Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Proceedings Series 2024/022

Newfoundland and Labrador Region

Proceedings of the Regional Peer Review on the 3Ps Cod Assessment Framework

Meeting date: October 8-10, 2019
Location: St. John's, NL

Chairperson: Karen Dwyer
Editor: Aaron Adamack

Science Branch
Fisheries and Oceans Canada
PO Box 5667
St. John's, NL A1C 5X1

April 2024

Foreword

The purpose of these Proceedings is to document the activities and key discussions of the meeting. The Proceedings may include research recommendations, uncertainties, and the rationale for decisions made during the meeting. Proceedings may also document when data, analyses, or interpretations were reviewed and rejected on scientific grounds, including the reason(s) for rejection. As such, interpretations and opinions presented in this report individually may be factually incorrect or misleading, but are included to record as faithfully as possible what was considered at the meeting. No statements are to be taken as reflecting the conclusions of the meeting unless they are clearly identified as such. Moreover, further review may result in a change of conclusions where additional information was identified as relevant to the topics being considered, but not available in the timeframe of the meeting. In the rare case when there are formal dissenting views, these are also archived as Annexes to the Proceedings.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1701-1280
ISBN 978-0-660-71184-3 Cat. No. Fs70-4/2024-022E-PDF

Correct citation for this publication:

DFO. 2024. Proceedings of the Regional Peer Review on the 3Ps Cod Assessment Framework; October 8-10, 2019. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2024/022.

Aussi disponible en français :

MPO. 2024. Compte rendu de l'examen par les pairs régional du cadre d'évaluation de la morue de la sous-division 3Ps; du 8 au 10 octobre 2019. Secr. can. des avis sci. du MPO. Compte rendu 2024/022.

TABLE OF CONTENTS

SUMMARY

A Regional Peer Review process was held in St. John's, Newfoundland and Labrador (NL), from October 8-10, 2019 to assess the methodology for estimating the population size and other stock status indicators of Atlantic Cod in Northwest Atlantic Fisheries Organization (NAFO) Subdivision 3Ps (hereinafter 3Ps). The assessment framework meeting reviewed and considered candidate models of population dynamics for 3Ps (Atlantic) cod that incorporated multiple data sources (including commercial catch). The data sources appropriate for modeling the 3Ps cod population were determined at a data review meeting held in May 2019 (Varkey et al. In Prep[1]).

This proceedings document contains abstracts from meeting presentations and summaries of discussions. The Terms of Reference, Agenda, and List of Participants are detailed in the Appendix. Participation included representatives from Fisheries and Oceans Canada (DFO) Science and Resource Management, the fishing industry, the provincial Government, academia, and non-governmental organizations.

In addition to these Proceedings, Research Documents produced from this meeting will be made available online by the Canadian Science Advisory Secretariat (CSAS).

[1] Varkey et al. In Prep. Report on data-review in preparation for the assessment framework of 3Ps Cod stock in southern Newfoundland. Can. Tech. Rep. Fish. Aqua. Sci.

PRESENTATIONS

OVERVIEW OF AVAILABLE INDICES AND LANDINGS

Presenters: D. Ings

Abstract

An overview of the landings and research data available for modeling the 3Ps cod population was provided. Reported landings were available from 1959 to present. The time series shows a decrease in landings during the 1970s followed by some increases in the 1980s and moratorium from 1993 to 1997. Reported landings increased to approximately 30 kt following the moratorium, but generally have declined since then. Stratified random bottom trawl surveys in 3Ps have been conducted during various periods by France, Canada, and the Canadian industry. France conducted winter surveys of offshore areas from 1978 to 1992, using a Lofoten trawl. Canada conducted winter/spring surveys of areas offshore and closer to shore from 1983 to present, but there was a gear change from an Engel to a Campelen (shrimp) trawl between 1995 and 1996. The Canadian industry also conducted an autumn survey of offshore areas from 1997 to 2007 using a Campelen trawl. Age disaggregated catch rates (i.e., catch rates by age) from the sentinel survey, conducted by inshore fishers using gillnets and longlines, during June to November of each year were available from 1995 to present. All research indices were quite variable with many showing year effects. Few surveys bridged the moratorium period.

Discussion

The timing of the DFO Research Vessel multispecies survey (RV survey) has varied, with spring data ranging from February to April over the time series. This varied timing may introduce uncertainty due to seasonal changes in fish distribution. A reviewer asked how data was selected for use in the model, given the changes in timing in the 1980s and early 1990s. The author explained that differences between the available habitat and occupied habitat were investigated to assess the impact of possible seasonal distribution shifts on survey results. Significant differences were not identified in cod habitat used during the period before 1993 (when the surveys using the Engel trawl in winter ended) compared to the period following.

The sentinel survey employs two different gear types: line trawl and gillnet. There was discussion about how the differences in the spatial application of gear types may impact survey results. Recent gillnet survey sites have been concentrated in Placentia Bay and along the coast. The line trawl survey, however, extends west from the Burin Peninsula. Despite the spatial differences in the use of these gears in the 3Ps sentinel survey, tagging data shows movement of cod throughout the sampled area. It was noted that the strength of the coastal components to the 3Ps cod stock may have changed over time. This calls into question the coherence between the time series. As a result, the model iterations presented later in the meeting included runs both with and without the sentinel survey data.

A meeting participant requested that the author elaborate on the objectives, strengths, and limitations of the sentinel and RV surveys. The sentinel survey is an inshore fishery survey. It's strength is that it covers the inshore area, but it is limited because it can only measure catch rate it is unknown how large a component of the stock the sentinel survey covers. The RV survey is a fishery-independent, randomly stratified trawl survey. While in most cases the two surveys target different areas of this stock (inshore vs offshore), in some years the RV survey has completed inshore sets. Participants noted that in the RV survey index years where there are both inshore and offshore sets, the inshore portion of the survey doesn't appear to be a significant driver of the index. The author added that the biggest differences observed between the inshore and the offshore survey sets were among younger age classes.

The meeting also discussed whether there was any correlation between the inshore RV survey and the sentinel survey. While this has not been investigated in detail, the author did not expect a match between the two surveys, primarily due to differences in gears employed. The RV survey catches smaller size classes than the sentinel survey, however the results might be more strongly correlated for older ages of fish. It was noted that the sentinel survey wasn't established as a recruitment index, but was established to cover inshore areas that the RV survey could not reach. Survey timing is also different between the inshore RV survey and the sentinel survey: the RV survey provides a snapshot of conditions in spring whereas the sentinel survey extends over several months (July to September) and includes the cod spawning period.

RE-CONSTRUCTION OF CATCH-AT-AGE

Presenter: H. Penney

Abstract

Catch-at-age is important for stock assessment modelling because it provides the age structure of the commercial and sentinel catch. Length frequencies are compared to age-length keys. The age length keys are generated on a subset of the catch that is sampled by fisheries observers (i.e., their otoliths are removed and used to age the fish). Using the two metrics (length frequency and age-length key) the age structure is determined for the catch each year. Length frequencies and age length keys are matched by each gear type, area, and time of year (quarter), to make age frequencies of the catch. The age frequencies are used to determine what proportion of the catch weight comes from each age. The quality of the match is determined by looking at the sum of products. When a perfect match age-length key was not available, a biologically relevant next best age-length key was used. This creates an estimate of the age distribution for the catch each year.

There were two main issues with catch-at-age:

1. before 2011 catch-at-age was completed in another software so the estimation and bump-ups (scaling the measurements from the sampled fish to the total catch) were conducted in a different way; and
2. the matching of length frequencies and age-length keys can be subjective.

The goal of this project was to reconstruct catch-at-age as far back as possible. Catch-at-age was reconstructed starting from the most recent years and working backward for 2017 to 1994, and 1992 to 1979 without issue. 1993 was missing data, and 1978 to 1974 should be revisited because the foreign catch amounts were suspect. For 1971 to 1973, the data documentation was not available, so it was not possible to reconstruct. Reconstructing 1959 to 1975 was possible but those reconstructions are likely not as accurate as after 1974.

Overall, there were few differences in the old vs new catch-at-age, except:

1. more age 2 fish occur early in the time series; and
2. cohort tracking is apparent using both methods. There was correlation among all age groups except for age 2 fish.

In recent years most of the fish are aged 5–8, because those are the ages that the fishery is largely targeting. Another interesting conclusion that arose from this process was that weights are decreasing for age 7+ cod over time, which was also reflected in the RV surveys.

Discussion

Cohort consistency was discussed: cohorts are generally consistent across analytical methods and over time.

One participant noted that the age-length keys presented were gear-specific and asked if gear was functioning as a proxy for location. The author explained that they are separated due to gear selectivity; for example, gillnets appear to catch faster-growing fish than other fishing gears. They agreed that there are possibly confounding effects between gear and geographic location.

Clarification was requested on matching of length-frequencies and age-length keys. Length-frequencies are prepared by month, gear, location, and country of vessel origin. The data in age-length keys are aggregated across multiple categories as only a fraction of the fish are aged (i.e., have otoliths removed and aged). When trying to match length-frequencies to age-length keys, there is not always an exact match. A decision tree was developed to identify the best match between length-frequencies and age-frequencies.

A participant observed that the mean weights-at-age indicated that the older cod seemed to have grown quite quickly to large sizes, which is the opposite of what is generally seen (e.g., young fish typically grow faster than old fish) and asked if this may be due to sample size of otoliths (how many fish were aged). Upon investigation, it did not appear that sample size is an issue; for example, 2,389 otoliths were aged across all gears for the catch-at-age analysis for 3Ps cod in 2016. Further clarification was requested on why growth rates seemed to have been faster for the older cod with age 7–8 cod weighing around 3 kg and ages 10–12 nearly double that within a few years (~5 kg). It was suggested that older cohorts appeared to have grown faster than more recent cohorts. This could have emerged through an issue with sample size by age (with the coefficients of variation for the older fish being higher than the coefficients of variation for the younger fish), it could be spatial variation in growth rate, or it could be the result of a period of poor conditions for the slower growing cohort in question.

Catch estimates were also discussed. A consistent method was used to bump-up the catches as there were not always representative age frequencies available for every location and time. Catches were bumped up by unit area and then for the quarter. If an age-frequency were missing for a particular area, gear, and time, the next closest quarter and/or the next closest area using the same gear would be applied to that area. For most years post-2000, the data were adequate to generate catch-at-age estimates. However, data are patchy for some gears. Earlier periods have more data gaps, with the 1990s during the moratorium being particularly challenging, as catches and sampling were low.

MODELING CATCH AND STOCK WEIGHTS

Presenter: N. Cadigan

Abstract

No abstract provided.

Discussion

A participant asked for clarification the divergence between spawning stock biomass and the catch weight at-age. Spawning stock biomass is representative of trends in the population whereas the trend in catch weights at-age reflects landings, i.e., these (catch) data are subject to length bias based on the fishery landings.

Participants discussed the differences between previously used stock weights and updated stock weights. The increasing trend in ages 4, 5, and 6 cod, from the commercial data is coming from inshore gillnets in Placentia Bay. This may be due to an increase in the use of 6-inch gillnets starting roughly in 2008. Another possible explanation provided for the discrepancy in stock weights for ages 3–6 fish may be ‘high grading’ (the practice of selecting for the most desirable) small fish. It was noted that the discrepancy does not appear in the line trawl data. It was also noted that there have been a number of changes in what gears are used, mesh sizes, and where the gears are used, over the years. There was agreement that if there is a change in procedure for calculating stock weights, the consequences need to be described when communicating about the stock changes.

INFORMING NATURAL MORTALITY

Presenters: G. Robertson

Abstract

Various tags have been placed on Atlantic Cod in the Northwest Atlantic for decades, and these programs have helped to determine fish movement and stock structure. However, unbiasedly estimating mortality from tagging data requires several assumptions that usually cannot be met with a basic tag and release program. Starting in 1997, an annual cod tagging program was initiated that addressed many of these confounds. Specifically, methods to estimate tag reporting rates, tagging mortality and tag loss were all developed. The Northern cod Assessment Model for NAFO areas 2J3KL (NCAM) incorporates all of these estimates in a space-state model framework and further addresses the seasonality of the cod-tagging program, the seasonality of fishing, and incomplete mixing of fish directly after tagging. The tagging program in 3Ps is operationally similar to 2J3KL, however, an integrated population model is not yet available for this stock. To estimate fishing and natural mortality in 3Ps, the (R) scripts within NCAM that addressed the tagging data were extracted and recompiled into a stand-alone model (TagEst). 3Ps is a stock complex, and tagging experiments in different parts of 3Ps may lead to different estimates of fishing mortality (F) and natural mortality (M). A further modification of TagEst was added to allow for a fixed effect of NAFO subdivision where the tagging was conducted. In general, TagEst was able to estimate annual F and M from the tagging data, over the period of 1997–2017 (using Fortune Bay, 3Psb, as the reference level). F was sensitive to the tagging subdivision used as the reference level, while M was not sensitive to where the fish was first tagged. These results show that M in 3Ps is a significant mortality source, and shows important annual variation. The challenge of using tagging data in an assessment model for 3Ps remains in how to estimate an average F for the entire division when tagging efforts and F appears to vary across 3Ps.

Discussion

One participant noted that there were spikes in M in the early 2000s, 2008–09, and 2015–16. A comparison of these results to M for Northern cod from NCAM found some broad similarities, with spikes in M between 1999 and 2004, 2010, and another increase beginning in 2017. It appears that M estimated for 3Ps cod was more variable than M estimated for Northern cod.

Participants requested more detail on the tagging program’s length and spatial coverage. The current program has been running since 1997, but the level and extent of effort have varied over time ranging from 100–10,000 cod tagged annually in the Region. The tagging data does not represent a randomized experiment with even distribution and effort. There are also issues around the timing of the tagging. As a result of all of these issues, this data is not ready for inclusion in the assessment model.

CONDITION-CORRECTED NATURAL MORTALITY FOR ATLANTIC COD IN NAFO SUBDIVSION 3PS

Presenters: P. Regular

Abstract

Starvation is a ubiquitous process in nature as all animals depend on finite resources to survive. Food resources often vary seasonally and, as such, individuals must endure times when they rely on energy reserves to fuel the basic processes of life. Limited food resources can ultimately lead to starvation-induced mortality and, depending on the scale of the food limitation, this can have population-level consequences. Previous research has indicated that starvation-induced mortality may be revealed by estimating the proportion of individuals experiencing severe emaciation in the population. Using data from both the RV and sentinel surveys, an index of starvation-induced mortality was derived from proportion of cod in poor condition. The greatest proportions of cod in poor condition were observed through the critical spring period and, as such, most starvation-induced mortality presumably occurs at this time. This index of starvation-induced mortality appears to be increasing in recent years (since ~2004) and the trends correspond with tagging-based estimates of M. These results indicate that starvation-induced mortality represents a non-negligible component of the M experienced by the stock and implies that prey availability may be a factor limiting the productivity of cod in 3Ps.

Fulton's condition factor (K) is an index of body condition, which is determined by measuring the weight and length of individual fish. It relies on the assumption that heavier fish of a given length contain more energy reserves in the form of body fat than those that weigh less. In this study, both K and a relative body condition were calculated.

Discussion

Clarification was requested on the selection of a relative condition threshold. For fish below the Fulton's threshold (K = 0.65) established by Casini et al. (2016), the mean relative condition value was 0.85. This value was conservative compared to the laboratory-based values from Dutil and Lambert (2000). The author suggested that the mean from the upper quantile of individuals with K below 0.65 may provide a closer link to the relative condition index. The threshold would be more comparable to laboratory findings. There is room to test different thresholds which will affect estimates of mortality, however, the broader patterns and trends in mortality estimates will remain consistent.

It was noted that the scaling used for this analysis was very important. Some meeting participants recommended considering the general trends in M as useful, without placing too much emphasis on the magnitude at this time.

One participant asked for clarification on a 0.2 correction to the estimate of M from starvation. The 0.2 represented an estimate of mortality from other causes including predation, parasitism, and various other components.

It was also suggested that the length-weight relationship should be rerun for the comparison of Fulton's K to relative condition using only fish with lengths in the size range of the fish used in Dutil and Lambert (2000).

One participant asked about the relationship between the tagging data and the condition data. It was suggested that an option may be to optimize the condition threshold using the relationship with the tagging data. Essentially, this would use the tagging data to define the magnitude of condition-corrected M.

Almost all of the predicted M values were lower than the observed. The author clarified that this was because the normal distribution applied to the predictions doesn't capture the tails of the observed distribution.

There were questions about the fish selectivity differences between the sentinel and RV surveys. The relative condition data is consistent between the two surveys throughout the year. There were no biases apparent in the residuals for either survey. However, further examination is warranted. It was noted, for example, that most of the estimated M was occurring in the spring, with low M expected in the summer and fall. It is possible that gear selectivity is therefore not an issue for estimating mortality, if most of the mortality occurs in the spring when fish are being sampled in the RV survey.

A participant asked for additional detail on the calculation of M from relative condition. The response was that the conversion is based on the proportion of cod classified as poor condition observed in the survey. The model also provided the mean and variance in condition by month, and a normal distribution is applied to calculate the proportion of fish in the population expected to fall below the critical condition threshold. This value is then converted to an instantaneous rate and summed across the year to estimate condition-based mortality.

This analysis relies on the assumption that fish below a critical condition threshold are going to starve and die. This is an extrapolation from laboratory work that shows that once a fish falls below a certain threshold they will die of starvation. Once a fish eats away its muscle protein it becomes likely to be eaten, simply die, succumb to parasites, and/or continue to decline in general health.

A participant noted that M has been increasing over time and asked about possible connections to ecosystem conditions. Condition can be considered an indirect measure of prey availability. Recent studies (i.e., Bernier et al. 2018; Koen-Alonso and Cuff 2018) show that the cod habitat in 3Ps is warming and more southernly, warmer-water species like silver hake are moving into the area. However, the direct relationships between temperature and condition or increased competition and condition are not clear. It was speculated that increased competition for prey could be a driver. It was suggested that perhaps ecosystem productivity and pelagic fish availability be examined further for connection to trends in cod condition and mortality.

Another participant questioned the condition threshold. In some cases, in the laboratory studies, fish that were starved were able to recover once feeding was resumed. This suggests that fish in the wild may be able to do the same. However, it was noted that recovery may be possible in a controlled environment (where fish are being fed); whereas, in the wild fish with limited muscle mass may struggle to capture prey once they fall below a certain body weight. Additional research would help finetune the threshold and determine if fish can recover from starvation in the wild.

The meeting was not able to reach consensus on whether condition-based M should be applied to the assessment within the time available. It was agreed that this discussion, and the model runs that used condition-based M, would be revisited later in the meeting.

FISHING INTERVIEWS TO INFORM CATCH BOUNDS

Presenter: E. Carruthers

Abstract

Fisheries catch had not been used as an input into the 3Ps cod assessment model, in part due to uncertainty in the catch time series. Other Canadian stock assessments have used censored catch models with catch bounds based on fishers' knowledge interviews or industry surveys.

Here we interviewed 23 fixed gear harvesters who, collectively, had used trap, handline, line trawl and gillnet gear to fish for cod from 1956–2018. Harvesters' home ports were broadly distributed from St. Brides, in the east, to Burgeo in the west. Information provided in these interviews can be used to inform:

1. the length of catch time series to be used for the 3Ps cod assessment;
2. uncertainty bounds for the catch time series, at least for the Canadian fixed-gear portion of the time series; and
3. identify opportunities to improve the assessment and management of 3Ps cod.

There was a clear distinction made between the pre- and post-moratorium periods with major differences in how the fishery was prosecuted and how landings were reported. During the pre-moratorium period, trap fisheries accounted for a large portion of fixed gear landings. Undersized cod caught in traps were either discarded and/or salted for personal use and not in receipted landings data. Discarding due to depredation is not size-selective and was consistent across the time series, except where there are specific discarding events. Fish harvesters estimated approximately 2% of the catch biomass was discarded from gillnet fisheries throughout the time series due to depredation, except in regions and time periods associated with long soak times and sea lice and/or hagfish depredation. Harvesters not only identified regions associated with higher depredation but also detailed their strategies to minimise waste due to depredation. Incentives for size-selective discarding of larger fish likely occurred between 1998–2002 and between 2006–10. The impact of these periods and/or regions on the overall fixed-gear time series appears to be minimal given the proportion of landings relative to annual 3Ps cod landings. Harvesters' comments on dockside monitoring corresponded with data from the dockside monitoring program, with close to 100% monitoring at high volume ports over the past 10 years. Fishers' knowledge interviews not only helped define catch bounds and different reporting timelines but also helped identify ways to improve assessment and management, such as through modifications to tagging and logbook reporting programs.

It should be noted, however, that any stock-wide estimates of catch reporting levels would need to include estimates from the mobile and international fleets in addition to the Canadian fixed gear fleet.

Discussion

One participant asked for clarification on how a licensing requirement for 100% dock-side monitoring interacted with the tiered port approach, where Tier 1 ports have 90^{+}% coverage, Tier 2 ports has 60^{+}% coverage, etc. The presenter clarified that the requirement for 100% dock side monitoring has not been achieved since the beginning of the program.

A participant commented on the finding that high-grading was a problem in the late 1990s/early-2000s. During this period there was a divergence in weight at-ages (3–6) of the commercial fish noted in the earlier presentation by N. Cadigan. However, the meeting could not identify a compelling reason that high-grading would result in the trend being observed where weights were increasing for early ages and decreasing for older ages. Further, the trend has continued to today with little evidence of high-grading. The fish survey includes a question on size-selective discarding however there isn't an estimate available of the prevalence of high-grading in this fishery.

It was noted that this presentation only covered the fixed gear fishery. Thus, care is needed when extrapolating the survey findings to the fishery as a whole.

DISCUSSION ON CATCH BOUNDS

Presenter: General Discussion led by D. Ings and D. Varkey

Abstract

No abstract provided.

Discussion

The presenter began the discussion by pointing out that they had worked with Fish, Food, and Allied Workers Union Unifor (FFAW Unifor) on the fisher interviews and was able to compare the outcomes from that work with other data being used for this modeling effort. Specifically, the Canadian and NAFO landings datasets. The wider bounds in the early periods were due to discrepancies in the Canadian and NAFO landings datasets. For the more recent years, the data used in the models seem to align with the information from the interviews.

One participant noted that when permitted, there was discarding in the offshore fishery, but the exact amount to inform catch bounds is not known. He also thought that there may be uncertainty in the landings data due to both under-reporting and over-reporting pre-1992. The problems have been resolved through increased catch monitoring, which has narrowed the catch bounds.

One participant commented that there seemed to be 3 periods where there needed to be wider catch bounds:

1. The period following the expansion of Canada's jurisdiction (1974–77). The reported landings for the stock differ from those used by NAFO during this time.
2. Pre-1992 there is a period when the French landings were much higher than those accepted by NAFO for the stock.
3. For the more recent period, there are issues with high-grading and some cases where offshore vessels over-shoot their quota (and subsequent dumping) during the reopening period.

The presenter felt that further investigation of the sources of uncertainty may ultimately support an argument for wider catch bounds. It was suggested that different widths for the bounds be tested for the earliest period due to high levels of uncertainty associated with the reported landings (1959–72).

For the pre-1992 period, the catch landings were less precise and there was some discussion about the support for the applied catch bounds. It was noted that for Northern cod, the upper bound is 1.5. However, unlike for Northern cod, there appears to have been no foreign fishing (e.g., fishing by countries other than Canada or France) on the 3Ps stock since 1978. Whereas for Northern cod, there was foreign fishing until 1992. It was also noted that there were very large catches at the beginning of the time series (prior to the 1970s), and a 1.1 multiplier on those early catches is a substantial amount of fish as opposed to later years when the landings were smaller.

For the late-1980s to early 1990s, misreporting of catch is largely attributed to the ongoing negotiation of catch shares between Canada and France. The stock was assessed by NAFO at that time, however the landings reported by France were not accepted due to evidence of over-reporting. Additionally, the boundary between France and Canada's Exclusive Economic Zone (EEZ) within 3Ps was set during this period. It can be rationalized that there was possible political pressure to inflate reported landings, while that decision was in deliberation, to influence quota shares.

For the early periods wider catch bounds were largely based on the discrepancy between the accepted landings for the stock and the landings reported to NAFO. One participant suggested that the bounds be kept wide as a precautionary measure, and there should be presentation of evidence before using narrower bounds. The chair noted model runs have been prepared with different sets of bounds, and these runs will be examined and discussed later in the meeting.

Following this discussion, the meeting reached consensus that the presented catch bounds were appropriately supported by the available data and history of the fishery.

STATE-SPACE STOCK ASSESSMENT MODEL FOR 3PS COD

Presenter: N. Cadigan

Abstract

The implementation of a state-space stock assessment model (called SSM) for 3Ps cod is described here. An assessment challenge for 3Ps cod is that there are uncertainties and possible biases in fishery landings information. The censored likelihood approach was used based on the best available information on landings bounds to address this uncertainty. SSM only uses information on lower and upper bounds of what the real fishery landings were. Information on the age-composition of the catches was included in the model fitting using a compositional data likelihood based on the continuation ratio logits of catch proportions-at-age.

Otherwise, the model described is formulated similar to the typical state space model SAM (Nielsen and Berg 2014, Berg and Nielsen 2016) used in International Council for the Exploration of the Sea (ICES) assessments, with some differences in the stochastic model for variation in fishing mortalities (F’s) and the likelihood function for survey indices. These indices have a normal distribution with a constant coefficient of variation, whereas SAM is based on the lognormal distribution. Both distributions have the same mean and variance structure, but an advantage of using the normal distribution is that indices with zero values can be used for estimation, whereas they cannot be used with the lognormal. This is an important issue when there are many zeros that are not distributed at random throughout the ages and years, which is the case for 3Ps cod.

The RV survey was extended to include inshore strata in 1997, and indices for all strata are called the RV_IO indices, while indices for the strata surveyed prior to 1997 are called the RV_OFF indices. The age-distribution of 3Ps cod varies spatially, such that more young fish are found closer to shore. The RV_OFF and RV_IO were treated as separate indices with different catchability parameters, but also included was a likelihood component for the age-based differences in the RV_IO minus RV_OFF since 1997 to constrain differences in the catchabilities for these two indices.

Basic model inputs are:

- Model years 1959–2019 and ages 1–14+.
- Landings bounds for 1959–2018.
- Two scenarios of uncertainty about landings:
    - Good landings multipliers.
    - Uncertain landings multipliers.
- Catch age compositions, for 1959–2017 and ages 3–14+.

- Survey indices:
  - Can_RV_IO, 1997–2018 (not 2006) and ages 1–14+.
  - CAN_RV_OFF, 1983–96 and ages 1–14+.
  - ERHAPS, 1978–92, ages 2–14+.
  - GEAC Groundfish Enterprise Allocation Council, 1998–2005, ages 2–13. Could not get a 14+ index for this survey.
  - SENT_GN, 1995–2017, ages 3–10.
  - SENT_LT, 1995–2017, ages 3–10.
  - RV_IO-RV_OFF, 1997–2018 (not 2006); ages 2–14+. Note: age 1's not used.
- Age 1 M=0.5, age 2 M=0.3, and M=0.2 for other ages. M constant across years.
- Stock weights from a simple model.
- Raw catch weights. Catch weight-at-age in 2018 assumed equal to 2017.

Many model formulations were presented and compared in terms of AIC (Akaike information criterion)/BIC (Bayesian information criterion) and model results. Retrospective diagnostics were presented for some model formulations. The patterns in the retrospective diagnostics were large enough to be a concern. The model formulations had trouble finding a good fit to the RV and sentinel indices. An acceptable assessment model formulation was not found, and it was concluded that further research on assessment inputs is needed before a reliable assessment model formulation could be recommended.

Discussion

One participant asked if it was possible for a negative index value to be entered into the model. The response was yes because the model never takes a log of the index value. However, the predicted index will still be positive. If you log-transform the estimator, the model is sensitive to small estimates and less so to large estimates, as the log-transformation turns very small estimates into very large negative numbers and makes large estimates into smaller numbers. The approach being used here is more intuitive, small estimates have little influence, while large estimates have a larger impact (similar to using a gamma distribution).

It was suggested by one participant that different data sources (e.g., numbers-at-age and weights-at-age, and landings) should have linked likelihoods as they were not totally independent of one another which was in contrast to the presenter's modeling philosophy. The presenter suggested in this type of situation, catch estimates could be increased to account for the associated uncertainty.

A participant asked if the standard deviations from a stratified analysis had been compared to those estimated by the model for each survey. This has not been done, but the presenter agreed that it would be an informative comparison.

The presenter was asked for clarification about how M is treated in the model. For the most part, it was assumed that M = 0.2, except for ages 1 and 2. It was noted that M is mostly a scalar, while is catchability is allowed to be free, without bounds. With the use of process errors, it was unclear whether changing the scalar value of M would have a significant influence on model results. The presenter suggested that if M varied over time, it would have a greater impact. The tagging data and the condition-based M work suggested that M >0.2. Participants noted that the model formulation with increased M provided better fit to the RV survey data. Increasing the value to M = 0.4 was discussed by survey participants. The meeting discussed if changing M to 0.4 would change the scalar value of F. It was decided that this likely would not be the case if catchabilities were aggregated, and improvement in the model is possible. It would be useful to examine this scenario to see if the fit is about the same. If so that would suggest that the available data are uninformative about M.

Changes made to the error structure within the model were discussed. The presenter clarified that the error structure was adapted to address the zero-inflated data. The model's diagnostics indicated that the changes were appropriate. However, for small, predicted values, the distribution of predicted values does not go below zero which may result in unrealistic confidence intervals.

As a follow-up question, it was asked if the model is being fit in real space would this affect how much weight would be given to a high number? Fitting the model in real space meant using non-transformed data. The presenter noted that this is an ongoing challenge for this kind of modeling; it is not ideal that the model is influenced heavily by high index values, however the alternative log approach was also rejected due to the unrealistic influence of small index values. It was agreed that it may be preferable to fit raw (untransformed) data. This may be investigated further in future work but was not possible at the time of this meeting.

One participant commented that some of the presented model formulations appear to be compensating for data and/or index issues, rather than statistical limitations of the model. It was acknowledged that this development of an appropriate assessment model is particularly challenging for this stock, due to diverging signals from the different survey inputs. The presenter suggested that one solution may be to move to a spatial analysis.

A participant commented that the current model formulation allows the landings estimates to be predicted close to the lower bounds, even in periods that are data poor and/or subject to higher uncertainty. The participant suggested a model formulation that would force the landings estimates away from the lower bounds during periods of high uncertainty, which could be considered more realistic estimates. The presenter clarified that under the current model formulation, there were no explicit preferences about where the landings estimates could fall within the bounds, and therefore the landings were equally likely anywhere between the lower and upper bounds. When asked if he could recommend any of the presented model formulations, the presenter was very clear that he did not find any to be appropriate as an assessment as this time, citing the ongoing issues estimating M, specifically the lack of age-based M estimates.

In terms of model performance, the presenter did highlight the model run using auto-correlated process errors had the best retrospective diagnostics and is effectively treating M as age-based, which may be a promising direction for further work. The chair asked if they had a preferred model now, as at the pre-framework meeting. The response was that there was no longer a preferred model due to unrealistic retrospective diagnostics.

The discussion about process errors then shifted into year effects. For example, the meeting discussed possible differences in the survey in 2013 and impacts on the model results. Standardized Proportions by Age across Years (SPAY) plots indicated a shift in the index for 8 year-olds in 2013 and 2014 but specific identification of the source of these changes would require investigation of the data that was beyond the time and scope of this meeting. Detailed model results and data for all recent year classes indicated that the indices are declining more rapidly at older ages than the model is able to predict. This suggests that the data sets are in conflict, in this case the sentinel line trawl and the commercial catch are presenting different signals.

This prompted further discussion on how different model formulations fit the diverging trends in each input dataset. For example, model formulation, M19 showed a good fit to the sentinel line trawl (coefficient of variation of 0.4), however this model was not able to fit the gillnet data (coefficient of variation of 0.8), nor the French survey (CV ~0.5). In general, all models will struggle with the GEAC survey and the gillnet survey, however removing these inputs had little effect as they are not given much weight in the model fit. Across model formulations, the line trawl data was consistently better fit. However, a participant expressed concern about consistently bad model fits to the RV inshore/offshore, as that survey has the best coverage of the stock and consistency over time. The presenter agreed, stating that the RV survey has the best potential of status of the stock as a whole. However, they noted that most of the fish migrate inshore to feed in the summer and may be available to the sentinel survey at that time, however the sentinel indices appear to be affected by changes in catchability over time.

Another participant suggested that sentinel data should not be included in the modeling due to large changes in weights-at-age, suggesting that catchability of the fixed gear has changed dramatically over time which is not being captured in the mode. Further concern was raised over inclusion of the sentinel gillnet index specifically, which is based on fishing a very small portion of the stock area. It was also concerning that the spatial extent of the sentinel survey has changed over time. The presenter clarified that the line trawl data shows a broad selectivity pattern, and may be less sensitive to changes, but agreed that the gillnet would be sensitive to changes in size-at-age.

One participant asserted that none of the surveys were fully tracking the stock, including the RV survey, citing fluctuations in the RV survey index that they proposed were not realistic. A participant suggested that this perception may be due to expectations that the survey should track catch, rather than stock status. This contributor emphasized that the DFO RV survey is a very good index for stock dynamics in 3Ps but trying to relate it to the structure of the catch seems to be a problem. This breakdown may be due to how the catch relates to the stock assessment model or the lack of coherency of the survey indices over time, which limits our ability to track strong cohorts. Although some cohorts can be tracked over time through the RV, sentinel, and catch data, there is not a direct match in cohort pattern across data sources. For example, sentinel data shows cohorts that start high, decline rapidly, and remain low, which is in contrast to all of the other indices. The contradictory patterns cause problems in fitting the model. The presenter disagreed with the idea of picking and choosing surveys, clarifying that all data sources had strengths and limitations.

The RV survey showed relatively strong year effects that should be mitigated by providing additional data from other indices to the model. However, they acknowledged that in this case, the available data may not be sufficient to achieve this goal. Further, they did not expect that removal of the sentinel data would have a significant effect on model performance. The presenter recommended keeping the sentinel survey in the analyses and thought that the work could be open to criticism at a later point if an excluded survey showed diverging trends. Another participant disagreed, however, and questioned including indices that do not effectively track the stock due to changes in catchability over time. They suggested that either the model must be adapted to address the changing catchability over time, or the sentinel survey should be excluded as it is not meeting the fundamental assumption of constant catchability. The presenter noted that it may be possible to address these modeling issues and further noted that the sentinel survey is able to track some year classes (both the gillnet and the line trawl surveys). However, they also suggested that if the peer review and meeting conclusions found that the sentinel survey was not appropriate for inclusion in the assessment model, that the survey program should be ended.

A participant argued against the exclusion of the sentinel data series, expressing the opinion that all available information should be used. This participant suggested that effort should be focused on better understanding the biology of the stock, specifically M and seasonal movements.

The meeting did not reach consensus on the appropriateness of the sentinel data for inclusion in a stock assessment model for 3Ps cod, however it was agreed that the meeting would look at results from a run of the state-space model presented by D. Varkey (HYBRID) including only the RV data.

SAM STYLE MODEL FOR 3PS COD WITH/WITHOUT POST-STRATIFICATION

Presenter: J. Champagnat

Abstract

A stock assessment model (called SAM) including commercial catches and survey indices has been developed for the 3Ps cod stock. This model developed by France was among three models developed in a collaborative project between Canada and France in 2018 and 2019. It allows for process and observation errors, and uses random walks in the mortality rate to estimate a time-variant selectivity. The model provides estimation for stock size, F, recruitment, as well as a stochastic forecast. The objective of this model development was to achieve an accurate estimation of the stock dynamics e.g., consistent with expert knowledge of the stock and with as little bias and variance as possible. Four runs of the model were presented in the framework meeting with different options on mortality and surveys inputs. Runs produced comparable estimation but lead to different measures of stock status in the recent period. Runs also performed differently in terms of retrospective diagnostics, residual patterns, and description of latent processes.

Discussion

A participant suggested that it would be useful to see the predicted landings from SAM plotted against the catch bounds. It was noted that the predicted landings were never higher than the observed landings. Upon further observation, it appeared that the model predictions were inconsistent with the catch bounds.

A contributor noted that the model run using condition-based M showed best fit to the RV survey. They suggested because the RV survey is considered the most reliable reflection of the stock status, the goodness-of-fit with this time series (represented by the standard deviation parameter for the inshore/offshore survey) should be included in model comparison summary tables.

The meeting revisited previous discussions on the condition-based M estimates. Citing recent increases in M and inconsistencies between M estimated from tagging data and model predictions, one participant suggested that there might be a non-linear relationship between M and condition. Time-varying and condition-based M seemed to improve the fit of the model by taking variance out of the process error and placing it on M. The predicted spike in recruitment may indicate a need for better correlates for M rather than a problem with the model itself. The author who presented on condition-based M earlier in the meeting, suggested that it might be possible to better estimate condition-based M by disaggregating age-classes which in turn may reduce the number of recruits the model has to generate.

It was noted that the model is struggling to resolve the conflict between the line trawl and the RV survey, and is ending up finding the middle ground between the two series. It’s unclear as to whether the predicted recruits are in the system or not. They do not appear in the sentinel survey; however, it is uncertain if that is due to spatial or selectivity issues. One solution may be development of an age-disaggregated condition-M estimate, but it’s unclear as to how long this may take.

Concern was raised that if M was age-disaggregated, that the model would ignore the RV survey at the young ages when the sentinel line trawl is included. Unlike the RV survey, the line trawl survey isn't designed to capture younger fish (ages 2–3). Including age-3 fish from the sentinel line trawl survey may unrealistically reduce the estimate of age-2 fish in the model compared to what the RV survey generates. Another participant commented that the commercial catch data could present the same problem. Concern was also raised about commercial catch-at-age relying on ages from the sentinel survey data, however it was clarified that due to differences in the gillnet mesh sizes, age-length keys from the sentinel survey are very rarely applied to the commercial catch data and this only occurs in years with low commercial catch when more suitable age-length keys are not available.

One of the participants expressed concern over the use of time varying natural mortalities, particularly towards the end of the time series where M = 0.8. Concern was also raised about the inability to develop projections for time-varying M without an established mechanism that drives the variation. While the researchers agreed that the mechanism remains uncertain, it is clear from the data available that mortality is varying over time and a lot of fish in this stock are in poor condition. Experts agreed that these trends are likely related to the prey base. It was suggested that the acoustic data from the spring survey should be used to estimate abundance of small pelagic fish to refine M. Practical limitations impeded this. For example, there is currently no way to verify the spring survey acoustic data, and there was neither time nor capacity available to complete work of this scope by the end of the year.

It was argued that the high values of M being seen in these models are reasonable. In other regions similar M's have recently been accepted: NAFO Division (Div.) 4X5Y – terminal M (age 5/6) = 1.5, Southern Gulf Cod terminal M = 0.81 to 0.85, 4S3Pn M = 0.7. Despite the uncertainties discussed above, it was suggested that high M values were believable, based on the biology of these populations, and known ecosystem pressures (i.e., predation by seals, availability of prey).

A meeting participant asked for clarification on the drivers of high M estimates in the other assessments. In Div. 4X5Y they were estimated by a virtual population assessment model with a random walk function on M. The catch and landings data in this system are considered to be reliable, largely because of observer coverage. In the other cases, the assessment model produces an estimate of M that is applied as a constant, with the exception of the Northern Gulf of St. Lawrence where they use an iterative approach. Notably, 3Ps seems to be the only stock where scientists are attempting to develop a time series for M estimates.

The importance of the conversion from condition factor to M was also raised by a reviewer who noted that there were very few years >10% of fish were observed with condition below the threshold condition of 0.85, and estimated that only 12–15% of fish being affected in the worst years, yet this estimate is then translated into M as high as 0.9 where half the fish are being affected. The reviewer expressed concern that the conversion was too sensitive, but agreed that it is better than applying a constant value for M.

It was pointed out that the uncertainty from condition-based M-values had not been considered by the meeting; specifically, that the uncertainty in the M values was higher in more recent years. A participant suggested that it might be useful to use the lower bounds on the M estimates to see if that helped with the recruitment issues. The author of the condition-based M work clarified that the proportion of cod with condition below the threshold in some of the years reached 20% and that level was sustained for several months. As these condition observations accumulate over time, the proportion of individuals dying of starvation add up. It was asked how long it took for the fish to die in the Dutil and Lambert (2000) study. The details of the experiment were discussed but it was unclear if there was a running total across time (of the number of fish dying) or if there was only a final number of dead fish provided. It was suggested that condition-based M be recalculated from an index of poor condition days.

A participant suggested using the tagging data to inform on M. While it would be difficult to get a stock-wide F from tagging because of the spatial heterogeneity of these data, a stock-wide M might be estimated if all fish in the Region are assumed to be exposed to the same natural mortality. Based on the correlation with the M estimates for Northern cod, derived from NCAM, it was suggested that a broad-scale M is a reasonable assumption. However, unlike condition data, which can provide an M estimate for a single year of data, multiple years of tagging are required to generate an M estimate. An option brought forward was to compare the 20 years of tagging data with the indices presented for condition-based M, effectively ground-truthing the condition index. If a relationship can be identified, it may then be possible to extrapolate back the time series of M with the tagging data. However, this would require work beyond the available time and scope of this meeting.

One compared the approaches to estimating M, describing the tagging-based M as a direct estimate, whereas the condition-based M is an inference. A NCAM approach to estimate M was suggested. Experts explained that exploring M in 3Ps using a process similar to what is done in NCAM would be difficult. This is due to uncertainty about stock mixing (for example with Div. 3L and the Gulf). It was suggested that if migration does not have a clear trend, but there seems to be a strong trend in M, the model driven by that M would work, and additional uncertainty would go onto error. This is the type of question that has driven recent work to use the tagging data to investigate spatial dynamics for this stock.

It was noted that the condition-based M work was also inspired by findings from the Northern cod stock. At the March 2019 Northern cod meeting, research was presented that showed correlations between M and condition. When this work was adapted to 3Ps cod, a seasonal model was required due to the difference in spring and fall data for these two stocks.

Clarification was requested about the accumulation of M; although there is a high rate of lethal conditions observed in spring, M is expected to be low for the rest of the year. The presenter explained when following the methods described by Casini et al. (2016), the high rates of poor condition (i.e., threshold of Fulton's K $\leq$0.65) in April and May were enough to generate cumulative estimates of M = 0.6 within a single year. However, further experimental work would be required to determine how many days the fish survived once they hit that starvation level. It was also noted that the tagging data provide historic M, but not recent M. It was unclear how spatially-varying F would be integrated into a spatially aggregated assessment. Further, one participant noted that a spatially explicit assessment would also be challenging as it remains uncertain whether the relatively low rate of tag returns from the offshore region of 3Ps is due to low F, fish movement, tagging mortality, or other unknown factors.

The chair noted that there seemed to be agreement that constant M was not ideal. They asked for further feedback on J. Champagnat's presented SAM model.

A reviewer asked if the SAM model struggled with sudden changes in F during the moratorium when fishing dropped quickly and then rose again. The response was that the model did seem to struggle with the high variance in F and the high peak in average F in 1992, just before the moratorium is probably an artifact.

The chair summarized the discussion noting that while the SAM model wasn't as customized as the SSM model, it still performs fairly well, depending on the data used.

The reviewer followed up noting that several varying M options were explored and asked for some comments on the different options tested. Six different methods of estimating M in the model were attempted: Time-blocks (tested if recent M was higher than pre-moratorium times), Lorenzen (1996; based on weights-at-age), Brodzlack et al. (2011; based on the ratio of length-at-age vs. length-at-maturity plus a baseline of 0.2), Brodzlack et al. (2011), Casini et al. (2016; similar to Brodzlack et al. [2011], but with a time-varying baseline), Casini et al. (2011), Casini et al. (2011; but with age groups). The reviewer followed-up by asking if the same M was used across all ages and to which length range it corresponded. For the age-disaggregated one, it was noted that it was an initial attempt, and it wasn't working properly as it was sensitive to the bin widths of the length groupings and the age groupings which results in the magnitudes of M being affected incorrectly, and thus this method was rejected.

HYBRID MODEL FOR 3PS COD

Presenter: D. Varkey

Abstract

This paper describes the development of a state-space model (HYBRID) for the stock assessment of the 3Ps cod stock. The survey-based assessment model (SURBA) previously used for the assessment of the stock (from 2009–18) fit data obtained from the RV survey only. The SURBA model does not fit to catch data (i.e., commercial or sentinel catch). The HYBRID model fits to the RV survey (1983–2005, 2007–19) as well as the following additional survey time series: the IFREMER ERHAPS survey (1978–91), the GEAC survey (1998–2005), and the sentinel gillnet and line trawl surveys (1995–2018).

The HYBRID model also fits to fisheries data with the expectation that the model can improve upon SURBA in separating F from M. Fisheries catch-at-age is fit using continuation ratio logits, and the fisheries landings are fit via censored likelihood. The use of censored likelihood for fitting landings allows the inclusion of expert opinion on reliability of landings throughout the model time series. The start year for the SURBA model was 1983. The HYBRID model presented here starts in 1959, which is the first year for which landings data are available.

The HYBRID model allows exploration of different forms for parameterization of time-varying fisheries selectivity, M, and approaches for estimation of catch-at-age data. The model formulations also differ depending on the survey series used in fitting the model. Seventeen model formulations are presented. Alternate formulations based on F included:

1. a logistic selectivity,
2. SAM style multivariate normal random walk,
3. a separable age-year correlated process for F, and
4. breaks in F pattern for the moratorium.

Alternate formulation based on M included:

1. invariant M across age and years, and
2. time-varying M related to fish condition.

The different model formulations are compared and evaluated on the basis of AIC values and their performance in retrospective diagnostics.

Discussion

A participant asked for a more detailed explanation on how condition-M works in the model. The presenter noted that the condition-M is being treated as an index rather than as an actual value for M. The equation used here is based on Kumar et al. (2013).

A participant commented that it was a big step to unlink the surveys for ages 2–7 based on the rationale that there has been a change in the spatial distribution of the stock at those ages. The explanation was that the change appears to be behavior driven, with fish selecting habitat based on temperature rather than depth. The fish distribute over the bank more in warmer years. For age 8+, fish are located in more offshore waters and the spatial shift doesn't seem to happen as much, thus adjusting the index for the age 8+ isn't a large leap of faith.

The coupling of 8+ in the SSM model was discussed in comparison. There, constant catchabilities were used for age 8+, but this is different from what is being done in the HYBRID model where the time series is being broken for ages 2–7 with different catchabilities, when offshore is being compared to inshore/offshore.

A more detailed description of fish behavior followed. Looking at the distribution of ages 8+ fish in the survey over the entire time series, most of the fish are located in the offshore area; particularly in the outer channel area and around Burgeo bank. A few fish are found in the inshore area occasionally. The habitat associations of 3Ps cod was examined using the full inshore-offshore time series.

A sharp change was noticed in the median habitat occupied in 1998, which was observed in the offshore data as well. For large fish, there was a change in thermal habitat occupied from 1998 to 2000, after which the fish moved to the range of temperatures seen in the past. For medium size fish, from 1998–99 they occupied the available temperatures more closely than the older fish. Temperatures occupied by small fish were not different from the available thermal habitat. Recent increases in the proportion of small and medium sized fish relative to larger fish drove the habitat associations observed (~1.5–~3.5°C).

Increases in the proportion of 3Ps covered by warm bottom temperatures may be driving cod aggregations, resulting in some of the large trawl catches observed. A participant commented that in linking the catchabilities across all ages for the inshore/offshore and the offshore only going up to 1996, the assumption is that the relationship between the fish in the inshore and the offshore is the same all the time. Also, looking at the temperature available versus the temperature occupied indicates there has been a change in the behavior of the fish and these differences may vary across sizes. When they attempted runs in SAM and the HYBRID model with unlinked data sets, they ended up with poor fits and bad retrospective diagnostics. Given that most of the fish ages 8+ were in the offshore, they decided to go forward with models where only the age 8+ were linked and the impact of the change in the offshore and the inshore would be minimal. This would also remove the linkage for the younger ages where impact of this change may be more pronounced.

A discussion on the rationale to split the data series occurred. In response, it was suggested to look at the differences in catchabilities as a result of the split. The catchabilities for the younger ages were higher after the split, but not by much. It was clarified that they were not actually splitting a survey, one survey included the offshore and the other survey did not. It was previously assumed that a constant relationship between the inshore and the offshore in the SURBA because they had to go back to 1983 and this was the only survey available. It was decided that the catchabilities were being split rather than the data series.

A participant commented that the HYBRID model fit the catch-at-age compositions less well than the SSM model, which might be positive as they thought it had been over fitted in the SSM model. This seemed to have been done through independent index errors and correlated continuation ratio logit (CRL) errors, which have the effect of down-weighting the CRLs relative to the surveys. It was noted that there were year effects in the bottom trawl surveys but not in the sentinel. This was because when the sentinel surveys are used, there are some year effects in the standardization process. The participant disagreed with the rational of not having year effects for the sentinel survey as standardization also happened for the RV survey.

Another participant suggested the difference in standard errors around the CRLs in the SSM and HYBRID models is the use of a standard deviation being estimated for the 1994 to 1997 period, in the SSM model, whereas there was more spread for those years in the HYBRID model. This was supported.

One participant noted that there was pattern in the retrospective diagnostics for total mortality rate (Z) but no retrospective patterns in spawning stock biomass or recruitment estimates (highlighted by the small Mohn’s rho values, a measure of the severity of retrospective diagnostic patterns). The participant wondered where the change in the numbers-at-age was coming from to match the change in the Z’s and the presenter hypothesized that it could be driven by the year effect estimate. In a follow-up comment, the participant noted that the population estimates remained similar for the various model formulations, but the mortality rates changed, and a look at fishing mortality (F) retrospective diagnostics could potentially provide more clarity on the root of the change. The participant wondered if both F and population size had to change together under the model formulation. Another possibility raised was that the scalar parameter for natural mortality (M) might be changing under various model formulations. The presenter reminded the group that there are acknowledged issues with the estimation of M and suggested looking at the closest model without M’s (model 16) as M is invariant in that formulation. When looking at model 16, the differences appear to be driven by ages 2–5, however for ages 6+ there seems to be similarity in estimation across the different model structures. A participant argued that this may be an indication that there is very little information for M at the younger ages and the model does not have enough information to distinguish between M or the year effect. A participant noted that it was puzzling how the model had the same recruitment with different Z’s and was getting the same spawning stock biomass. After further discussion, it was suggested that the process error might be masking the retrospective error.

The presenter was given time to look at the model in more detail to provide an explanation, as it was important for determining the acceptability of the model.

DISCUSSION ON MODEL FORMULATIONS

A participant asserted that before it can be determined if the meeting has accepted a model for the 3Ps cod stock assessment, the meeting needs to decide if condition-based mortality will be included in the model formulation. It has not been decided how to estimate time-varying mortality for each of the model structures.

It was suggested that condition-based M could be included as a scalar index given that some meeting participants were not comfortable using the raw calculations. It was also noted that other components of M, such as predation, may (or may not) be related to condition-based M. Participants agreed that while they saw problems with using condition-based M, constant M was unrealistic. There was general discussion about the recent high estimates of condition-based M and how it was estimated. It was suggested that laboratory experiments could overestimate mortality through extreme of starvation being imposed on cod. The argument that condition-based mortality could be overestimated based on laboratory experiments was immediately countered by participants suggesting that it is equally plausible that it could be an underestimate, due to stressors on wild populations such as predation effects and the diminished ability of weakened cod to capture prey. Another participant noted that based on past laboratory experiments, it takes a long time for cod to die of starvation (Dutil and Lambert; 2000). They thought scalar condition-based M could be useful rather than directly using the condition-based M as we do not know how to accumulate M over the year (i.e., it is not an instantaneous process). It was suggested that the assessment could use a scalar estimation of M, based on condition information and/or tagging data. There seemed to be agreement that there was potential for the work on condition-based M, and that a time invariant M was the incorrect option.

The chair noted that we do not need exact numbers at this meeting, the purpose is to have a way forward to be evaluated further at the 3Ps stock assessment in November. There seemed to be agreement that the method developed had potential, but it needed work based on feedback from the meeting before being presented at the assessment. For example, it was noted that at this point, the retrospective diagnostics were not understood. If this issue was resolved, participants would be satisfied by the HYBRID model. It was further suggested that the twenty year time series of tagging data maybe able to provide an estimate of M and this should be tested to see if that resolves some of the retrospective problems. It may be possible to determine the proper scaling of the condition-M time series using the tagging data.

A participant brought up the concern of spatial differences in the tagging over the period and the low return rates for tags. Uncertainty generated from these issues may impede using the tagging data for estimation of M in the assessment model. It was suggested that removal of the offshore tagging experiments from the analysis may be a solution. This would create a gap in the time series in the mid-2000s because there were few inshore tagging experiments then. Due to the fixed effect on F by the tagging sub-region that is applied in the model, it is implicitly assumed that disappearances of fish are part of F. However, as it is known that fishing on the offshore fish is very low. The participant mentioned that uncontrollable issues like tag loss, tagging mortality, etc. are incorrectly assumed to be part of F.

There was some discussion of specific SSM runs and how to improve model performance. Participants suggested: splitting catchability estimates for young fish, modeling catchabilities by age, and further examination of the temporal autocorrelation in process error to better inform interpretation of model projections. It was suggested that the process error is reflecting problems with the surveys being used.

If the survey data were comparable year to year, it would be expected that the auto-correlation of the time varying process errors would be positive, however the model is estimating almost no auto-correlation in time process error, but a lot of auto-correlation in age process error. The model may be adapting to very noisy early survey years through the auto-correlation in process error, which may not be estimable. As a solution, a fixed auto-correlation in process error of 0.5 was imposed for some preliminary runs. It was noted that this was similar to previous work on stock assessment models in Iceland, however none of the model formulations discussed here were recommended by the author to bring forward for stock assessment.

A meeting participant commented that there were faults with all of the models reviewed, but that they are providing the same advice. They asked if it was part of the Terms of Reference for the meeting to examine the relative performance of these models with respect to providing advice. He suggested that it might be worth considering using multiple models for 3Ps in an ensemble approach, rather than trying to identify a single best model. The chair responded that the Terms of Reference did not specify and rather dictated just that “we are to provide direction for an approach for estimating reference points.” A multiple model approach was discussed prior to the meeting and not pursued; however, this decision could be revisited. However, the outcome of that discussion had been that there could be a bias towards choosing the model that gives the most optimistic projection. The participant responded that there are approaches to objectively combine model projections and remove the subjectivity in how multiple models are used for management.

It was asked how the models would be compared. There are many possible approaches for setting Limit Reference Points ([LRPs] such at the Biomass Limit Reference Point [BLim]), for example, BLim can be based on the lowest spawning stock biomass of the time series from which the stock demonstrates recovery. A participant suggested that an exploitation rate based reference point is preferred for an assessment that relies on ensemble modelling since spawning stock biomass/BLim could be inconsistent across models. Another participant suggested initially comparing based on diagnostics first but noted this will be difficult as it may not be possible to do a comparison with multiple models.

A participant advocated for the use of multiple models. They noted that over the last 15 years, statistics has moved away from searching for the single best model, which was essential when models took weeks or months to run. There are several model averaging approaches that could be useful.

A challenge with using these “ensemble approaches” with the models is that averaging approaches assume models are developed independently. The models developed here share commonalities in their frameworks, and so averaging them would lead to false precision.

A reviewer noted that the multiple modeling approach has been difficult to implement elsewhere (in Iceland). They expressed a preference to present concrete advice rather than having to use an algorithm to explain the decision. One way around may be to try a Management Strategy Evaluation (MSE) that considers the model’s problems when giving the advice. The chair noted this approach would require another meeting.

Another participant suggested that the SAM model seemed reasonable enough and wondered if there was sufficient payoff from the increasing complexity of the SSM and HYBRID models. A participant commented that using the simplest statistical catch-at-age model likely would not work due to the complexity of the indices (i.e., variety of survey inputs).

The participant asked the developer of the HYBRID model what the main challenges were when trying to fit their model to the data and what improvements could be made to the model. The response was that there were conflicts between the different indices (e.g., the RV and sentinel surveys, particularly for the younger ages). The model is estimating F’s for age 2s, removing this could improve the model. There may be some possibilities for improvement with the models, without the M estimation, that are using the sentinel and RV surveys. Some of the problems may be coming from estimating M and F. The participant asked for additional information on the retrospective patterns in Z. They may be from the process errors, it is unclear as to how large it is, whether it is directional and how large are the retrospective error standard deviations. The size of the standard deviation of the retrospective error needs to be determined in order to decide if it can be dealt with or not.

The participant noted that in the other models there were retrospective errors for spawning stock biomass, F, and recruitment, but not for the Z. The response was that they could consider the retrospectives of F, as that is what is being seen in the other models. They could also investigate process error retrospective diagnostics as well. It was noted that the retrospective diagnostics for spawning stock biomass and recruits in the HYBRID model are good, but the retrospectives for Z are not. If stock status can be given with reasonable certainty, despite problems with Z, then it may be sufficient. Another participant noted that the model has to be projected and that is where the retrospective errors become a problem.

There was some discussion about which variation of the HYBRID model should be looked at given the limited time available to prepare the models for discussion. Seventeen different versions of the HYBRID model were presented.

It was suggested that a table comparing the best model from each developer be produced. The table will include spawning stock biomass with uncertainty, recruits, F, and M. A number of questions were raised about how many models from each developer should be provided, which datasets should be included, and differences in model configuration. One of the model developers noted that you cannot select models based on their outputs. The developer said that the models need to be picked based on how well they fit the data. It was noted this works within a modeling framework, but not across modeling frameworks. A participant suggested using cross validation, but that this would not be ready for the following day. It was suggested that using the same data, they might be able to look at Bayesian Information Criterion (BIC) which are approximations of cross-validation. It was also suggested that residual plots be examined. Furthermore, the rationale for selecting a preferred model needs to be provided (e.g., what diagnostics were looked at, what does the preferred model do better than the models not selected). A developer raised a number of potential issues with the residuals from SAM and suggested that observed vs. predicted plots be prepared as well, possibly by cohort. It was clarified that year effects would not be included in the models to be examined the next day. It was noted that in the SSM model, there were auto-correlated errors, which is different from year effects, but might be considered year effects. A participant agreed with the concern about focusing solely on the model outputs, but also noted that the meeting had been looking at model fit as well. They pointed out that a number of models had already been eliminated from consideration over retrospective errors, the developers have been told the group's concerns about the different models, and they encouraged the developers to keep in mind the feedback they had received about the group's preferred model from each developer. They suggested polishing the model version that had withstood criticism the best and they proposed that it was time to bring the three models together to be compared.

The chair agreed with these comments and asked the developers to share their best model. A developer asked for clarification on what was being requested for the following day. It was suggested that the meeting wanted to see an overlay of the outputs: spawning stock biomass, recruitment, and the average F. There needed to be clarification on what F was needed. Diagnostic plots were also requested: the fit to the survey so long as there was not a year effect in the survey, observed and predicted. It was suggested that an overlay plot be produced for observed and predicted. A comparison of the process errors between the models was also requested, but it was noted that this would be a bit more difficult to do, so the developers would have to agree how to do it. There was additional discussion as to whether there should only be one model from each framework considered, or if multiple models per framework could be considered. From the SAM model, the base + sentinel model was suggested. From the HYBRID model, Model 11 was suggested. From the SSM model, Model 20 was suggested. The SSM developer expressed some concerns about Model 20 and expressed plans to bring a new version of the model that considered some of the earlier discussion from the meeting. It was asked over which ages average F should be provided. Ages 5–8 were selected. Another participant requested that the developers identify what was good about each model, what was bad, and what is not understood. An email was circulated to model developers summarizing the requests from the meeting.

The chair began the third day of the meeting by noting that there had been a turn of events and invited the developer of the SSM model to speak. They announced that they had run the wrong retrospective diagnostics in preparation for the day's efforts. They went on to say that they thought the decision point for the meeting is what to do with M and none of the models they are presenting directly address M and for that reason they felt that the meeting should focus on the models that do. For that reason, they withdrew their model from consideration.

The chair stated that the meeting was going to take a look at HYBRID model M11 and a SAM model and compare the output from the models and their diagnostics side by side to determine if there was clear evidence that one model was preferred for 3Ps cod. Following such an assessment, three scenarios are possible: conclude that one model is better, use one model and continue to work on the other, or neither of the models are acceptable.

A participant noted that the alternative option is to use the ICES model where they could look at model trends. The chair agreed that this also is an option.

A participant asked if the meeting would make a direct comparison of the assumptions going into the models. It was agreed that that this would be done.

Another participant asked if the option of using both models in an ensemble approach was still on the table. The chair stated that the option was still available.

A participant asked if both models are rejected, did this mean that the SURBA model was the default option? They were concerned, as evidence was presented that there has been a change in catchability which would mean the SURBA model is no longer appropriate for the stock, and it should not be the default option. Another participant noted that they were close to rejecting the SURBA model at the last assessment. A participant suggested that there be a comment saying that SURBA is no longer a reliable option. This suggestion received support from others.

For the model comparison, two versions of the SAM model were examined (the base model + sentinel, and the model that uses the raw condition-based M’s + sentinel) and one version of the HYBRID model which had the scalar on the condition-based M’s. The outputs examined were F, recruits and spawning stock biomass followed by the model diagnostics. It was noted that there was a fairly significant difference in the recruitment patterns, particularly in the periods of the 1980s and from 2005 onward.

Some meeting participants expressed confusion over the reason why a SAM run without scalar M was being compared to a HYBRID model run with scalar-M. The response was that this was done to see how a model that included a changing M compared to a model that did not.

Multiple participants asked for additional detail on how M was implemented in the SAM and HYBRID models. A participant asked if the differences between the two models could be summarized as: one model (SAM) tries to capture all the model misspecification via process error while the other model (HYBRID model) attempts to model M and includes process error to capture any missing processes not captured. The developer of the HYBRID model generally agreed with this description but noted that there are several other subtle differences in the two models. It was noted that there needed to be strong evidence of the M-trend relationship. If this evidence was absent, this approach is likely to receive heavy criticism.

The chair noted that there was discomfort with the condition-M approach and suggested that the alternative approach being tried in the HYBRID model was an alternative and suggested that the meeting look at the diagnostics and process errors and go from there. A participant noted that there was an intention to provide the models with additional data going forward (e.g., mortality rates from the cod tagging data) and so the absolute numbers being seen today would not be correct nor would the examination of the shapes of the mortality. It was suggested that the meeting needed to see how the models would respond to the additional data being provided.

The diagnostics and errors of the two models were then presented for comparison. The first comparison was the fit of the models to the survey data. It was noted that the HYBRID model has more age classes and that the scales differed in the plots. A participant noted that these outputs were not comparable.

The Chair asked if the SAM model was offering anything that the HYBRID model does not. It was suggested that it gives an objective way of projecting M wrapped up in the process error, which may be a practicality rather than best practice. For the stock assessment, it allows the group to produce M for 2021 and 2022. It was argued that SAM does not provide a way to project process error as it is an independent and identity distributed (IID) process error, the best projection of process error for the next year is 0 and will always be 0. It was noted that the built in SAM projection procedures would not be able to project forward the recent low process errors, a procedure would have to be developed for this. It was asked if the HYBRID model would have correlated process errors projected forward. This had not been done yet but could be. It was noted that SAM's projection process had been peer reviewed whereas the HYBRID model still needs to be peer reviewed at a framework meeting or a stock assessment meeting.

A participant requested clarification as the recent discussion had been criticizing SAM because it does not have variable M, however earlier in the meeting there was a plot from SAM shown with variable M. The response was that the variable M in SAM was produced in a way that was not thought to be correct. The participant noted that this was an issue with the variable M in the HYBRID model. The response to this was that presumably it would be possible to do variable M in the HYBRID model in a way that the meeting thought was correct soon. It was suggested that this could probably also be done in the SAM model using a macro. However, uncertainty around the M scalar would not be accounted for in the SAM model.

It was asked if the meeting wanted to use scalar M. It was suggested that the meeting was drifting in that direction with some participants suggesting that they could not accept a modeling platform that would not be able to include variable M in the near future. It was argued that both models should be able to deal with variable M in the near future.

A participant asked if the SAM and HYBRID models were basically the same model with some minor differences. It was argued that with tight catch bounds, then yes they are very similar, but with wider catch bounds, they would be quite different models. It was pointed out that SAM does not include the ability for a user to declare how precise catch is. This is an advantage of the HYBRID model over SAM. A reviewer mentioned the example of a SAM model for the North Sea cod where errors in the catches were modeled explicitly using catch multipliers. It was argued that this was a different scenario as SAM requires a period of time where there was very good catch, but that does not exist in the 3Ps scenario. It was attempted with a variant of the SAM model used here but struggled with the really high levels of catch in the past. It was suggested that it may be worth looking at the two models when the assumptions are the same. Another participant suggested that at the minimum, there should be a future work recommendation that the HYBRID model undergo a simulation self test given that it has not been peer reviewed. The SAM model has been peer-reviewed multiple times with other stocks, so this is less of a concern.

A participant noted that they desire the possibility of M-varying in the model. Decisions have been made on what they want in M. They like the direction that the condition-based M work is going but have decided that the instantaneous M option for condition-based M is not the way to go. Further, there should be some comparisons of the condition-based M work with the tagging data analysis. There seems to be some agreement with the idea of using a scalar M, but the current implementation is not supported. Based on this, it was suggested that the meeting list the items that they want reviewed. If HYBRID M11 is the chosen model, but the things they want looked at fail, then the alternative is HYBRID M10, which is the same as M11, but lacks the M-scalar. Furthermore, the SAM model should be kept as an alternative model.

One concern raised was that the models are giving a lot of weight to the sentinel line trawl. This has been an issue in ICES in frameworks for a number of stocks where a lot of weight has been given to a series in a new model and then in the following year when the model is rerun, the results are substantially different. Having more than one model to look at in the run-up to the assessment could be useful. In response, one participant noted the meeting needed to consider the usability of the models for the current year. It was noted that SAM is ready to project while the HYBRID model will hopefully be ready. This was taken as an argument for having multiple models available. The need for wide bounds on landings still needed to be determined.

A discussion was also held on what the catch bounds should be for future model runs as it was argued that the initial catch bounds might be too tight and in other cases they might be too wide. The initial catch bounds used were:

- 1959–93: 0.5–1.5 multipliers
- 1994–96: 0.9–1.1
- 1997–99: 0.35–1.75
- 2000+: 0.75–1.25

It was argued that some of the bounds were consistent with D. Ings and E. Carruthers work (unpublished data). It was asked if the bounds should be wider in the earliest time period. It was asked what the logic was for the wide catch bounds in the earliest time as there was no Total Allowable Catch (TAC). The response was that in 1983 (earliest time period in NCAM) there were issues around foreign fleets and the area of reporting (e.g., fish may be caught in one area and reported in another). Some of this would be captured through the purchase slips. A participant also raised the issue of fish being caught in one area and being landed in another and further noted that species might be mixed in some of the landing records. With respect to the use of landing slips to generate numbers, they noted that this only addressed the fish that made it to the plant and did not cover any fish that were discarded. On this basis, the participant recommended that the bounds be wider for the early years.

Catch from personal use by fishers was raised, but it was argued that this would likely represent less than 1% of the catch. With respect to discards, it was noted that a study by Kulka and Stevenson (1986) showing 10% of small fish being discarded was from a single month, in a single year. A second participant noted, based on his discussions with fishing captains, that 3Ps cod had been landed in other areas as flatfish over a period of many years.

There was agreement that the 10% bounds were too low. It was argued by another participant that this discussion was a bit arbitrary as documentation was lacking for many aspects of the discussion (e.g., landings, discards, misreporting, etc.). Another participant asked what the rationale was for the 1.3 bounds in 1976–77. This was based on the discrepancy between the NAFO reports for that period and what was now being used for that stock. It was suggested that this might be grounds for using a 30% bound for periods of high uncertainty as 1.3 was the upper bound in a period of uncertainty. A participant argued that there were problems with both of the numbers used to produce the 1.3 and that it would likely be useful to look more closely at historical assessment documents and possibly discuss the issue more with people who participated in management at that time. One of the developers suggested running a sensitivity analysis on the bounds. It was argued that this would affect the entire time series when the concern is mostly for the earlier years. The decision was to use 1.3 as the upper catch bound from 1959 to 1977 while the lower bound would remain the same. This would be implemented for the assessment.

A participant argued that the logic for the lower bound was inconsistent as a lower, lower bound was used in a later period and the current lower bound is more precise. Another participant agreed and noted French fishing in 3Pn and on the West Coast of Newfoundland. A participant who had spoken with a French plant manager did not believe that there was a problem with misreporting of areas at that time but did suggest that there were problems with people cheating on conversion factors during that time period. A participant asked why the bounds were wider from 1997 to 2002. The bounds were largely based on the interviews that were performed with fishers. There was a concentration of effort in the Bar Haven area and lots of nets and gear were lost resulting in a lot of ghost fishing. Additionally, there were periods of high grading. When the trawlers started up after the moratorium there were indications of quotas being exceeded and dumped. It was asked why the bound changed from 9 to 10. This was based on the perception of high grading. It was decided to set the lower bound to 0.9 rather than 0.99 given the uncertainty around what was happening. This decision was extended from 1978 to 1986. The decision for the wider bound scenario at this point was:

- 1959–75: 0.9–1.3
- 1976–77: 0.7–1.3
- 1978–86: 0.9–1.3
- 1987–93: 0.5–1.5
- 1994–96: 0.99–1.05 (widened in later discussions)
- 1997 to 2009: 0.99–1.3
- 2010+: 0.99–1.1

It was suggested that for 1994 to 1996, the moratorium period, the catch bounds should be widened given the tiny catches, which make it possible for relatively small differences in the catch to have a somewhat larger effect than in other years. It was suggested that this decision be brought to the stock assessment meeting. The Chair argued that this had already been reviewed and so a decision could be made now. A participant argued that for Northern cod, the decision had been made at the stock assessment and that this decision should be made at the assessment as well. It was agreed to bring this to the stock assessment for a final decision.

The Chair noted that participants had been asked to consider the suggestion to go forward with HYBRID model 11 and a list of recommendations. As an alternative or addition, work should continue with the SAM model.

A participant noted that the inconsistencies in the retrospective diagnostics need to be understood. There is a need to understand why the inconsistencies are there and their implications.

The model developer advocated that recommended actions be attempted and a decision could be made from there. It was suggested that the inconsistencies in retrospective diagnosis was probably a result of Z vs process error, but it was unclear how it would affect projections, and this would have to be investigated. A participant suggested for the stock assessment meeting that the model developer attempt to run retrospective projections as a means of demonstrating the reliability of the assessment. Another participant noted that with the shift to time series models from more empirical models with population dynamics structures, there is a need to be able to explain the process errors to stakeholders.

The Chair then opened discussion on how the meeting participants recommended work on developing condition-based M to proceed going forward. A participant noted that in addition to taking an age disaggregated M approach, there was also a need to make an annual estimate of M, assuming the monthly estimates of M may not be appropriate. Additionally, there was concern about the use of the 0.85 condition threshold. It was suggested that a sensitivity analysis be attempted with the threshold in order to see how robust the trend is given that the proposed plan going forward is to use a scalar-M based on the trend of condition-based M. There was some discussion that the use of scalar M in the SAM model should be explored further, but there were concerns that the uncertainty estimates from this would not be carried forward in model projections.

Another participant suggested looking at condition-based M as a seasonal death rate rather than an annual one. It was suggested that the rate of improvement in condition may be faster in years with higher mortality as poor condition fish are dying rather than in years where poor condition fish are able to recover as feeding resumes. This idea may be explored. Tagging M's will be investigated further.

It was asked how these alternative approaches to producing a condition-based M will be used going forward. What would be the implication if the trends disagree? Could a trend in M be estimated if M is provided for a block of years from the tagging? This has been done with the HYBRID model where they tried a random walk with the condition-based M data. It was argued that they lacked the survey data to do this effectively. There was some discussion about how the tagging M data will be used in the SAM and HYBRID models, but an exact plan was not clear at this point. It was suggested that it might be possible to compare the age disaggregated condition based M trends to the tagging-M values as one test. They could also look for correspondence in the means. Another participant argued that this is a poor approach as condition-based M is solely mortality from poor condition while tagging-M is an estimate of M from all sources of mortality. It was argued that if starvation-induced mortality is a large component of M, then this would be a reasonable comparison. If the two mortality time series do not correspond, then the situation is simply more confused. If this work fails, it was suggested that they look at HYBRID model 10 rather than model 11. It was suggested that they use this condition-based M work as a test for how things would be different if they had time varying M. Another participant argued that as a longer term plan we should be looking at what new biological data needs to be collected and/or identify laboratory experiments that should be performed in order to gain a better understanding of condition-based M rather than only relying on existing data.

A participant suggested that for HYBRID model 11 it may be worth trying to estimate the scalar on the M in order to generate a time series of total M from the model and compare those values to the tagging M values as a check. If they are totally different, that indicates that there is a problem with how mortality is estimated. A participant suggested that this would be informative on scale but would not help for confirming the trend. It was suggested that for a new data source, when tagging occurs, the weights of the tagged fish should be collected. This would provide the tagging analysis a field-based covariate for the mortality rates that are being obtained from the tagging program. This might help to identify a breaking point for condition-based mortality in the wild.

The short term research recommendation for mortality was to go forward with HYBRID model M11 with M10 as the backup. M for model M10 is 0.3. Two suggested that it would be interesting to compare HYBRID model M11 with M10, possibly adjusting the mortality in M10 so that it approximates the average M from scalar-M trend in model M11. It was argued by the model developer that making this change would make it difficult to compare between models as an M = 0.3 in model M10 is what anchors the scale down between the two models. Changing the M = 0.3 makes it difficult to know what has caused differences as you do not know if the differences are because the scale has been changed or because of the scalar in M11.

The Chair began discussion on LRPs. It was suggested that lowest observed spawning stock biomass is a poor BLim. The model runs presented at the meeting appear to have some apparent stock-recruitment relationships. A stock-recruitment relationship should be explored for either a parametric or visual interpretation of BLim. If there is a parametric stock-recruitment relationship, the possibility of determining maximum rate of fishing mortality (FMSY) should be explored. It was asked if there was agreement on this recommendation. A participant asked if the model should be rerun with the stock-recruitment relationship being fit inside the model as it can affect everything else inside the model. It was argued that doing this is incorrect as it introduces bias in the model. Another participant argued that this is done all the time, but also saw the issues with including it in the model. For the SAM model it would be possible to estimate the stock-recruitment relationship internally. It would be possible to compare the model with and without the stock-recruitment model. If they are the same that is good, but if they are different it is problematic. It would be better to fit the stock-recruitment relationship internally in the overall model where the errors could be accounted for as compared to determining the stock-recruitment model externally and then adding it to the model. It was suggested that the first step is to examine the stock-recruitment relationship and set an interim LRP visually.

The Chair opened discussion on projection methods. A participant noted that there are two components to the projection: biologicals and uncertainty. It is known that there are trends in the weights-at-age, and the maturities related to these trends would have to be projected. This may not be possible. An alternative approach would be to use the average of the most recent 3–5 years. A decision needs to be made on this latter point. Furthermore, all of the uncertainties in the model need to be projected. The participant argued that given it is a one-year projection, the last year of the model might be the best option. The participant suggested this depends on how large a trend there is and that using 3–5 years of data might be better but in the past only 3 years of data have been used if 5 cannot be done. The Chair asked if the meeting needed to be any more prescriptive but given the Terms of Reference it was thought this should be sufficient.

Another participant asked about M. It was suggested this would have to be projected forward if M is being estimated, as would its uncertainty. If it is not being estimated, then it would just be whatever the constant is. Similarly, selectivity would have to be projected, probably the average of the last 2–3 years. It was asked if the catch-multiplier was being projected. It was argued that catch multipliers would be on the landings, and it is not a trivial projection to do. The Chair noted that projections of F would likely be requested. A participant noted that the fishing year is April 1 to March 30. It was asked how people would feel if this was ignored as it can be difficult to deal with as this requires part-year model simulations, and the model uses full year simulations. Modifying the HYBRID model to do this could be complicated. A reviewer noted that in Iceland the fishing year runs from September 1 to August 31, and this comes up for Icelandic cod. In their case, they ignore the fishing year, and assume its not a large problem. A participant argued against doing this because the stock is not managed on the calendar year, it is managed on the fishing year.

When they used to use ADAPT (a kind of virtual population assessment) they did the year and then projected for the first few months. It is something they will have to be aware of and either provide advice ignoring the fishing year or doing 2-step projections.

The Chair then began discussion on whether the assessment methodology has the potential to support quantitative evaluation of Harvest Control Rules (HCR). The Chair commented that the quick answer was yes. The HYBRID model developer argued that with the scenario analysis it would be possible, but with more than that, it would be a larger exercise.

It was asked if everyone was ok with the statement on the SURBA model:

Statistical Analysis System (SAS) SURBA is no longer a reliable assessment model as used in the assessment of this stock in 2010–18 relies on a constant ratio of distribution of fish in the inshore and offshore. Research presented at this meeting indicates that the ratio has not been constant over time. In addition, the constant selectivity for total mortality is "going on 40 years", but evidence is to the contrary. Therefore SURBA, as is currently set up, is no longer a valid assessment model.

There was agreement on the statement.

The Chair then asked which working documents should be upgraded to Research Documents. The documents identified as needing to be upgraded to research documents included: D. Varkey's write-up for the HYBRID Model, J. Champagnat's write-up for the SAM Model, and E. Carruther's work on the fishery interviews to inform catch bounds. A reviewer recommended that the documents undergo an internal review before they go forward. It was determined that the work on condition-based M would be included as an appendix on D. Varkey's Research Document for the HYBRID Model. A reviewer requested that N. Cadigan's analysis of the survey stock weights-at-age be included as a Research Document even though it was not used.

REVIEWER COMMENTS – A. MAGNUSSON

Review of 3Ps cod

Arni Magnusson

Main Concerns

1. The models presented do not fit the RV survey well. The high estimated sigma for the RV survey effectively puts a very low weight on this important data source.
2. It is not clear whether using time-varying M as assessment model input is helpful for the management of this stock. The estimation of time-varying M is statistically challenging and the evidence for extreme M values is somewhat weak. The state-space models can be fitted without explicitly modelling time-varying M, thus indirectly accounting for time-varying M as process variability. Such models would use a simple M value or age-specific M vector as model input.

Secondary Concerns

1. All models presented are quite similar state-space models. Including one or two simpler models might have helped to clarify and highlight the challenges in this assessment. These would also be a valuable reference to diagnose the effect of the process error on model results.
2. The uncertainty about the landed catches has been modelled using bounds with a flat likelihood inside the bounds. The algorithm used to implement this seems to introduce computational challenges that were seen in somewhat odd artifacts in the results and increased run time. If the catch uncertainty was modelled with a normal curve, centered around a catch multiplier of 1, that might help with computations and the ability to run comparable assessments on standard software platforms.

Background

The 3Ps cod fishery has been subject to several important changes over the years, such as shifts in gear composition, a moratorium, survey area expansion, and years of low body condition factor. Weights-at-age in the data have diverged over time, with the survey weights declining while the catch weights remain high. There is considerable uncertainty about the magnitude of the landed catches, especially in the years shortly before the moratorium.

Models

Three assessment models were presented at the framework meeting: SSM, SAM, and a HYBRID model. They are all state-space models that resemble each other, and for each model several variants were presented with model run diagnostics.

The model variants explored alternative ways to model fixed or time-varying M, including or excluding the coastal sentinel survey, statistical correlation in the model estimation, and the uncertainty about annual landings.

RV Survey

In general, the models do not fit the RV survey data well, which is a concern as this is a standardized survey that covers the stock area. This seems to be a challenge with the survey data, regardless of what assessment model is used, with considerable year effects (in a given year, all ages are high, or all ages are low) that makes it difficult to calculate the relative size of cohorts. There is indication that the survey catchability may vary annually linked to oceanographic conditions. No large tows tend to occur in cold areas in a given year.

Catch Bounds

All years have uncertain catches in tonnes, expressed as lower and upper bounds. The widest bounds are in the years 1976–77 (multiplication factor 0.7–1.3, when the EEZ was established) and 1987–93 (0.5–1.5, before the moratorium). Some models used a censored likelihood to establish these bounds, with a flat likelihood between the bounds.

Natural Mortality

Time-varying M has been estimated based on annual variations in body condition. The working hypothesis is that when many fish have a relative K factor below 0.85, there is an increase in starvation-induced mortalities. This analysis predicts annual values of M for different length groups, which may not be straightforward to apply to age groups.

The conversion from body condition to M seems somewhat sensitive, with annual shifts in condition causing M to fluctuate from 0.2 to around 0.9 for larger fish. The resulting M seems to react strongly to slight changes in the condition factor. It is somewhat surprising to see that years with the most extreme spikes of low condition factor (12–15% of fish below 0.85 relative K) result in M values around 0.9 (killing around 50% of fish). The temporal trend in this estimated M is a long-term increase, with the highest values in recent years.

Some model variants use a fixed M, such as 0.2 or 0.3, sometimes age-specific with higher M values for younger fish. Another model variant uses time-varying M based on the body condition study. Yet another model variant uses this condition-based M as input, but estimates a multiplier to inflate or deflate M, with one M time series for younger fish (ages 2–5) and others for older fish (ages 6+), modelled as deviations from M=0.3. This model estimates M for younger fish as very steady around 0.3, but M for older fish increasing from around 0.25 in 1980–2005 to around 0.5 in 2010–18.

Time-varying M is known to be hard to estimate, especially for a fishery where the assessment data do not seem very informative to begin with.

State-space models can use fixed M as input, while indirectly taking account of underlying changes in M as part of the general process variability. At the meeting, some participants indicated their preference to have the final assessment model using time-variable M as input, while others (mainly reviewers) indicated their preference to use fixed M as input and leaving fluctuations in the population dynamics to process variability, with no pre-set pattern.

Practical Issues with the HYBRID Model and SAM

Two special features of the HYBRID model were designed specifically to handle the estimation of:

1. scaled time-varying M and
2. the catch bounds.

SAM has some ability to estimate scaled M, by running the model iteratively to search for the M multiplier that has the best overall model fit. SAM also has the ability to estimate a year-specific catch multiplier for a given period, such as 1987–93, but these are not expressed as pre-set bounds.

Both modelling platforms have some pros and cons for the 3Ps cod assessment. The HYBRID model is tailored for this particular assessment and the stock assessors have the expertise to work with this model and make modifications based on modelling decisions. SAM is used widely, has been peer-reviewed and extensively self-tested, and has the ability to run projections. Projections are not yet implemented in the HYBRID model.

Analyzing the Effect of Time-Varying M

To quantify and visualize the effect of estimating time-varying M, one can compare HYBRID model runs 10 and 11. The only difference between runs 10 and 11 is the explicit estimation of time-varying M.

As expected, these model runs are quite different in how they fit the data, the estimated stock size and fishing mortalities, retrospective diagnostics, etc.

Final Comments

Overall, the research presented at the framework meeting was of high quality, well prepared, and clearly communicated. The meeting was characterized by constructive dialogue, good progress, and general agreement on the major issues.

Discussion

There was some discussion during the reviewer comments about some strange shapes in the random walk portion of the exploratory SSM model. The reviewer asked if there were computational challenges with the flat catch bounds. The model developer said that they had to use a two-step process where the model fits to the mid-point of the catch bounds as if the initial value ended up really far out of the catch bounds you ended up with NA (missing data) problems. For the HYBRID Model, an atomic function was developed to calculate the derivative for the censored bounds so that it could be done in one run. In the exploratory model, pnorm-pnorm (probability distribution) is being calculated and if Z >37 then you are effectively calculating 1–1. Using the derivative resolves the problem. In summary, there is some potential for computational issues. It was noted that an advantage of this method is that it allows for asymmetric bounds.

The developer of the HYBRID model noted that on the point of the statistical catch-at-age, the HYBRID model could be reworked with a very simple formulation with no process error and there would be no stock-recruitment, it could be fit to catch numbers-at-age and ignore proportions and the censored landings. One of the first models attempted, the logistic was trying to do that, but it had process errors on it. However, you could turn off the process error. Doing this would allow for the difference between having process error in the model and not having it to be clarified.

The developer of the HYBRID model asked if you considered the process error to be M, would you factor it into the catch calculations, or would it be ignored there or would the process error feed into Z? The response was that it was in the stock equation. The developer suggested that this does not apply directly to M. The reviewer suggested it might get directional changes. The developer suggested that it affects M but does not get applied to the decay in that year.

The developer of the SAM model asked if you could capture or account for the M as a time-varying process error, would the process error which is normal, capture a trend as there is a directional bias, while process error should not have a bias. The reviewer suggested that a comparison of models 10 and 11 would show that it does not do it in nearly the same way as explicitly estimating the time-varying M. It is almost 3 levels, you could completely ignore with a statistical catch-at-age with no process variability, the next step is model 10 which would try to track it indirectly, and then model 11 fully modeling the changes in M. All 3 are interesting, but he thought that the reviewers would prefer model 10. He heard very mixed thoughts on modeling time-varying M. They felt that the reviewers were more concerned about time varying M.

A participant asked if the process error is accumulating and positive in recent years, doesn't this violate the underlying assumption of the process error that it is a stationary process? If you have autocorrelation in your process error, you need to directly account for that. The reviewer agreed.

The developer of the HYBRID model asked the reviewers about their experience with removing data series from their models. How do you decide to do this? A reviewer responded they had done this where they had both a spring and an autumn survey, and they dropped the autumn survey as there was a lot of noise. They noted that it was nice to have two surveys for age effects. As time passes, the catch-at-age matrix takes over and so the old surveys do not mean so much. They understand that there's pressure to include survey to justify their being done. However, from a data perspective, there must be a neutral approach to ask why it is being included.

A second reviewer said as a scientist based in Iceland and now seeing a lot of ICES assessments, it is very common to use the RV surveys that cover all of the stock. For their equivalent of the sentinel survey, it is looked at on the side. The final reviewer said it was tricky to use a model to judge whether a survey should be excluded or not. On the other hand, a priori a lot of thought should be put into whether the survey should be used or not. They noted that with a 20–30% decline in weights-at-age since 1995, catchability-at-age in a gillnet survey has assuredly changed. The comment to put a series into the model and if it is wrong the model will ignore it is problematic. If you know the series is wrong and somehow the model does ignore it, you have a real problem. In many cases, the models do not have an ability to ignore data series as they average through data series and the ones that are longer and have stronger signals will dominate. Questions that should be asked include is catchability constant over time? If catchability changes over time you have to model it or exclude it. Is the series representative? Are the survey indices correctly standardized? This is a problem for the longline survey where the locations of sites have changed over time and the design is highly imbalanced. Standardizing with a Generalized Linear Model (GLM) does not solve the problem. It may be worth going back to make sure that the standardization calculation for the longline survey is valid. The Chair noted that there had been a recommendation that had came out of the Northern cod meeting in March on the sentinel survey and there will be a review done on sentinel data where this will be looked at.

REVIEWER COMMENTS – G. THORDARSON

The work presented at the meeting was of high quality and mostly well documented. Most of the material was available for review a week before the meeting. As is to be expected some further work went on by the assessment scientists after that. But the material provided gave a good overview of the assessments presented at the meeting.

The process as set up is in many ways understandable and makes sense. That is to present various assessment models and scrutinize fit to data, residuals and retrospective diagnostics and not to focus on reference points and prognosis. However, including that part gives opportunity to do sensitivity testing by including assessment error in a MSE framework. Additionally, there are other things that complicate using models such as the assumptions used when projecting biomass one year ahead. This can result in considerable changes up or down when comparing advice from year to year.

The three modelling frameworks all appear to be an improvement on the current SURBA assessment that uses only survey data and produces estimates of Z but not F. Additionally the SURBA has shown strong patterns in the retrospective diagnostics in recent years.

An exploratory assessment SSM-model looked promising. The model assumes constant M and deals with various data inconsistencies with process error. One challenge is that in one year the model "lost" around 20 thousand tonnes of spawning stock biomass out of a stock of around 70 thousand tonnes. A SAM model with the sentinel survey and the HYBRID model 11 are similar. A proposal to use the HYBRID model as a basis for advice and run SAM alongside looks like a sensible one.

The framework meeting spent considerable time on discussing time variant M. Work was presented where condition is used to estimate changes in natural mortality. Additionally, mortality signals were extracted from tagging data. The approach looks promising but various questions were raised about the validity of including these estimates directly into the assessment. Therefore, it is premature to take it in directly as this signal may be confounded with other factors that affect M. Accounting for changes in M through a process error term sounds like the most sensible approach at the moment.

All the models presented included all the available data except in some cases the sentinel survey data was left out. Including surveys that stopped several years ago may be questioned. In most cases they do not get much weight in the fit and the catch-at-age matrix takes over as normally the signal in that data set is much stronger. It would be interesting to see the effects of just running SAM and the HYBRID model with the RV survey and the sentinel line trawl. The use of the gillnet sentinel survey as a tuning series is questionable as it shows very different signals and seems to have high variance and has difficulty in tracking cohorts going through the population. The data collected in the gillnet survey may be thought to be useful in another context.

It might be interesting to see the results from completely different model frameworks i.e., length-based models such as Stock Synthesis or Gadget. Production models might possibly be informative.

Patterns in retrospective diagnostics are a challenge for most of the assessments presented for 3Ps cod but that is a very common issue, for example in the stocks assessed by ICES. Obviously, this should be addressed as has been tried at this framework meeting. However, at low levels of F the probability of such patterns increases as the model takes longer to converge back in time. Especially for a relatively long-lived species such as cod. One way to tackle this is through a HCR which dampens annual changes in TAC. In the HCR for Icelandic cod previous years TAC weighs half against the 20% of the estimated biomass this year. At the time the scientists thought this was good so that sudden increases in the assessment would not lead to overexploitation, the industry took this as a damper against sudden decrease in the assessment.

Discussion

A model developer asked why the reviewer said it is hard to starve a cod to death, given that it had been done in the tank experiments. The reviewer responded that there are always some prey available that they can consume in order to survive, not thrive. As soon as they get to a certain size, they are able to eat smaller cod.

REVIEWER COMMENTS – H. BENOÎT

I found the three working papers to be very well done. The collective group of highly competent authors working on this one stock is very impressive. The authors have done a tremendous job modelling what is clearly an analytically problematic stock. I am impressed by the diversity of models explored and the degree to which different hypotheses were evaluated.

Perhaps my largest criticism of the process is not the modelling framework review itself, but the fact that reviewers were not invited to the critical review of model inputs, nor were the details of that review available to us. Model inputs were therefore essentially immutable at the meeting even though some clearly present problems, and there was little possibility to fully understand the potential limitations of different inputs. It is well known that assessment outputs can be quite sensitive to the inputs, and poor quality of the latter cannot necessarily be compensated by sophisticated models, particularly if the problems with the inputs are not made explicit.

A key example is the dependence of model results on the sentinel indices. It is clear that the nature and intensity of sampling for these indices has evolved over time in ways that cannot necessarily be compensated for in their standardization. Furthermore, and notwithstanding the preceding, it appears that the design matrix for the standardization for the longline index (at least) is highly unbalanced. It is therefore not clear to me that the standardized series provides a true relative index as it is assumed to. Revisions to the standardization should be considered and the ability to construct consistent indices should be evaluated critically.

In addition, and importantly, what is clear is that the gillnet index (perhaps longline also) is now based on catches in a very spatially restricted coastal area. Furthermore, given non-negligible changes in population weights-at-age (20–30% decline) and lengths-at-age (10%), we would expect changes in catchability for sentinel gillnet survey (especially) and longline survey indices. The question must be posed whether it is realistic to assume that the indices proportionally reflect abundance at the population scale even if the index were properly standardized.

Some of the authors used model diagnostics to decide on the appropriateness of retaining or not indices or on the model treatment of the indices. This practice is certainly not unique to 3Ps cod, but is one with which I disagree. I believe it is important to carefully consider the pertinence, representativeness (bias) and variability of indices a priori and on their own merits first. A model can provide bad diagnostics for an index because that index is wrong and because the model is wrong. Using the model to discriminate assumes that model is correct.

The three modelling frameworks appear to be a significant improvement on the former SURBA model and constitute a huge shift from the virtual population assessments of not too long ago. We’ve added a lot of flexibility to the models using random effects, but in very general terms I wonder if we fully understand their consequences. There is clearly an improved ability to explain/fit the data, but this can produce strange output that will be difficult to explain: such as retrospective patterns in Z, but not recruits or spawning stock biomass, that might be offset by process error. It is also not clear to me that this improved ability to fit data does not come at the price of overfitting.

The framework review concentrated on model suitability and the model fit to the observed data. However, the review lacked evaluation of model performance through simulation. Similarly, the ability of the model to provide reliable projections in support the development of science advice for the stock was not assessed.

It is apparent from the different models that M has varied over time in a non-stationary manner for the stock, as it has for other cod stocks in the Northwest Atlantic. In other stocks, failure to (correctly) model M variation has been an important source of model misspecification. I am concerned about the use of a condition index by D. Varkey as a proxy for M. The authors have used it in a manner that the original authors (Lambert et al.) did not intend based on their experiments, as is reflected in these authors' use in their own population models, where condition contributes only to a component of M, not all of M, and in a manner different from the thresholds used for 3Ps cod (see Chassot et al. 2009). Fixing M as a strict function of condition as was done here is an exceedingly strong assumption, the consequences of which were not well explored. The rationale for doing so was also not entirely convincing. While low condition might make cod vulnerable to different sources of mortality, clearly the magnitude of these other sources is also important, e.g., predation risk may increase with declining condition, but likely increases with predator abundance.

Other assessments, both with and without tagging data, estimate M trends independently within the model. Apparently, this is not possible for 3Ps cod when M was freely estimated, although I wonder whether there isn't sufficient tagging data to estimate M for a period of time, which might allow the model to estimate the trend independently for the remaining periods. If this is not possible, and given model uncertainty particularly with respect to M, it may be wise to favour models with autocorrelated process error which will indirectly account for time-varying M.

Discussion

A participant noted that they had not talked about projections at this meeting and it is going to be a key point at the assessment. They thought that they would need some extra time at the assessment to work through the projections with the new modeling framework and have the time to really understand what is going on and that the uncertainties were captured appropriately.

The Chair agreed that they could have used more time for the framework process. It would have been a good idea to have the reviewers at the data input meeting. Perhaps similar to the Northern cod meeting where data was discussed prior to the meeting and then further by email. Some firm conclusions on what they had decided on would have been helpful.

A participant said that they thought that the meeting could have been structured differently to make it easier to look at the three models. They suggested that more time could be spent on where the models differed. The Chair agreed.

The developer of the SMM model noted that in 2004 they were trying to have a framework meeting for 3Ps cod, and they had 5 models. They noted that 2 of the models were being advocated for by no one, 1 was known not to work. It has taken 15 years to get to this point. The recommendation to DFO was that framework review meetings need to happen more frequently. In the United States that it is a normal part of business and is scheduled, often on a 5 year basis.

The developer of the HYBRID model noted with regard to model complexity, model complexity is a concern, particularly when the data is very noisy. This is why the model was developed as a

HYBRID model, so that it can be both a very complex model and a very simple model. The model has been designed to be generalizable. It would be fairly easy to adapt this model to another stock where there are similar data sources. Z retrospective diagnostics has been an issue. Part of the Z retrospective diagnostics problems are coming from the desire to chase M in a meaningful. It may or may not be the right way. They will do the comparison of M10 and M11 and see where that goes.

RESEARCH RECOMMENDATIONS

In order to refine the M levels included in the model, there was a list of recommendations to examine in the near future:

1. Ideally M should not only be time-varying, but it should also be variable over ages. There was concern over the trialed method of converting condition into a level of M (additive over months), and it was recommended that a seasonal instead of a monthly approach should be tested. The sensitivity of the cut off value (level below which a fish does not recover from poor condition) should be examined. In addition, the group felt it would be useful to compare estimates of M from both condition and tagging. Although these would be expected to differ because the estimate of M would presumably include removals from all factors (rather than just F), the estimate of M from condition would only be removals due to starvation-induced mortality – and while this would likely be a large proportion of M, it would not be everything. Tagging could be used to compare levels of estimated M with the levels estimated from Model M11. Further refinement should go into scalar M levels used in several of the models. Further, to link condition to M through tagging, weights of tagged fish should be collected, and cod should be tagged in spring.
2. Investigate density dependent and other components on M.
3. Short term research recommendations were made for the HYBRID and SAM models with a targeted completion date of the assessment meeting in November. For Model M11 of the HYBRID model, determine the causes and implication of the observed retrospective diagnostics (Z vs. numbers-at-age vs. Process error), assuming they are still present following completion of the planned changes to the model. If the questions about patterns in the retrospective diagnostics are not resolved successfully, the HYBRID model to apply would be run “M10”.
4. Development of the SAM model should continue as a backup to the HYBRID model. Scenarios that should be run include the base model using the sentinel data and constant M, and to explore the work that is being done using a scalar value for M. This is new work, and uncertainty wouldn’t be carried forward in the projections with this model/method.

Medium term recommendations:

1. For the upcoming stock assessment in November 2019, the options for changing the implementation of stock recruitment methods are limited. Going forward, if a stock-recruitment relationship can be identified, that relationship should be implemented and fitted with the model in order to allow the model to account for errors in both variables.
2. Additional investigations are needed for the spatial components of both the available data and the models. This includes temporal variation in the spatial patterns of fishery effort (e.g., the effects of changes in gear usage and whether fishing effort is focused on inshore or offshore areas of 3Ps).

Long term recommendation:

1. Uncertainty about input data (e.g., age composition, survey indices) for the models should be included in the assessment. This will be difficult.

REFERENCES CITED

Berg, C.W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J. Mar. Sci. 73(7): 1788–1797.

Bernier, R.Y., Jamieson, R.E., and Moore, A.M. (Eds.). 2018. State of the Atlantic Ocean Synthesis Report. Can. Tech. Rep. Fish. Aquat. Sci. 3167: iii + 149 p.

Brodziak, J., Ianelli, J., Lorenzen, K., and Methot Jr., R.D. (Eds.). 2011. Estimating natural mortality in stock assessment applications. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-119. 38 p.

Casini, M., Eero, M., Carlshamre, S., and Lövgren, J. 2016. Using alternative biological information in stock assessment: condition-corrected natural mortality of Eastern Baltic cod. ICES J. Mar. Sci. 73(10): 2625–2631.

Chassot, E., Duplisea, D., Hammill, M., Caskenette, A., Bousquet, N., Lambert, Y., and Stenson, G. 2009. Role of predation by harp seals *Pagophilus groenlandicus* in the collapse and non-recovery of northern Gulf of St. Lawrence cod *Gadus morhua*. Mar. Ecol. Prog. Ser. 379: 279–297.

Dutil, J.-D., and Lambert, Y. 2000. Natural mortality from poor condition in Atlantic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. 57(4): 826–836.

Koen-Alonso, M., and Cuff, A. 2018. Status and trends of the fish community in the Newfoundland Shelf (NAFO Div. 2J3K), Grand Bank (NAFO Div. 3LNO) and Southern Newfoundland Shelf (NAFO Div. 3Ps) Ecosystem Production Units. Serial No. N7023. NAFO SCR Doc. 18/070. 11 p.

Kulka, D.W., and Stevenson, S.C. 1986. Trends in discarding of Divisions 2J3KL cod by the Newfoundland offshore fleet for 1981–86, with particular reference to size of fish. CAFSAC Res. Doc. 86/101. 19 p.

Kumar, R., Martell, S.J., Pitcher, T.J., and Varkey, D.A. 2013. Temperature-Driven Decline of a Cisco Population in Mille Lacs Lake, Minnesota. N. Am. J. Fish. Manag. 33(4): 669–681.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J. Fish Biol. 49(4): 627–642.

Nielsen, A., and Berg, C.W. 2014 Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158: 96–101.

APPENDIX I: TERMS OF REFERENCE

Terms of Reference

3Ps Cod Assessment Framework

Regional Peer Review – Newfoundland and Labrador Region

October 8-10, 2019
St. John's, NL

Chairperson: Karen Dwyer
Rapporteur: Aaron Adamack

Context

The assessment framework meeting will focus on methodology for estimating the population size and other stock status indicators of cod in Northwest Atlantic Fisheries Organization Subdivision 3Ps.

The most recent (2018) stock assessment for 3Ps cod utilized a survey-based assessment model to determine stock status. This method relies only on a single index of stock abundance (from Fisheries and Oceans Canada's multi-species survey), and generates relative estimates of total mortality rates (Z) and stock size. Caveats to this approach include an inability to incorporate commercial catch or other data sources, and an inability to provide estimates of fishing mortality or catch-based management advice. An assessment framework meeting will review and consider candidate models of population dynamics for 3Ps cod, specifically those that can incorporate multiple data sources (including commercial catch). The data sources available for modeling the 3Ps cod population were determined previously at a data review meeting held in May 2019.

Objective

Evaluate potential assessment models for possible use as the basis for providing science advice for 3Ps cod. If a model is found to be appropriate for the assessment of 3Ps cod it would be used subsequently during the stock assessment from November 19-22, 2019. Specifically, the framework meeting will address the following:

- Evaluate potential assessment models to determine if they provide a sufficient framework for providing scientific advice on the impact of exploitation on 3Ps cod, in particular estimating the stock size (biomass and abundance), recruitment, fishing mortality, and potentially natural mortality of the population.
- Provide direction on projection methods for future catch options.
- Provide direction for an approach to estimating reference points for this stock.
- Discuss whether the assessment methodology has the potential to support quantitative evaluation of harvest control rules.
- Identify uncertainties and knowledge gaps.
- Identify priority short and medium-term research recommendations to improve data sources, assessment model formulation and estimation, and projection methods.

Expected Publications

- Proceedings
- Research documents

Expected Participation

- Fisheries and Oceans Canada (DFO) (Science and Fisheries Management Branches)
- French Research Institute for Exploitation of the Sea (IFREMER)
- Provincial Department of Fisheries and Land Resources
- Industry
- Academia
- Non-Governmental Organizations
- Other invited experts

Reference

DFO. 2019. Stock Assessment of NAFO Subdivision 3Ps Cod. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2019/009.

APPENDIX II: AGENDA

Framework project on population models for Northwest Atlantic Fisheries Organization (NAFO) Subdivision 3Ps Atlantic Cod

Chairperson: Karen Dwyer
October 8-10, 2019

Memorial Room - Northwest Atlantic Fisheries Centre
80 East White Hills Road, St. John's

Tuesday, October 8

Time	Topic	Presenter
09:00	Opening remarks and overview of Regional Peer Review Process by chairperson	K. Dwyer
-	Overview of available indices and landings	D. Ings
-	Re-construction of catch-at-age	H. Penney
-	Modeling catch and stock weights	N. Cadigan
-	Informing natural mortality - Cod condition - Tagging data	P. Regular G. Robertson
-	Fisher interviews to inform catch bounds	E. Carruthers
-	SSM for 3Ps cod	N. Cadigan

Wednesday, October 9

Time	Topic	Presenter
09:00	Continued… SSM for 3Ps cod	N. Cadigan
-	SAM style model for 3Ps cod with/without post-stratification	J. Champagnat
-	HYBRID model for 3Ps cod	D. Varkey

Thursday, October 10

Time	Topic	Presenter
09:00	Continued – discussion on model formulations	All
-	Reviewer comments	A. Magnusson, G. Thordarson

Time	Topic	Presenter
		H. Benoit
-	Drafting of Summary Bullets	All
-	Closing remarks and *ADJOURN*	Chairperson

Notes:

- Health breaks will occur at 10:30 a.m. and 2:30 p.m. Coffee and tea can be purchased from the cafeteria.
- Lunch (not provided) will normally occur 12:00-1:00 p.m.
- Agenda remains fluid – breaks to be determined as meeting progresses.
- This agenda may change.

APPPENDIX III: LIST OF PARTICIPANTS

NAME	AFFILIATION
Aaron Adamack	DFO Science – NL
Arni Magnusson	ICES
Bob Rogers	DFO Science – NL
Brian Healey	DFO Science – NL
Chelsey Karbowski	Oceans North
Danny Ings	DFO Science – NL
Darrell Mullowney	DFO Science – NL
Devan Archibald	Oceans North
Divya Varkey	DFO Science – NL
Erin Carruthers	FFAW
Eugene Lee	DFO Science – Centre for Science Advice
Geoff Evans	DFO Science – Emeritus
Greg Robertson	DFO Science – NL
Gudmundur Thordarson	MFRI
Heather Penney	Memorial University of Newfoundland
Hugues Benoit	DFO Science – Quebec
Jennica Seiden	DFO Science – Groundfish
Joanna Mills Fleming	Dalhousie University
Joanne Morgan	DFO Science – NL
Joel Vigneau	IFREMER
Jonathan Babyn	DFO Science – NL
Juliette Champagnat	IFREMER
Karen Dwyer	DFO Science – NL
Keith Lewis	DFO Science – NL
Kris Vascotto	Atlantic Groundfish Council
Krista Baker	DFO Science – NL
Laura Wheeland	DFO Science – NL
Luiz Mello	DFO Science – NL
Mark Simpson	DFO Science – NL
Nicole Rowsell	Gov. NL – Department of Fisheries and Land Resources
Noel Cadigan	Marine Institute – CFER
Paul Regular	DFO Science – NL
Rajeev Kumar	DFO Science – NL
Shelley Dwyer	DFO Resource Management – NL