Updates and Thoughts on Panel Recommendations:
EXTERNAL PEER REVIEW OF ECOSYSTEM-BASED FISHERY MANAGEMENT STRATEGY

Summary

In the following we provide updates on the status of work conducted to date to address issues raised during the External Peer Review of the Ecosystem-Based Fishery Management Strategy held in Woods Hole May 31-June 3, 2018 and offer thoughts on next steps. We would like to express our gratitude to the review panel, Dr. Keith Brander, Dr. Villy Christensen, and Dr. Daniel Howell for their many constructive and insightful comments during the review and in their individual reports and to Dr. Lisa Kerr, chair of the review panel who efficiently ran the meeting and expertly coordinated the preparation of the summary report of the panel.

A major focus of our activities since the review has been to implement efficiency changes in the coding of the existing multispecies operating model Hydra. These changes are necessary to facilitate more intensive simulation studies to test the performance characteristics of a proposed multispecies management procedure for Georges Bank. A full Management Strategy Evaluation is planned for 2019. We have also continued development and parameterization of two ecosystem models, Ecopath with Ecosim and Atlantis to permit evaluation of broader ecosystem-level issues and concerns.

During the review simulation results were presented as a general proof of concept for the prototype management procedure using hypothetical management objectives. The full MSE will elicit specific guidance on management objectives and concerns from managers and stakeholders. This advice will be used tailor the final selection of appropriate models structures to address these issues and the scenarios to be tested

The multispecies-multifleet simulation model Hydra has been revamped to substantially reduce run times for each simulation of the management procedure from 12-14 hrs to approximately 40 minutes, greatly increasing the number of simulations that can be conducted in a management strategy evaluation. Major enhancements have been effected by increasing the efficiency of how intermediate output is handled in the course of the simulations. The program is also configured to allow greater efficiency through parallel processing for scenario testing.

In response to specific structural recommendations made by the review panel, we have implemented the following changes in Hydra to date:

1) The fleet structure has been expanded to allow a small-mesh otter trawl fishery. The fixed gear fleet has been partitioned to permit separate treatment of gillnets and longlines. The total number of fleets considered is now five.
2) The harvest control rule now allows a continuous ramp function as an option to augment the step function initially employed
3) Species and guild status indices are now calculated to represent the number of species that collapse in a given time period rather than the number of times each species falls below a threshold.
4) The stock recruitment sampling routine has been modified in several ways:
    a) Additional parametric forms for recruitment have been included (Beverton Holt, Ricker, and Shepherd)
    b) Sampling takes place around the optimal fits for each species (MLE). For asymptotic parametric recruitment functions, the sampled slopes are constrained to have the same asymptote. For overcompensatory parametric recruitment functions, the sampled curves are constrained to pass through the same maximum recruitment level. For the segmented linear function the sampled slope is constrained to be within a specified bounds (either side of the fitted slope)
    c) The amount of "other food" is also randomly sampled

In the remainder of this document, we address specific comments and concerns raised by the review panel in their summary report and additional changes to be implemented. Below, the elements of the Review Panel Summary Report dealing with each of the terms of reference are provided and, where appropriate, thoughts and comments on progress to date and next steps are provided in italics. Reference is made to the 'Ecosystem-Based Fishery Management Strategy Georges Bank Prototype Study Summary Document' (hereafter Summary Document).

EVALUATION OF TERMS OF REFERENCE

ToR 1: Evaluate the approach used to identify Ecological Production Units on the Northeast Shelf of the United States and the strengths and weaknesses of using these Ecological Production Units as the spatial footprint for Ecosystem Based Fisheries Management in the region.

The Panel reviewed the written materials and presentations on the methods used to identify ecological production units on the Northeast Shelf of the United States for application in EBFM in the region. The aim was to identify geographically-defined ecological units based on: 1) physical oceanography, 2) hydrographic variables, and 3) biological variables (including primary production, but not upper trophic levels). Multivariate analysis was applied to reduce dimensionality of the data (principal components analysis) and identify clusters of data (disjoint cluster analysis) that represent major ecological production units. This process led to the identification of four ecological production units: 1) Mid-Atlantic Bight, 2) Georges Bank, 3) Western-Central Gulf of Maine, and 4) Scotian Shelf-Eastern Gulf of Maine. These were put forth as the spatial management units that would underpin the EBFM approach in the region. The Panel identified strengths and concerns with the approach and made recommendations for consideration in future work.

Strengths

- Scientifically rigorous method: The Panel recognized that the approach was rigorous and allowed for objective identification of ecological production units (i.e., the data defined the geographic structure).
- Comparable to previous findings: The results of the analysis aligned well with previous approaches to define ecosystem management units using alternative methods (e.g., Clark and Brown 1977, Higgens et al. 1985). This provides support for the ecological production units.

Concerns

- Dynamics of boundaries: One of the concerns of the Panel was that that the boundaries of ecological production units are dynamic and will need to be revisited and updated at some interval. The EBFM technical team should consider an approach for dealing with this concern.

  We agree entirely with this statement. In the Summary Document we suggest a time frame for updates and would welcome comment: 'We further anticipate that the production unit boundaries defined here will be subject to periodic updates and reanalysis as climate, human use patterns, and other factors change. Practical considerations dictate that re-evaluation on 5-10 year time scales would be reasonable."

- Connectivity between ecological production units: The EBFM team will have to develop an approach for estimating the exchange of productivity across ecological production units. Many fish stocks will span these boundaries (i.e., migratory species) and this will need to be considered.

Again we agree this is important. In the Summary Document we noted "The EPU boundaries defined here are conceived as open and interconnected, reflecting dynamic water mass movements, seasonal and longer-term migration of organisms within the regions, and shifting fishing patterns. These considerations will play an important role in the practical implementation of this place-based management scheme. It will require that interchange among EPUs be addressed."

Currently in Hydra, we treat recruitment processes of three migratory species on a coastwide basis: herring, mackerel. To treat biomass and yield of these species, we integrate seasonal survey information over an annual cycle to compute the mean annual presence of each on Georges Bank and apply exploitation rates to this available biomass.. We are exploring options involving computation of transfer coefficients between EPUs and are attempting to understand the possible magnitude of uncertainties in these estimates.

- Missing information on upper trophic levels: The approach did not include upper trophic levels (e.g., fish) in the definition of ecological production units. However, given the desire to have management units that are relatively stable, the focus on physical, oceanographic, and lower trophic data is advisable.

  Agreed

- New management boundaries may create new difficulties: Re-definition of management boundaries may create difficulties in assigning historic fisheries information (both fishery independent and dependent data) and allocating catch shares. This concern will need to be addressed as the EBFM strategy moves forward toward application.

  Agreed. A major reason we chose spatial units of 10' latitude by 10' longitude in our analysis is that this is the finest spatial scale at which historical catch data can be extracted. For fishery-independent information, we have used post-stratified survey information that conforms to the EPU boundaries in all analyses.

Recommendations

The Panel found the methods for defining ecological production units to be reasonable and recommends that the approach continue to be refined to consider the details of implementing new management units. For example, the Panel recommends consideration of how exchange across ecological production units would be estimated and the appropriate method and timeline for revisiting the boundaries of ecological production units in the future.

Options for estimating exchange rates are now under consideration for use in Hydra and in Ecopath with Ecosim.

If the EPU concept is adopted, the boundaries will be re-evaluated and then updated on a 5-10 year time frame. If evidence suggests rapid rates of change in key input variables, 5 year intervals will be adopted.

ToR 2: Evaluate the methods for estimating ecosystem productivity for the Georges Bank Ecological Production Unit and advise on the suitability of the above methods for defining limits on ecosystem removals as part of a management procedure.

The Panel evaluated the proposed method for estimating ecosystem production potential of ecological production units. The method was a bottom-up approach that was applied to determine fisheries production potential and exploitation for various ecosystem components. The approach utilized information on the: 1) net primary production for two functional groups (nanopicophytoplankton and microphytoplankton), 2) pathway of energy flow in the system, and 3) energy transfer efficiency to estimate total ecosystem production potential. Potential fishery production was then calculated based on applying a 20% exploitation rate on each functional group as described in Moiseev (1994). The approach was illustrated for the Georges Bank Ecological Production Unit. The Panel identified strengths and concerns with the approach and made recommendations for consideration in future work.

Strengths

- <u>Scientifically rigorous method:</u> The basic approach to estimating ecosystem production potential is straight forward and grounded in the scientific literature. In addition, there is good information on lower trophic level productivity in the region to support application of this approach.
- <u>Appropriate for tracking trends</u>: The Panel suggested that the approach is useful for tracking trends in primary production and for understanding how this might impact production at higher trophic levels (considering the lag in transfer of energy through system). This information could be used as a warning sign of changes in the ecosystem and could provide a general context for fisheries management decisions.
- <u>Comparable to previous findings:</u> The initial estimate of Georges Bank fisheries production (220,000 mt) seems to be in the ballpark of estimates produced by others (e.g., 90,000 mt; Link et al. 2008 and 130,000 mt; Collie et al. 2009), although somewhat higher. However, given that the Fogarty et al. estimate includes latent fishery resource production, it is expected to be higher than realized production.

Concerns

- High uncertainty in estimate: The approach of using primary production to estimate fishery production potential is highly uncertain. This estimate was viewed as an appropriate approximation of fishery production; however, the Panel was concerned about the use of this number as a reference point (i.e., a ceiling/overfishing limit). Furthermore, when this number is reported, the associated information on uncertainty should also be reported.

  The panel makes an important point. We provided estimates of total potential yield [including suspension feeding bivalves (expressed as live weight), benthivores, planktivores and piscivores] for 4 time periods in Table 2.3 in the Summary Document. The information is expressed in terms of two measures of central tendency (mean and median) and the upper and lower 25th percentiles. For example, the median potential yield for the period 2013-2016 was 465 kt with upper and lower 25th percentiles of 374- 566kt. Although not reported in Table 2.3, the coefficient of variation of the estimate for this time period was 31%. A broad comparison of the uncertainty in related reference points in current assessments where available (e.g MSY) would be informative to judge the uncertainty in the potential yield estimates in a relative sense.

- Alternative approaches: The Panel suggested that other approaches to estimating fishery production (e.g., multi-species surplus production models, Ecopath model) be explored for comparison. Furthermore, different metrics of potential fish production should be considered (e.g., potential fish production vs. fished species production).

  In the Summary, we reported comparisons with earlier network models for Georges Bank. We also have previous results using multispecies production models implemented at the individual species levels and the functional group levels for Georges Bank that can be brought into the comparison.

  We agree that the distinction between production potential and fished species production is important and have made previous calculations for the latter employing production to biomass ratios and estimated biomass for exploited species.

- Missing information on upper trophic levels: This method is a bottom-up approach and does not utilize information on upper trophic levels in the estimation of ecosystem production potential. It should be noted that the estimates of fisheries production includes both exploited and non-exploited species.

Recommendations

The Panel viewed the methods for estimating ecosystem productivity for Georges Bank as a useful means of tracking an important and dynamic metric of ecosystem status. However, they did not advise using this for defining limits (i.e., reference point) on fishery removals at this time due to the uncertainty in this method. The Panel suggested that the EBFM technical team explore other methods and metrics of estimating fishery production and continue simulation testing limits on removals defined from multiple approaches to resolve the best approach.

We will compile alternative measures based on other network models, multispecies production models, catch histories to provide a range of measures related to production dynamics in this system.

ToR 3: Evaluate the approach and rationale for specifying Fishery Functional Groups as proposed management units.

The Panel evaluated the approach and rationale for specifying fishery functional groups as proposed management units. Fishery functional groups were described as species that are caught together by specified fleet sectors, have similar life history characteristics, and play similar roles in the ecosystem with respect to energy transfer. The approach required characterization of: 1) catch characteristics and targeting practices by fleet, 2) trophic guilds (e.g., benthivore, piscivore, planktivore), and 3) issues of differential risk to species within functional groups based on life history characteristics. The approach is designed to address both technical and biological interactions of species in the definition of the management unit.

Strengths

- Scientifically rigorous method: The approach of using fishery and biological characteristics is reasonable and aspects of this method have been previously published (Garrison and Link 2000, Lucey and Fogarty 2013).
- Addresses technical interactions: This approach enables consideration of biological and technical interactions together in the definition of a management unit. Well-defined fishery functional groups may help alleviate some of the current issues associated with technical interactions in the mixed stock groundfish fishery.

Concerns

- Appropriateness of fishery functional groups as management units: It is not clear if fishery functional groups are the most appropriate management unit. Further work needs to be done to understand whether grouping by both trophic guilds and fishery characteristics will improve and/or simplify management of the system. These units do not map onto existing management units (single-species stocks) or the scale at which harvest is allocated (sectors), and the transition may be a challenge. Furthermore, the appropriateness of the fishery functional group as a management unit will depend on the management objectives which are currently not determined. Therefore, the definition of management units may need to be revisited after final definition of management objectives.

  The fishery functional group approach seeks to identify operational fisheries as high-

level management units. Our objective to establish avenues for direct consideration of the fish-catching process and fishing effort in relation to potential management options. We then identify trophic guilds within operational fisheries to allow checks on system structure. Finally we overlay consideration of species vulnerability to harvesting to afford opportunities to consider additional protections for vulnerable species.

The reviewers quite properly note that this is very different from management of single species stocks. It is however directly connected to the concept management of species complexes under U.S. regulations. Currently, approximately 75% of just over 900 species/stocks managed in the US fall into designated stock complexes; the remainder are managed as individual stocks. Two species complexes are currently managed in this way in the Northeast U.S. – the skate complex and the silver and offshore hake complex. There is accordingly a national precedent to consider this approach and direct experience for two species complexes in this region providing direct experience.

- Dynamics of fishery functional groups: As the availability of fish to the fishery and fisheries practices change, fishery functional groups will change. Due to the definition of these groups being based on historical targeting and catch composition of fisheries in the region, this approach could be inflexible to future changes. The EBFM technical team should consider a method for modification of fishery functional groups to consider future change (e.g., distributional shifts of species or change in fishing behavior). Furthermore, they will need to evaluate the potential changes in fisher behavior associated with the change to EBFM in the region (i.e., quota allocation at the fishery functional group level may change targeting practices).

  This is an extremely important point. Our example of identification of operational fisheries (Lucey and Fogarty 2010), if employed to specify the starting point for defining fishery functional groups, would be periodically updated for the reasons identified by the reviewers. Operational fisheries are defined by species compositions and spatial location and seasonal patterns. The operational fisheries so defined hold certain elements in common with the specification of métiers now employed under the Common Fishery Policy of the European Union for management of mixed species fisheries.

  An alternative approach to deal with shifting species compositions related to factors such as changing targeting practices, environmentally driven changes in distribution, and other factors could entail a focus on specifying fleets with identifiable and persistent characteristics (defined by gear type (including mesh size or other relevant characteristics) and vessel size as primary fishery units with the understanding that the species comprising the catch will be subject to change.

  The trophic guild component of the specification of fishery functional groups focuses on ecosystem roles in energy transfer. It permits direct consideration of a key aspect of ecosystem structure in EBFM -- an issue not addressed in traditional single species management. The underlying idea is that the relative abundance of different species will *change but the biomass of the guild will be more stable than its component parts – and that the overall ecosystem function of the guild can be met. It can accommodate alteration in species composition related to environmental change (e.g. the 'invasion of a new species into the EPU) presuming that the species can be mapped to their trophic role and guild membership. This part of the story depends more on functional roles than species identity per se.*

We emphasize that the difficulty of predicting the effects of environmental change and human behavior must be kept closely in mind. As we strive to anticipate the direction and magnitude of these changes, we must also define strategies and approaches that are robust to change that can only by broadly predicted. Elements of the management procedures outlined in the Summary are intended to meet this challenge. In particular by demonstrating the importance of 'applying the brakes early' in response to system and species level changes and treating the problem at a system level rather a purely species level we hope to define a robust strategy. We believe the simulation results conducted to date point to a promising avenue of approach.

- Individual species/stock concerns: It will be important to make sure that monitoring and attention to single species will not be lost in this approach. There may be stocks that managers would want to continue to monitor and assess at the individual-level based on management concerns.

  Continued monitoring of the status of individual species through fishery-independent and fishery-dependent sources is central to our overall approach, particularly with respect to identifying safeguards for individual species. The focus on Fishery Functional Groups as one approach to dealing with the mixed-multispecies nature of the problem is not intended to downplay the importance of tracking the status of individual species.

Recommendations

The Panel found the definition of fishery functional groups to be a reasonable approach that would enable consideration of biological and technical interactions together in the definition of a management unit. However, the Panel recommends further examination of the appropriateness of this unit for management through simulation testing with a more realistic representation of the fishery functional groups on Georges Bank. The Panel recommends further research into the dynamics of fishery functional groups over time and development of an approach to update management units with changes in the system. In addition, practical considerations of implementing new management units will need to be addressed as these units do not map onto existing management units (single-species stocks) or the scale at which harvest is allocated (sectors), and the transition may be a challenge.

To permit a focus on core aspects of the management procedure, Hydra was implemented *in a simplified form in which three fleets were simulated (without explicit seasonal or spatial considerations as in the full operational fishery specifications. Thorpe et al. adopted a similar strategy in their simulations of the North Sea fishery ecosystem (consolidating 88 metiers into 4 fleets). The simulation model used in the North Sea simulations (LeMANS) is similar in structure to Hydra. We also consolidated the trophic guild component into 4 elements (planktivores, benthivores, piscivorous teleosts, and piscivorous elasmobranches). The latter group also encompasses our vulnerable species categories. We are currently adding two additional fleets to Hydra and can consider finer partitioning of the trophic guild component. We note however, that inclusion of more grouping factors in general would result in some categories with a single representative under the current configuration of 10 species and run counter to the purpose of our analysis.*

Our operational fishery analysis is amenable to consistent periodic updates and these will be undertaken unless an alternative approach (e.g. fleet-based management) is considered preferable.

Should NEFMC decide to retain the sector system in a transition to EBFM, the definition of the fishery component could, in principle, be harmonized with sector management for groundfish, potentially reducing transition costs and difficulties. The current system comprises sectors representing a single gear type (trawler or fixed gear vessels) and some with representatives of both major gear categories. Further modification in permitting structures within sectors, however, may be necessary and desirable.

ToR 4: Comment on the applicability and utility of the strawman management objectives and associated performance metrics which were used to guide the development of operating models.

The Panel reviewed a presentation of the strawman management objectives and associated performance metrics for the EBFM procedure. The strawman objectives were used to guide the development of operating models and outputs of the management procedure. The strategic management objectives presented included:

1) maintain/restore sustainable production levels (ecosystem),
2) maintain/restore biomass levels (functional group/species), and
3) maintain/restore functional trophic structure.

A range of operational management objectives were also presented. These included:

1) Ecosystem and community/aggregate fishing mortality and or total catch is below a dynamic threshold,
2) Fishing-related mortality for threatened/endangered/protected species is minimized,
3) Managed and protected species biomass is above established minimum threshold,
4) Maintain ecosystem structure within historical variation recognizing inherent dynamic properties of the system,
5) Maintain habitat productivity and diversity,
6) Habitat structure and function are maintained for exploited species, and
7) Minimize the risk of permanent habitat impacts.

The performance metrics presented were:

1) Functional group status (proportion overfished/depleted)

2) Species status (proportion overfished/depleted)
3) Landings
4) Biomass at species and functional group levels
5) Stability of landings
6) Large fish index (population)
7) Large fish index (landings)
8) Revenue

The presenter indicated that this was a sample list of potential management objectives and ultimately these objectives would be determined by the NEFMC through outreach and engagement with stakeholders. The presentation also discussed the Magnuson-Stevens Act and outlined how EBFM is consistent with new National Standard 1 guidelines (i.e., NS 1 would allow for using an aggregate approach to estimate the maximum sustainable yield of a fishery).

Strengths

- Reasonable approach: The strawman management objectives were reasonable, high level objectives, but will need to be refined for operational use. The expectation is that these will be refined and expanded upon through the stakeholder engagement process.

Concerns

- Limited in scope: The strawman objectives should not limit the full scope of objectives considered in the MSE. For example, economic and social management objectives should be considered more fully.

  We agree that social and economic objectives must be an essential part of the suite of objectives considered and implemented by managers. Although we have introduced some economic objectives in out treatment of the portfolio approach to minimizing risk, fuller consideration of social and economic objectives will be necessary.

- Single species metrics: Another concern is that the only metric of single species stock status being tracked is reduction below 20% of unfished biomass (B_{lim}). This provides information on reduced stock reproduction potential, but does not give information on reduced yield potential. The fraction of stocks falling below the higher trigger point of the ramp-down harvest control rule (point at which fishing is reduced) should be tracked as a metric as well.

  Although the radar plots presented Figure 4.8 and 4.9 provide only one metric of single species status as properly noted by the panel, in Figures 4.10-4.11 and 4.13-4.14 we show the landings and biomass by individual species in the form of box plots and we will seek to expand the representation of single species information. To this end, we will add metrics for the fraction of stocks falling below the trigger point for the ramp-down scenarios.

- Strawman objectives limit model structure: The Panel notes that the strawman objectives have, in part, defined the metrics that are output from the current MSE framework. As the management objectives evolve, there may be a need to revisit the structure of the model and HCRs as management objectives will need quantifiable outputs to track performance from the model. Furthermore, some of the operational objectives presented (i.e. habitat objectives) are not integrated into the MSE or linked to performance tracking.

Agreed. Once we have guidance through stakeholder engagement on objectives, it will be necessary to reconsider performance metrics, harvest control rules, and model structures (whether new models altogether are necessary and/or the ways in which the current models in use must be modified). Without this guidance we could only hope to provide a proof concept tailored to a sample of possible objectives.

- Strategic and operational objectives not linked: When management objectives are finalized, there should be a clear linkage made between strategic objectives, operational objectives and the associated performance metrics.

  For this exercise, we chose to explore broad strategic objectives "…to maintain overall system resilience and to optimize yield and revenues subject to conservation constraints" consistent with the tenets of EBFM and current fishery practice. Our operational objectives are embodied in the near term sequence of actions related to status determination (including assessment methods and timing of assessments), and actions to be taken in response to change in condition of the system. The performance metrics chosen relate to our strategic ecological (e.g. functional group and species status, biomass, proportion of large fish in the population, diversity, and ratios of functional group biomass levels) and human (revenues, landings, stability of landings, proportion of large fish in the landings) considerations.

  As properly pointed out by the panel, the final strategic and operational objectives will emerge from managers based on stakeholder engagement.

Recommendations

The Panel viewed the strawman management objectives as a reasonable starting point for the EBFM procedure, however, the Panel expects that these will be refined and expanded upon in the future through the stakeholder engagement process. The Panel recommends that additional objectives are explored based on input from stakeholder engagement, these should include biological, economic, and social objectives. Expansion of management objectives may require iteration of the model to accommodate performance measures which are not currently quantified in the current structure.

We will ensure that these important considerations will be an essential part of the stakeholder engagement process to guide the overall process and future model development.

ToR 5: Evaluate the utility of the proposed management reference points as part of a management control rule for ecosystem-based fishery management. These include: an

overall catch cap at the Ecological Production Unit level conditioned on environmental conditions, ceilings on catch for each Fishery Functional Group (defining overfishing) conditioned on aggregate properties, and biomass floors at the single species level (defining overfished conditions).

The Panel reviewed the proposed management reference points for the EBFM management procedure, which included: 1) an overall catch cap at the ecological production unit level conditioned on system productivity, 2) ceilings on catch for each fishery functional group (defining overfishing) conditioned on aggregate properties, and 3) biomass floors at the single species level (defining overfished conditions). The definition of the ecosystem overfishing limit was proposed to be based on the dynamic 'carrying capacity' of the ecosystem as a function of production at the base of the food web. The methods for estimating this value were reviewed under ToR 2. It was not clear how the ceilings on catch for fishery functional groups would be calculated, just that their sum would not exceed the overall cap. Biomass floors were proposed to be calculated at either the fishery functional group (biomass of fishery functional group not to fall below 20% of unfished biomass) or individual species level (biomass of any species not to fall below 20% of unfished biomass).

Strengths

- Reasonable approach: The Panel viewed the proposed approach to define management reference points (i.e., floors and ceilings) as a reasonable approach, however there was substantial concern regarding the details of how reference points would be calculated. The implementation of these reference points will require simulation testing.

Concerns

- Definition of biomass floors: The Panel had concerns about biomass floors for single species and how are these will be defined (e.g., the use of unfished biomass to define the limit, and what percentage of unfished biomass should be used as a limit [i.e. should all species be at 20% ?]).

  Under current US fisheries management, direct relationships between unfished biomass (B_0), biomass and maximum sustainable yield (B_{MSY}), and overfished status are posited. For a symmetrical production function, B_{MSY} is taken to be one half B_0 and stocks are considered to be overfished when biomass drops below one half B_{MSY} or one quarter B_0. Other production functions of course yield different quantitative relationships;

  In the current analysis it must be recognized that in a multispecies context, B_{MSY} for any species is a function of the biomass of all interacting species and not its own internal dynamics alone. In this sense, expressing results in terms of unfished biomass avoids the need to provide an estimate of B_{MSY} conditional on other species (and management actions affecting that species).

  To define the overfished level in our simulation, we selected 20% as one potential choice among many possibilities. It is in the range of choices now made under US fishery management law. In comparable multispecies models, other analysts (e.g Thorpe et al 2015, 2016) have chosen 10% of the unexploited case to indicate an *overfished or depleted condition. Ultimately, the choice will be made by managers. Issues arising from the case where B_0 cannot be estimated from empirical data will be discussed in the section on TOR 6.*

We agree that the issue of choosing different limits for different species deserves careful consideration. Although it differs from current practice in US fisheries management, strong arguments can be made for considering alternatives.

- Definition of ecosystem ceiling: The concept of the overall catch cap is useful, but the Panel was concerned about using primary production as the basis for limiting fishing and it was unclear how the ecosystem ceiling would be applied in fisheries decision making. In theory it seems like the catch cap should not be breached, however, there was concern that this could be risky if this value is viewed as a target. Further work needs to be done to define the role of the ecosystem ceiling in management and the corresponding action that would occur when the ceiling is breached (HCRs need to specify this). The simulations only included action when biomass dropped below floors.

  As noted in the Summary, we stress that the catch cap based on energetic principles is a limit rather than a target. It is intended to provide a general context for production potential and as way to evaluate the implications of changes in primary production. The important concerns raised by the reviewers here are related to issues raised in the discussion of TOR 2 and we will re-evalaute this issue in relation to the recommendations by the panel in relation to TOR 2

- Definition of fishery functional group ceiling: There is a need to clarify the calculation of the catch cap for fishery functional groups. What was proposed in the general description of the management procedure and what was implemented in the worked example for Georges Bank (sum of single species MSYs) were different approaches. If the MSY approach is pursued for this purpose, the MSY for fishery functional groups, should be calculated based on a multispecies model (not sum of single species MSY).

  In our simulations, we test a range of exploitation rates spanning recommendations by Moiseev (1994), and Ivlev (1990) as further developed by Ware (2000) based on energetic considerations. For exploitation rates range from 0.05 to 0.4 in increments of 0.05. For each exploitation rate, there is a corresponding equilibrium catch that gives the ceiling for each functional group. Hydra is effort-based, however, and as an operational matter, we track exploitation rates and the specified floors to implement remedial action. There is an apparent misunderstanding in that we do not use the sum of single species MSY to specify catch ceilings.

- Dynamics of reference points: The Panel was concerned whether these reference points will be responsive to ecosystem change. This concern is not specific to an EBFM approach, but the EBFM team should carefully consider the data used in estimation, how linked reference points will be to historic production, and how often values will be reestimated to reflect current conditions.

Again this is an important point. In Section 2, we tracked one aspect of ecosystem change (change in the f-ratio in overall primary production) and explored its implications for fishery production potential. In this case, re-evaluation occurred at 5 year intervals.

We do believe that tracking changes in primary production (and the f-ratio in particular) can provide critically important insights into overall system productivity and resilience quite apart from the Ecosystem Production Potential modeling approach. Recently, Trenkel (2017) proposed an approach in which exploitation rates can be adjusted according to changes in primary production in a simple management procedure. In the summary document, we suggested that changes in the f-ratio could be used in a similar way.

Recommendations

The Panel approved of the general approach of defining floors and ceilings for use as reference points in an EBFM procedure. However, there was substantial concern about how these numbers would be estimated and applied in operational management. In addition, the Panel recommends further examination of how ceilings will be used in a real-world application (e.g., what action would be taken when an ecosystem or fishery functional group ceiling is breached.

In practice, we anticipate that functional group ceilings would be set based on selection of a target exploitation rate and an estimate of total biomass for the functional group. If a ceiling is breached, it is only known in retrospect after annual catch statistics have been compiled. As in current management practice when single species quotas are exceeded, options include deducting the overage from the next years catch, or some multiyear 'payback' approach.

Once a floor is selected as some percentage of unfished biomass for each species (where this is possible), fishery-independent and fishery-dependent sources and models will be used to provide status determinations relative to the fractional level chosen. This can include the possibility of different percentages for different species),

It must be acknowledged that in many instances, it will not be possible to make any reliable estimate of unfished biomass and alternative approaches must be developed. This situation is routinely faced in current management in which MSY-based reference points are not available. Of the species currently included in Hydra, 60% do not currently have accepted MSY-based reference points and a substantial number of species currently under management on the Northeast US Continental Shelf (approximately 40%) do not have MSY-based reference points or proxies. In these instances, indexed based assessments are employed using fishery independent and/or fishery-dependent sources of information. Typically, a reference period is selected from the available time series and current status determinations are made relative to these reference periods (usually employing a moving average of the final 3-5 years in *the series). We have explored related options for our multispecies analyses, using a Kalman filter to smooth the empirical time series and to examine the recent observations relative to a reference period or percentile of the historical series.*

ToR 6: Review harvest control rules embodying the proposed floors and ceilings approach using the ceiling reference points in ToR 5 to cap removals at the Ecological Production Unit and Functional Group levels, while ensuring that no species biomass falls below the single species floor reference points.

The Panel reviewed potential harvest control rules embodying the proposed floors and ceilings approach to management whereby overfishing is determined at the fishery functional group level and the overfished status is determined either at the fishery functional group or individual species level. Two main forms of harvest control rules were explored: 1) threshold exploitation, whereby exploitation rate is constant until a threshold biomass level is reached (i.e., a fishery functional group or individual species floor), and 2) ramp-down exploitation whereby exploitation rate ramps down (step-wise approach) when a trigger point in reached and ceases then threshold is reached (i.e., fishery functional group or individual species floor). In addition, scenarios were examined which provided additional protection for vulnerable species (e.g., skates and sharks). For each scenario, system-based exploitation rates were simulated ranging from 0.05 to 0.4. The evaluation used performance metrics for revenue, functional group status, species status, landings, biomass, stability of landings, the proportion of large fish in the population, and the proportion of large fish in the landings. Overall, ramp-down harvest control rules, structured with a reduction in exploitation prior to declines in biomass approaching overfished, performed better than threshold harvest control rules. Early intervention preserved resilience as measured by species diversity and representation of large fish in system.

Strengths

- Reasonable approach: If reasonable floors and ceilings can be defined, the Panel indicated that the shapes of HCRs investigated make sense. The Panel expects that the current HCRs would be expanded upon and refined as the approach develops.

Concerns

- Definition of triggers and thresholds: The Panel was concerned about the estimation of reference points that define the triggers and threshold within the HCRs (see ToR 5). How to calculate reference points in an operational manner remains a serious concern.

  Please see our responses for TOR 5
- Lack of status quo comparison: The EBFM technical team has built the EBFM MSE for the purpose of testing fishery functional group HCRs. However, there is no comparison of the performance of this multispecies approach to the current single species management.

  Although not presented in during the review in the interest of time, we have examined in simulation what we believe to be the key issue in the performance of the single-species approach – the difficulty in hitting single-species targets in a mixed-species *fishery with strong technological interactions. For each species in Hydra, we computed single species reference points by fitting Schaefer production models to model output. We focused on the* F_{MSY} *reference point. Hydra is effort-driven and the resulting fishing mortality rate for each species will differ as a function of the selectivity curve employed and the species-specific catchability coefficients. It then becomes necessary to prioritize which species should be the focal point(s) for management. In our simulations, we sequentially set each species as the priority species and examined the realized fishing mortality rates for all of the other species relative to its single species reference point. Consistent with empirical observation for these species, the general results show a pattern of under- or overfishing for the non-priority species with the magnitude of the discrepancy of function of the magnitude of the difference in the* F_{MSY} *targets for the priority and non-priority species. This indicates sub-optimal performance relative to the stated single-species objectives. This fundamental outcome was conveyed verbally at the meeting and simulated results would have been provided on request.*

Several important caveats are necessary. First, in practice the single species reference points are computed using a number of different approaches and at the time the simulations were conducted, six out of ten of the species in Hydra were not managed using MSY-based reference because of data limitations or because single-species assessments using age-structured models had been rejected. Our use of a single assessment approach, while allowing a consistent method of comparison, differs from actual practice. Second, we did not attempt to postulate a direction or magnitude of change in enhanced selectivity or catchability for each species in its role as a priority species. Given proper incentives, fishers can improve targeting but we did not feel we had sufficient information to quantify this factor or to understand the knock-on effects on the non-priority species. Third, while current management for species with specified control rules generally rely on a threshold criterion rather than a ramp-down strategy, a wide variety of remedial measures are employed for species classified as overfished in different management plans. A common set of remedial measures is difficult to identify within and between current management plans. Accordingly, we did not attempt to incorporate any additional measures.

<u>Form of harvest control rule</u>: In general, the form of HCRs investigated was reasonable, however, the use of step functions within the ramp-down HCR was not supported by the Panel. The use of a step-functions can have unintended consequences when applied in management, with small changes in an assessment producing large changes in quotas. This places stress on the reliability of the assessment and can lead to implementation difficulties. The Panel recommends that step functions within HCRs be replaced with a slope.

This suggestion has been implemented in Hydra.

- Ramp-down trigger: The Panel recommends further consideration of the appropriate trigger point (currently 40% B_0) for use in the ramp-down harvest control rules through simulation testing.

  We chose this trigger for teleosts based on the idea of a 75% precautionary buffer. We will carry out simulations on a broader range of trigger points as suggested.

- Hybrid approach: The Panel suggested consideration of a hybrid approach whereby in addition to overall quotas for a fishery group there is a more specific constraint on one (or several) key species (not necessarily only related to life history vulnerability).

  This is very useful suggestion that can be implemented in a number of ways. We assume the motivation is principally directed at species that are at risk under the harvest control rules examined to date. Sub-quotas can be implemented for these species. We are actively exploring incentive/disincentive structures based on a points system in which allocations are made in terms of a total number of points rather than a weight-based quota and different species are assigned different points contributing to the total depending on their status (e.g. species at risk 'cost' more points than healthy stocks).

- Simulation testing: The Panel noted that HCRs were only tested using the Hydra operating model. Ideally, HCRs would be tested using multiple operating models (e.g., Kraken, Atlantis).

  We will implement HCRs selected for testing in Kraken. Work is currently underway on modificatons to Atlantis and EwE (implemented as RPath) and we will explore extensions to tests alternative HCRs in these platorms.

Recommendations

The Panel viewed the proposed harvest control rules as a reasonable starting point, provided the stepwise changes in fishing level are removed from the ramp-down HCR, but recommends that more harvest control rules are explored and that alternative control rules are simulation tested and compared to the performance of current single species harvest strategies. The Panel was concerned about the estimation of reference points (floors, ceilings, and trigger points) within the HCRs and recommends this as an area requiring more development and simulation testing.

As noted above, we will implement these suggestions. As in so many other important recommendations provided in this review, we have been hoping for additional council and stakeholder guidance to allow us to appropriately focus of our efforts. Our intention in presenting a generalized proof-of-concept was intended to solicit additional guidance to ensure our efforts meet the needs of the council. We see this unfolding as an outcome of an MSE process. In the interim, we will explore options recommended to help provide additional background for the MSE process.

ToR 7: Review the structure and application of operating models for Georges Bank.

The Panel reviewed the written materials and presentations on two operating models for Georges Bank: 1) Hydra, a multispecies-multifleet length-structured simulation model; Gaichas et al. 2017) and 2) Kraken, a multispecies production model; Gamble and Link 2009.

Hydra is a ten species, size-structured model, implemented for three fleets: demersal trawl, fixed gear (longline and gillnet), and pelagic trawl. Hydra traces population trajectories of a multispecies assemblage as a function of size, growth, recruitment and survival. Hydra was applied as a basis for testing the EBFM management procedure. Hydra includes technical and biological interactions as the fish species have size structure, which determines interactions and catchabilities.

Kraken is a ten species production model that requires biomass/abundance time series or survey index and a catch time series as inputs. The Kraken surplus production function acts as an operating model, simulating biomasses for 10 species. In the worked example, Kraken was applied for the purpose of portfolio analysis. The portfolio approach involves the application of financial portfolio theory to multispecies fishery management. The approach allows economic risks and returns to be calculated across varying combinations of species' harvest and allows for simulating an optimal harvest strategy for the system. Kraken was also used as the basis for assessing the use of catch ceilings which limit total removals from the ecosystem in the EBFM procedure (work by A. Hart).

Strengths

- Hydra model: The Hydra model provides a good basic structure for this purpose, combining detail and potential realism with moderate run times. This is a peer-reviewed, published model (Gaichas et al. 2017).
- Kraken model: The Kraken model is simpler in form and thus enables different applications due to the speed of model runs (e.g., portfolio analysis).
- Alternative models: There are two potential operating models (Hydra and Kraken). It is good practice to have multiple operating models.

Concerns

- Hydra scope and structure: The Panel suggests that the EBFM technical team evaluate the appropriate number of species for the operating models and expand on the fleet structure to ensure they are able to emulate realistic biological and technical interactions. It is not necessary that the model completely matches the "real world", but it may be necessary to increase the level of detail in the model to approximate population and fishery dynamics for robust testing of HCRs. Another concern with the Hydra model structure is whether the model is stable when moving away from the base scenario (e.g., is there a tendency for populations to crash in the model?).

  We have restructured Hydra to accommodate more fleets. Species can be added to Hydra and we have mapped out the data requirements and coding to do this. It was our understanding based on interactions with the NEFMC EBFM Committee that the species selected met the needs of the proof of concept worked example. In the current implementation of Hydra, we begin by applying three filters to reduce the parameter *space: (1) all species in the model must persist in the unexploited state, (2) the biomass of each species under exploitation must fall within empirically determined ranges using real world data, (3) the catch of each species under exploitation must fall within empirically determined ranges using real world data the latter two constraints ensure that Hydra is producing results consonant with real world observations. Species in Hydra can oscillate (notably for species with over-compensatory stock-recruitment function) but we do not see a general tendency to crash.*

- Hydra trophic interactions: Ideally, the key food components for species within the model should be fully modelled. If this is not possible, then care should be taken with modeling "other food", giving as much realism as possible and checking for model sensitivity to this input. In addition, the trophic interactions in the model do not include interactions at early life history stages and it would be worthwhile for the team to consider how important this may be to the realism of the model.

  Although not presented at the review in the interests of time, we have explored the consequences of employing two levels of 'other food' in the model offering starkly different levels. In these simulations, the model output was not unduly sensitive to this component. We are currently exploring this in more detail

  We will explore options to include interactions at the early life stages. We have less empirical information to go for interactions among larval and post-larval stages but they may well be important. Hydra is currently set up to handle predator-prey but not competitive interactions (as is the case for most multispecies fishery models).

  It would be possible to consider the role of predation by pelagic fish on the eggs and larvae of other species – an issue not now covered in Hydra but one that has been postulated for cod-herring interactions.

- Hydra stock recruit relationships: The Panel questioned the form and range of S-R models included in Hydra. The Panel was concerned with the use of a hockey stick form, as it tends to produce lower compensation than Beverton and Holt models at low spawning stock biomass. On the other hand, the range of curves explored were all to the left of the fitted function, which will provide stronger compensation and perhaps spurious robustness to the effect of fishing in the model (i.e., making it hard to fish-down stocks). In addition, the variability included on the recruitment functions are currently lognormal. This may be too restrictive for some stocks, such as haddock where other methods may be better at approximating erratic high recruitment. The Panel recommends exploring different forms and a balanced representation of possible S-R curves around the fitted function.

  During the review, we focused on the generalized hockey stick model (designed to cover compensatory and over-compensatory forms) because it is somewhat novel and we were interested in reactions to this approach (and of course hoping for a more enthusiastic response!). The standard hockey-stick model is now employed in LeMANS (Thorpe et al. 2015, 2016), a length-structured multispecies model similar to Hydra. However, Hydra is designed to allow use of a range of alternative S-R models and we have fit standard

Ricker, Beverton-Holt, and Shepherd (a three parameter B-H model) to each of the data sets. In our presentation, we showed Hydra simulation results using the Shepherd model first before showing results for the generalized hockey-stick model.

In initial tests, we found that reducing the slope at the origin for the hockey stick models often resulted in rejection of these parameters when passed through the three filters (persistence, biomass, and catch) described above. However we will revisit this issue in a more comprehensive set of simulations in response to this important concern. In a full MSE process, we can readily accommodate the range of S-R models currently included as options in Hydra.

We will implement the alternative function for stochasticity in recruitment as an option in Hydra.

- Further development of Kraken model: The Panel suggests that further development of the Kraken model is needed, including work to evaluate the appropriate number of species in the model and incorporation of more realistic fleet structure, as well as simulation testing of the performance of the operating model.

  These recommendations will be implemented. Currently, simulation tests are being performed on the operating model and programming is under way to incorporate multiple fleets and fishery functional groups.

- Hydra and Kraken model performance uncertain: The Hydra and Kraken models seem appropriate in structure, but realizations of operating models have not been checked. There is a need to evaluate the model against real world observations/trends to demonstrate that these models can produce credible results (e.g., when model is informed by high catch levels on the order of historic catch does the model demonstrate a decline for those species).

  As noted above, filters are employed to ensure that parameterizations of Hydra provides empirically grounded results. In an earlier implementation of Hydra (Gaichas et al. 2017), fishing effort levels ranging up to one and a half times mean effort were explored and provided reasonable responses to exploitation. However, the parameterizations of Hydra in this earlier analysis were quite different than the ones now in use.

  Kraken models have also had sensitivity analyses of increasing or decreasing effort and provided reasonable responses (Smith et al. 2015). As in the case of Hydra, the parameterizations and additionally species were different in these runs, so will need to be re-checked.

- Range of model complexity: There are trade-offs in modelling between providing a detailed representation of ecosystem dynamics as compared to a simple representation that captures the dynamics that matter for a specific question. The Hydra and Kraken simulations could be regarded as an example of each. It would be worthwhile to explore other models that varying in their level of detail and complexity (e.g., models that include the full size spectrum of fish life histories and therefore take account of early life interactions).

We will examine options for the use of other model types including size-spectrum models that provide much finer resolution of the size structure than that employed in Hydra. Models of this type typically focus (as in Hydra) on size structured predator-prey interactions but not competitive interactions.

- Application of alternative operating models: Kraken was used for the portfolio analysis and testing ceilings and Hydra was used for harvest control rule testing. If feasible, the operating models should each be utilized as a basis for the portfolio analysis as well as testing of harvest control rules. However, it is important to note that the two models are not truly independent as Kraken was tuned to results from Hydra. Ideally, the two models would be independent and applied for each purpose. Furthermore, additional alternative operating models could be utilized that include greater complexity (e.g., Atlantis model once update is complete and ecopath model).

  We are currently examining the possibility of testing the portfolio model in Hydra. Independent parameterization of Kraken is possible and we will explore this along with options for testing the HCR. As noted earlier, we do hope to use Atlantis and EwE in future simulation tests.

Recommendations

The Panel viewed the development of two multispecies operating models (Hydra and Kraken) with varying levels of complexity as good practice for testing aspects of the EBFM procedure. The Panel recommends specific areas for improvement for each model. The biggest concern is the need to evaluate the model output against real world observations/trends to demonstrate that these models can produce credible results. The Panel recommends further work evaluating the output of both operating models (Hydra and Kraken) to evaluate how well they can approximate current and past stock dynamics given similar fishing conditions. The Panel also recommends that the operating models should be used for cross purposes if possible (i.e., each be applied for harvest control rule testing and portfolio analysis). In addition, the Panel recommends that additional operating models for the Georges Bank ecosystem (e.g., the Atlantis model which is being updated and Ecopath model that is in development) be considered as a basis for simulation testing.

We will continue to evaluate Hydra and Kraken model results against real world data. We note that attempts to recreate past population and community-level trajectories face (at least) two important challenges: (1) important elements of past dynamics reflect changing environmental conditions and so direct inclusion of historical environmental change would be required in the models, (2) the system has been subject to a complicated sequence of regulatory change with major interventions occurring in 1977 (implementation of extended jurisdiction and passage of the Magnuson-Steven Fishery

Management and Conservation Act), 1984 (World Court resolution of US-Canada boundary dispute), 1994 (establishment of large-scale year round closed areas and change from qualitative management measures to effort (days-at-sea) management and 2010 (implementation of sector management and Annual Catch Limit management of groundfish. Superimposed on these major changes has been an array of regulations such as a series of mesh size increases, changes in minimum legal size limits etc.

We agree entirely with the recommendation to use models for cross-purposes wherever possible and to add additional operating models to the mix.

ToR 8: Review ecosystem assessment models and required data sources, as applied to the simulated data from the operating models in ToR 7.

During the meeting, the Panel reviewed a presentation on ecosystem assessment models and their required data sources. The proposed alternative assessment methods included a: 1) model-free simulated survey index, 2) multispecies production model, and 3) multispecies delay-difference model. The models require biomass and catch data as inputs. The proposed models range in their complexity, enabling evaluation of whether simpler assessment models can capture population dynamics of a complex underlying model. A modeling efficiency index used in evaluating the performance of the stock assessment. The performance of assessment models was tested with white noise only, however, in the future, bias can be added to performance testing. Simulation revealed that the more complex delay-difference model performed similarly to the simpler production model.

Strengths

- Comparison of multiple models: The comparison of multiple alternative models is a good approach to understand the appropriate model and level of complexity for the ecosystem assessment model.

Concerns

- Multispecies vs. single species assessment models: The Panel noted that multispecies assessment models were examined, but no comparison was conducted between the performance of multispecies and single species assessments.

  Our intention was to test the performance of two simpler models against the information provided by the more complex multispecies model Hydra. These comparisons included the use of white noise observation error and two levels of autocorrelated error. The results indicated that the simpler models can capture population and community-level trajectories well under moderate levels of stochasticity and observation error but, as to be expected, performance degraded with higher levels of error.

  We did not attempt to replicate age structured single assessments in Hydra since Hydra is size based. Because we include a model-free assessment model based on *simulated survey data in Hydra, we did replicate the basic process for index-based assessments (6 of the 10 species).*

The idea of testing the simpler production and delay-difference multispecies models against the actual assessment results could be profitably employed using real world data. Although we have been fitting these simpler model types to actual survey and catch data a deeper comparison against the specific results from stock assessments is worth pursuing. This is a different issue than the one raised by the panel but hopefully would contribute to the addressing the question underlying the recommendation.

- Testing alternative assessments and HCRs: The testing of alternative assessment methods (e.g. multispecies assessments) should be conducted separately from testing of alternative HCRs.
-

Agreed

Recommendations

The Panel viewed the comparison of alternative models as a good approach to understand the appropriate model and level of complexity for the ecosystem assessment model. The Panel recommends that the alternative multispecies assessment models be compared to single species models. Furthermore, the Panel recommends that evaluation of new assessment methods and new harvest control rules not be conducted simultaneously, as this will make it difficult to evaluate what was causing any successes or failures in the simulated management.

ToR 9: Review simulation tests and performance of the proposed management procedure incorporating the floors and ceilings approach, given the set of EBFM goals and objectives.

The Panel reviewed written materials and presentations on simulation tests and the performance of the proposed management procedure as implemented for the Georges Bank example. The Panel was instructed that performance was not being reviewed for the context of implementation for management, but to evaluate the approach.

Strengths

- Reasonable performance: The Panel noted that the initial results presented during the review seem reasonable in terms of performance based on their response to different forms of harvest control rule, although more critical evaluation of performance is required.
- Evaluation of ceilings: The Panel found the simulation testing of a range of ceilings and their impact on the performance of the EBFM procedure to be very useful and this work should be continued (A. Hart presentation)

Concerns

- Limited simulation testing: The Panel suggested that a broader representation of simulation results is needed to fully evaluate the performance of HCRs in the future. This should include a status quo comparison where the current single species management approach is approximated for comparison to the EBFM approach. Furthermore, one factor within the EBFM procedure should be changed at a time to be able to fully evaluate its impacts on performance. More generally, wherever there is a simplification (e.g., thresholds, trigger points, global exploitation rates, FFG structure) in the model, the Panel recommends that the effects of adding realism are investigated for each simplification separately. It may be that some of the current simplifications are justified, increasing speed and robustness without harming accuracy, but this needs to be tested.

Given additional guidance from the Council and a formal stakeholder process, we hope to define the full scope of issues to be addressed by further simulation as appropriately recommended by the panel. Our intention had always been to conduct more intensive simulations beyond the ones included in the proof of concept results conducted for and presented at the review.

- Presentation of HCR testing results: It is important to note that the performance metrics shown in radar plots were normalized to the highest value across simulations (i.e., highest value was defined as 100%) which can lead to potential misinterpretations of performance. Further work resolving management objectives with stakeholders may help to define the desire performance of the system and allow for performance to be evaluated relative to these values. The box plots will need some refinement for clarity (e.g., labels, similar scales, titles, etc.) in final reporting.

  We will adjust the box plots as suggested and will explore other options for the radar plots. Our intention with the latter was to permit a standardized representation across the scenarios test.

- Exploitation rates in HCR testing: In the current presentation of results the initial global exploitation rates used in the simulated scenario were shown, but not the realized exploitation levels. Information on realized F and realized F/nominal F would help identify the degree to which catch in a given FFG was being reduced by the single species protections within the HCRs.

  We do compute the realized exploitation rates and inspect these in a series of diagnostic plots that also include predation mortality rates, total mortality rates etc. We can bring these types of information forward as requested

- Alternative performance metrics: The current overfished metric tracks the fraction of time spent in a depleted state. This is problematic as it is influenced by the recruitment at low stock sizes. Alternatively, this could also be assessed by counting how many stocks crash at least once in any given 10-year reporting period.

  We will implement this suggestion

- Portfolio analysis It was unclear how the portfolio analysis will be used in the EBFM procedure. Further linkage and description of the role portfolio analysis could play is needed.

Recommendations

The Panel noted that the initial results presented during the review seem reasonable in terms of performance, however, the performance of the EBFM procedure cannot be fully evaluated at this stage due the preliminary state of the work (i.e., many decisions need to be finalized both on model details and management objectives) and the limited nature of simulations run. The Panel suggested that a broader representation of simulation results is needed, including a comparison of EBFM to single species management, to fully evaluate the performance of the EBFM procedure. Furthermore, one factor within the EBFM procedure should be changed at a time to be able to fully evaluate its impacts on performance and the impact of model simplifications should be critically evaluated. The simulated output is an example of how performance would be evaluated, and the Panel provided specific suggestions on the presentation of results.

It is our intention to fully address the many helpful comments provided by the panel and to follow this advice once we have further guidance on some of the issues raised by the panel. It is our hope that this guidance will emerge from a formal MSE process