done.**
3.12. L244. Should "DFO, 2017" be "DFO, 2017b"? **Citation confirmed (DFO, 2017).**
3.13. L265. No need for "accent-less", so suggest deleting this. **Done.**
3.14. L269. Replace "Deviation parameters" with "Deviates" to link to earlier descriptions? **The sentence was removed; the mathematical notation should now be clear in the text and Table 2 to describe the distributions of the parameters in Equations 8 and 9.**
3.15. L284-5. "...calculated from the variance of the index residuals in the MOST RECENT assessment." - this makes it clear that the value is taken from the most recent assessment and kept at this value until the next assessment - is this correct? **Yes, fixed.**
3.16. L290. "For THE Averaged Index,..." **Fixed.**
3.17. L296. "...as described in Tables 1 and 2." - both tables provide a description. **Sentences revised. Table 3 is cited to describe the differences between the scenarios. Table 3 also cites descriptions in Figure 2, and Tables 1-2, as needed.**
3.18. L320-1. "...(i.e. the proportion of years in projection years 26-50 AND all simulations, I.E. PROPORTION OF 25 YEARS X 250 SIMULATIONS) that B>=0.5BMSY." I think this makes it clearer. **This performance metric has been removed.**
3.19. L325. Along the same lines as 3.18 above: "The proportion of years OVER the entire projection period AND ALL SIMULATIONS (I.E. PROPORTION OF 50 YEARS X 250 SIMULATIONS) in which F<=FMSY". **This performance metric has been removed.**
3.20. L359. "...metrics were used, CALCULATED FOR EACH SIMULATION i:" - I think it is better to say this upfront as it helps interpreting the metrics. **This performance metric has been removed.**
3.21. L362. "A positive value indicates that the stock IS depleted LESS often under the MP j." - I think this is easier to follow! **The performance metric has been removed.**
3.22. L372. I suggest omitting the first part of the sentence to avoid repetition (see 3.20 above), and simply saying "The median (over simulations) of the pairwise metrics is used to..." **This performance metric has been removed.**
3.23. L378. "...trend OF biomass, fishing mortality,..." **Fixed.**
3.24. L385. The word "below" is not needed. **Fixed.**
3.25. L391. "...(Figure 5-7, first columns IN FIRST ROWS)." - since you are referring to Capelin. **Results have been re-organized.**
3.26. L409-10. "...was very likely to BE AT OR ABOVE 50% BMSY IN THE FINAL 25 YEARS (based on..." **This performance metric has been removed.**
3.27. L413. "...slightly smaller than STY IN MOST CASES, indicating..." - e.g. does not apply to Capelin in some cases. **This performance metric has been removed.**
3.28. L416. Should the ">50% BMSY" at the start not be ">=50% BMSY"? **This performance metric has been removed.**
3.29. L417. Replace "Figure 7" with "Figure 8"? **This figure has been removed.**
3.30. L418. Could reference Figure 7 here as well to illustrate the optimistic yields at the start of the projection period. **This section has been removed.**
3.31. L419. "..to those in the Base Scenario (EXCEPT FOR POP),..."? - the top right plot does indicate POP to have a lower PB50 for the depleted scenarios, perhaps due to its slower recovery? **This performance metric has been removed.**
3.32. L421. "OY was GENERALLY lowest for POP..." - e.g. it was not the case for the fixed TAC. **This performance metric has been removed.**

For Review Only

3.33. L425-6. "However, for Capelin...and OY." - this is not entirely true for the average index in the depleted scenario... **This performance metric has been removed.**
3.34. L435. Replace "compare" with "compares". **This section has been removed.**
3.35. L449. "...were observed between THEM with..." **This section has been removed.**
3.36. L451. "...Annual Assessment MP FOR THESE TWO SPECIES, but..." **This section has been removed.**
3.37. L477. "...biomass to higher stock levels THAN AT THE START (Figure 5..." **This section has been removed.**
3.38. L480. Remove the word "while". **This section has been removed.**
3.39. L486. Remove the words "first column," - it applies for all stocks. **This section has been removed.**
3.40. L509. "..., but was ALMOST unchanged for POP." - it wasn't completely unchanged. **These performance metrics has been removed.**
3.41. L545-6. "...under the Average Index MP, THIS IS due to the differences..." - otherwise the sentence does not make sense. **These performance metrics has been removed.**
3.42. L564. Replace "avoids" with "avoid". **This section has been revised.**
3.43. L566. "...in our study, e.g. FOR Vermilion Snapper." **This section has been revised.**
3.44. L576. "...can be quite high for both INTERIM MPs, although..." **This section has been revised.**
3.45. L577-8. "... were to show a decline OVER TIME BETWEEN ASSESSMENTS." - correct? **This section has been revised. In the Discussion, we get straight to the point that hyperstable indices pose a problem regardless of the MP used.**
3.46. L581-2. "..., the Fixed TAC MP LED TO reduced fishing mortality..." **Similar to the previous point, this sentence was a very detailed observation that neglects the overall idea that we should have more frequent TAC updates in rebuilding fisheries. This is possible with interim MP (regardless of assessment frequency) or fixed TACs with more frequent assessments. The point is that the former is more cost-effective in terms of assessment time.**
3.47. L624. "SETTING THE ISSUE OF benchmark assessments ASIDE, this study..." **TWe noticed there was quite a bit of repetition in the Discussion. Upon revision, we no longer need this paragraph and it was deleted.**
3.48. L653-4. Reference needed for the "40-10 control rule". **Deroba & Bence (2008) is now used to cite ramped harvest control rules such as the 40-10.**
3.49. L670. Remove the words "the trend in" - are they needed? **That sentence was superfluous and removed.**
3.50. L671. "...can be developed IN ORDER to evaluate potential Ftarget values..." - "among" not needed. **Fixed as suggested.**
3.51. L682. Add hyphen to "10-year". **That sentence was superfluous and removed.**
3.52. L845, L849. Replace "four" with "three". **Done.**
3.53. L852. "...parameters are lognormally..." - remove "in". **Done.**
3.54. Figure 1 caption. "DFO, 2017", is this a or b? Reference list only lists these two. **The fishery management plan has been updated so the appropriate citation is DFO (2019).**
3.55. Figure 4. "and the dotted," is repeated twice in the caption. Remove one of them. Also, subscript the reference points, as in the plot itself. Finally, the dotted horizontal line indicates BMSY/B0 for the bottom plots (don't used the work depletion, as there is another context for this in the paper related to one of the scenarios. I would also suggest you indicate that it is set at 1 for

For Review Only

the other plots (e.g. this is not clear for the top-right plot). **Fixed as suggested (and moved to supplement).**
3.56. Figures 5-7. Use proper subscripting for the reference points (as in the plot itself) and not e.g. B_MSY. Also, I suggest using colours that, when printed in back and white, will still differentiate sufficiently in greyscale to not have to go to the original coloured plots. **Subscripting has been re-checked for consistency. With trajectory plots where the lines are close together, it's unfortunately difficult to find a scheme which works well in black-and-white. Greyscale did not improve readiblity on paper or on screen. Here, we used the gplots::rich.colors function in order to obtain a color scheme that shows contrast on screen, but we can add text in the figure caption to refer to the online version for color.**
3.57. Figure 10. Caption, final sentence. "In all panels, positive values indicate..." - this should be "In all panels, values greater than 1 indicate..." because you are comparing ratios. [Check Figure B3 as well for the same problem.] **Figures have been removed.**

4. Appendix A (suggest relegating to supplementary materials).
4.1. L929. Remove dash: "age 1" **Done.**
4.2. L932, 954. Is it better to use ln (natural logs) instead of log, where the base of the log is not given? **We now use 'ln' to remove ambiguity on the base of the logarithm.**
4.3. L932. Place log 19 in parentheses: "(log 19)" - then there is no confusion about whether the ratio that follows should be included in the log (it should not be). **Done.**
4.4. L933. Start a new paragraph after "respectively", with "THE VULNERABILITY, assuming a double-Gaussian function, is" - this is then consistent with the preceding text. **Done.**
4.5. L941-2, 945, 947-8. Avoid the use of two symbols (VB and SB) when referring to a single quantity. Rather use superscripts and subscripts (e.g. B superscript V, and B superscript S). Blame Doug Butterworth for this... **This has been fixed throughout the manuscript as suggested. I too will blame Doug the next time I see him.**
4.6. L942, 949. Are stock and catch weights the same (same symbol used)? Please clarify. **Yes, by using the same symbols, we believe it's clear that the stock and catch weights are identical.**
4.7. Equations A.8, A.9 and A.11. These are not likelihoods, as indicated in L952-3, but distributions. Please give the likelihoods instead (as intended). **We now have the likelihoods.**
4.8. Equation A.8. You seem to assume that surveys have the same vulnerability as the fishery (compare with equation A.4) - is this a reasonable assumption? **We use Pope's approximation so that both the catches and index are taken and observed, respectively, at the midpoint of the year. We think this is a reasonable vulnerability assumption, e.g., fishing vessels are sometimes contracted to do surveys or fishery CPUE is used as an index. For simplicity, we adopted this assumption. However, future case studies need to evaluate how well interim MPs perform when fishery and survey selectivity are substantially different (see lines 566-569).**
4.9. L962. Is this needed, given that summing over a in equation A.10 will lead to 1 anyway? **The summation isn't needed (implied by A.10) and removed.**

References
Kell, L. T., Pilling, G. M., Kirkwood, G. P., Pastoors, M., Mesnil, B., Korsbrekke, K., Abaunza, P., Aps, R., Biseau, A., Kunzlik, P., Needle, C., Roel, B. A., and Ulrich-Rescan, C. 2005. An

For Review Only

evaluation of the implicit management procedure used for some ICES roundfish stocks. ICES Journal of Marine Science, 62: 750-759.

For Review Only