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 2021/048

Pacific Region

Proceedings of the Pacific regional peer review on Walleye Pollock (*Theragra chalcogramma*) stock assessment for British Columbia in 2016

November 14-15, 2017
Nanaimo, British Columbia

Chairperson: Dominique Bureau
Editors: Matthew Grinnell and Dominique Bureau

Fisheries and Oceans Canada
Science Branch
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

November 2021

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 978-0-660-40567-4 Cat. No. Fs70-4/2021-048E-PDF

Correct citation for this publication:

DFO. 2021. Proceedings of the Pacific regional peer review on Walleye Pollock (*Theragra chalcogramma*) stock assessment for British Columbia in 2016; November 14-15, 2017. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2021/048.

Aussi disponible en français :

*MPO. 2021. Compte rendu de l’examen par les pairs de la région du Pacifique sur l’Évaluation du stock de goberge de l’Alaska (*Theragra chalcogramma*) pour la Colombie-Britannique en 2016; les 14 et 15 novembre 2017. Secr. can. de consult. sci. du MPO. Compte rendu 2021/048.*

TABLE OF CONTENTS

SUMMARY

These Proceedings summarize the relevant discussions and key conclusions that resulted from a Fisheries and Oceans Canada (DFO), Canadian Science Advisory Secretariat (CSAS) Regional Peer Review meeting of November 14-15, 2017 at the Pacific Biological Station in Nanaimo, BC. A working paper to assess the stock status of Walleye Pollock in BC was presented for peer review.

In-person and web-based participation included Fisheries and Oceans Canada Science, DFO Fisheries Management, and commercial fishing sectors.

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report providing advice to Fisheries Management to inform fishery planning.

The Science Advisory Report and supporting Research Document will be made publicly available on the Canadian Science Advisory Secretariat website.

INTRODUCTION

A Fisheries and Oceans Canada (DFO) Canadian Science Advisory Secretariat (CSAS), Regional Peer Review (RPR) meeting was held on November 14-15, 2017 at the Pacific Biological Station in Nanaimo to review the Walleye Pollock stock assessment for British Columbia in 2017.

The Terms of Reference (TOR) for the science review (Appendix A) were developed in response to a request for advice from Fisheries Management. Notifications of the science review and conditions for participation were sent to representatives with relevant expertise from DFO Science, DFO Fisheries Management, and commercial fishing sectors.

The following working paper (WP) was prepared and made available to meeting participants prior to the meeting (working paper abstract in Appendix B):

> *Starr, P.J. and Haigh, R. Walleye Pollock (Theragra chalcogramma) stock assessment for British Columbia in 2017. CSAS Working Paper 2013GFR03*.

The meeting Chair, Dominique Bureau, welcomed participants, reviewed the role of CSAS in the provision of peer-reviewed advice, and gave a general overview of the CSAS process. The Chair discussed the role of participants, the purpose of the various RPR publications (Science Advisory Report, Proceedings and Research Document), and the definition and process around achieving consensus decisions and advice. Everyone was invited to participate fully in the discussion and to contribute knowledge to the process, with the goal of delivering scientifically defensible conclusions and advice. It was confirmed with participants that all had received copies of the Terms of Reference, background information, and supporting documents.

The Chair reviewed the Terms of Reference and the Agenda (Appendix C) for the meeting. The Chair then reviewed the ground rules and process for exchange, reminding participants that the meeting was a science review and not a consultation. The room was equipped with microphones to allow remote participation by web-based attendees, and in-person attendees were reminded to address comments and questions so they could be heard by those online.

Members were reminded that everyone at the meeting had equal standing as participants and that they were expected to contribute to the review process if they had information or questions relevant to the paper being discussed. In total, 19 people participated in the RPR (Appendix D). The rapporteur for the meeting was Matthew Grinnell.

Participants were informed that reviewers Dr. Doug Swain (DFO, Gulf Region) and Dr. Sean Anderson (DFO, Pacific Region) had been asked before the meeting to provide detailed written reviews for the working paper to assist everyone attending the peer-review meeting (Appendices E & F, respectively).

The conclusions and advice resulting from this review will be provided in the form of a Science Advisory Report to Fisheries Management to inform fishery planning for Walleye Pollock. The Science Advisory Report and supporting Research Document will be made publicly available on the Canadian Science Advisory Secretariat website.

REVIEW

Working Paper:	Walleye Pollock (*Theragra chalcogramma*) stock assessment for British Columbia in 2017. Starr, P.J. and Haigh, R. CSAS Working Paper 2013GFR03.
Rapporteur:	Matthew Grinnell
Presenter(s):	Rowan Haigh and Paul Starr

PRESENTATION OF WORKING PAPER

Rowan Haigh and Paul Starr presented the working paper. The authors provided background information on Walleye Pollock (WAP), and described the two BC stocks considered in this assessment (BC North and BC South, respectively). They noted that this was the first quantitative stock assessment of Walleye Pollock in BC. They described the surveys that collect data on Walleye Pollock abundance, and described the commercial catch data for Walleye Pollock. The authors described the delay-difference (DD) stock assessment model (iSCAM DD), and indicated that the model code had been slightly modified from the original code supplied by Robyn Forrest.

The authors described the process used to select data for inclusion in the model, and described the effects of the included data. They described the process used to develop growth models for the two BC stocks, which involved using growth data from areas outside BC because there were no adequate data from BC Pollock. Then they described sensitivity analyses with respect to variations in natural mortality, selectivity, and which survey indices were included. The authors described model results for the North and South stocks.

The authors defined the reference points used to determine stock status, and provided decision tables that indicated the probability of projected stock status exceeding reference points under a range of projected levels of catch. Finally, the authors provided recommendations for future research.

PRESENTATION OF WRITTEN REVIEWS

Doug Swain and Sean Anderson provided reviews of the Working Paper (WP) in written and oral presentation formats. Both reviewers commended the authors for a well-written and presented WP. Both reviewers reiterated information included in their written reviews. Main bullets indicate reviewer comments, and sub-bullets indicate author responses when applicable.

DOUG SWAIN

Please refer to Appendix E for the full written review.

Swain acknowledged that WAP is a difficult stock to assess for several reasons, including the lack of ageing data and incomplete spatial coverage for surveys. The authors provided a solution to this problem by using ageing data from other areas, using a population model of intermediate complexity, conducting sensitivity analyses, and comprehensively including uncertainty. This assessment provides a basis for advice on stock status and management options.

During the meeting Swain offered or asked authors for clarification on the following topics:

- There is a large difference in mean weight between North and South stocks, perhaps due to slower growth in the South, migration from South to North, and/or higher exploitation in the South. What are some of the theories about why growth would be slower in the South?
  - Authors suggested that unrepresentative sampling in addition to small sample sizes contributed to the uncertainty with respect to the differences in mean size. Given the sampling program, it's not possible to conclusively decide which hypotheses are the source of the observed weight difference between North and South stocks. Unrepresentative sampling in the early part of the time series may be the reason that early weights were similar for both stocks. The lack of representativeness can be seen in the poor spatial coverage in the early years: samples for the South stock primarily come from minor areas. The authors are not sure why growth would be different between North and South stocks. Migration from South to North is possible, but such a migration would be difficult to demonstrate. The authors noted that young pollock are present in the Gulf of Alaska, indicating that there is local recruitment in the North.
  - Harvesters agree that fish in the North were larger than fish in the South, even in the 1970s.
- What are the implications for this assessment if the North stock is a subset of a larger Alaskan stock?
  - Authors note that assessing the North stock as being an independent stock would be incorrect if the North stock were part of the larger SE Alaska stock. Sustainable harvest rates would likely be higher than predicted in the current assessment, given the low levels of harvest in the Alaska panhandle which is closed to bottom trawling and the much larger implied stock size.
  - However, at this time there is no evidence available to corroborate this theory nor do the resources exist to perform joint stock assessments.
- Add more description about survey vessels to the document: How consistent are survey vessels? Do vessels have similar size and power? Do they use the same trawl net?
  - Authors: Several surveys have been mostly done by the CCGS Ricker. All four synoptic surveys are specified with standard instructions, including survey vessel size (e.g., minimum 1,000 horsepower) and the net (Western 2A) specifications.
- Surveys cover a small part of the stock area, but are used as indices for total stock biomass. Does WAP biomass vary spatially?
  - Authors note that it is not known whether WAP biomass varies spatially. But the surveys show comparable trends which gives confidence that survey coverage is representative. The authors also noted that Pollock is primarily a mid-water species, resulting in high levels of uncertainty in the indices generated by a bottom trawl survey.
- Goose Island Gully (GIG) historical survey has small spatial coverage but extends back to the 1960s. Although there is no good alternative to using the GIG survey in the assessment, it's difficult to justify using GIG biomass index for the North stock.
  - Authors agree, and note that most of Queen Charlotte Sound (QCS) catch comes from the GIG. The authors also noted that five of the eight sensitivity runs for the North stock omitted this survey.
- Although catch per unit effort (CPUE) is standardized for some factors, it's impossible to standardize fishery CPUE indices for some important factors such as technology and harvester behaviour. CPUE can be hyperstable because harvesters target fish. However,

CPUE trends tend to match survey index trends, which provide some weight to using CPUE indices in the assessment models. Therefore, CPUE does not show evidence of hyperstability.

  - The authors agree that targeting and other behavioural tactics will affect CPUE. However, they note that Pollock is not a target species in the bottom trawl fishery used for the CPUE analysis. Rather, it is an occasional by-catch in a more generic fishery, targeted at a range of groundfish species, which is why the presence/absence component of the CPUE is also analysed. Shifts in fishing behaviour will undoubtedly affect this fishery, but hopefully will have less impact on the by-catch of Pollock. The authors noted that there was no apparent contradiction between the CPUE series and the overlapping survey indices, especially when taking into account the large uncertainty associated with this mid-water species in the bottom trawl surveys.

- Delay-difference model is a good choice, given the limited ageing data. This model has a key assumption of knife-edge selectivity at age *k* (the age at which fish are assumed to become fully vulnerable to fishing;), which is a strong simplification.
- Estimates of F_{max} (maximum fishing mortality rate) are too high to be credible for all model runs except S00, S03, and maybe S01 for the North stock, and S00, S03, and S06 for the South stock. Models with non-credible fishing mortality rates (*F*) should not be included in the `model averaging` process.
- Because it's difficult to justify including GIG survey data for the North stock, Swain would like to see a model run that omits GIG data.
  - As noted above, five of the eight sensitivity runs for the North stock omitted this survey.
- Models with *k*>3 are not acceptable for both North and South stocks. Do models with *k*>3 fail because catch includes a significant number of 3-year-old fish? If so, would these models work better if catch was restricted to just 4- or 5-year-old fish and older using a size cut-off?
  - Authors agree that this may be a reasonable suggestion, but given the considerable uncertainty in growth, there would be an equally strong uncertainty in the size at length associated with age 4 or age 5 Pollock.
- Because fixed values are used for standard deviations of all components of process and observation error, omit equations E.27 and E.28.
- Observation error fixed at a small value for all surveys, which forces a tight fit to the indices and allows a poor fit to weight data. Try estimating a single survey index error for the next assessment.
  - Authors noted that doing this for the Shortspine Thornyhead stock assessment did not affect conclusions, and may not have affected the fit to weight data. Could try this for the next assessment when more data are available.
- Model averaging is an important and novel component but should only include models with credible F_{max} estimates. For the next assessment, it might be possible to increase the number of acceptable models by revising fishery catch and indices to match the choice of *k*, estimate rather than fix observation error sigma, and estimate *F* in a way that penalizes unacceptable estimates.

  - Authors believe that high *F*s are because there are not enough aged fish so the model is by definition specified incorrectly. The Hecate Strait (HS) survey might have enough length/weight samples to look into this.
  - As noted above, the uncertainty in the growth model leads to uncertainty in the size associated with specific ages.
- Suggestions for future research include collecting ageing data, and better understanding stock structure.

SEAN ANDERSON

Please refer to Appendix F for the full written review.

Anderson agreed that this is an extensive stock assessment and represents a lot of work, especially with respect to the sensitivity analysis.

During the meeting Anderson offered or asked authors for clarification on the following topics:

- Concern with high *F* values: Is this from one or two years, or more of a systemic pattern? These unrealistic values might mean a mismatch between priors and posteriors.
- The authors use Monte Carlo Markov Chain (MCMC) diagnostics, including correlation plots, trace plots and autocorrelation plots to select models to include in the model averaging exercise. Anderson would like to see some common MCMC diagnostic metrics, such as unfished equilibrium recruitment ($\hat{R}$).
  - Authors agreed that MCMC convergence tests like $\hat{R}$ will be explored in future assessments.
- Some of the trace plots seem to show insufficient burn-in, and might require more burn-in to trust the values. Anderson suggested not showing decision tables for models that haven't converged and to help get to convergence the authors might try to run the MCMC for longer, or allow more burn-in.
- One hypothesis for larger fish in the North is that fish migrate from the South. How would this hypothesis affect model output, and how could it be interpreted in the future? Would ageing data or genetic testing help resolve this issue? Closed-loop simulation might also help resolve this issue. What about implications of the potential rescue of this stock from the larger Alaskan stock?
- Would like to see more explanation and interpretation on how process error is affecting model output. Again, closed-loop simulation is another way to investigate if process error affects model output (aside from sensitivity analyses).
- Concern about why model projections have similar uncertainty range for each year; typically in stock assessment projections, each successive year is more uncertain.
- Discrepancy between 3-year projections and doing a review every 5 years: is there concern about what to recommend after the 3-year projections?
  - Authors agree, and state that they have concerns about the reliability of projections using this model.
- Regarding Figure 10: shouldn't each model have its own unique reference points?
  - Authors agree, and will add this into the final version.

- There is some danger to averaging across different models because of the strong influence that extremes can exert. The authors should consider showing the range of values for all models for interpretation in case there are extremes.
- There are too many decimals in reported probabilities in decision tables, which give a false sense of precision. Probabilities should be reported to 2 or 3 decimals
  - Authors agree and will update this in the document.
- Coefficient-distribution-influence (CDI) plots (e.g., Figure C.30) are interesting, but need more explanation in the figure description.
- Delta-lognormal equation: are the coefficients in log space, or log-odds space?
  - Authors will confirm this for the final version.
  - Authors will also confirm that equation C4 is correct.
- Figure 10: removing CPUE index changes the results. This outcome needs more explanation and interpretation in the document.
  - The authors note that CPUE index, given the long length of the series, the availability of an index in each year and the relatively low coefficient of variation (CV) associated with each index, provides a great deal of stability to the model. Removing the CPUE index series allows the high CVs associated with fishery independent series to interact in unpredictable ways, but which do not strongly contradict models which include the CPUE index.
- Were midwater generalised linear model (GLM) outliers dropped from the analysis due to their heavy tails?
  - Authors noted that the midwater GLM was based on a targeted fishery that seemed unlikely to reflect changes in abundance, given issues such as: market forces which influence targeting, inconsistent fishing spatially year to year, and inconsistent fishing vessels year to year.
- Recommend using management strategy evaluation (MSE) with closed-loop simulation for this stock to choose among the three hypotheses regarding why the North has larger fish.

DISCUSSION

- Swain indicated that all of his issues have been addressed except for high F_{max} values, and doing a model run with k=3 and M=0.35 (natural mortality) excluding GIG data.
  - Group consensus was to not do another model run at this point.

Consistent uncertainty in successive year model projections

  - The authors noted that model projections are not reliable, given the lack of ageing data in the model. The only stochastic parameter is recruitment, which might not have a large impact on short projections like those in this assessment.
  - Another problem with these projections is that they all predict that the stock will crash, which seems unlikely given the long history of this fishery. The projected stock crash might be caused by the model using average recruitment over the time series, or improper growth models.

- It is not known why the uncertainty in the model projections does not increase as the number of projection years increase. As noted by one of the reviewers, this model behaviour is counter-intuitive.

Mismatch between 3-year projections and 5-year reviews

Because the document provides 3-year projections, but the authors suggest waiting 5 years before another assessment, there will be no advice in years 4 and 5.

- The authors are not sure what to do in years 4 and 5, however even the 3-year projections are not reliable because the model does not have an age structure. Therefore, model projections need to be interpreted with care.
  - Because of high uncertainty and low confidence in 3-year projections, meeting participants agreed to omit the 3-year projections, and to only provide 1- and 2-year projections.
- Fisheries Management needs advice more frequently than it's provided, and Fisheries Science has insufficient resources to provide more frequent advice.
  - Fisheries Management is unsure if/how the advice from this paper will be incorporated into decision making.
- In other fisheries, managers use the last year's total allowable catch (TAC) combined with survey indices.

Model projections show low biomass

- Recruitment is modeled using a stock-recruitment relationship, and recruitment deviations.
- Low mean recruitment or wrongly specified productivity could cause projected stocks to 'crash.'

PRESENTATION ON HIGH *F* VALUES

The model calculates annual fishing mortality (*F*) to match the observed catch for each MCMC iteration. *F*s are calculated using the Baranov catch equation in the assessment model. *F* is instantaneous, and *F*=20 means that all recruited fish are caught, which implies that all spawners are caught because the model assumes that the age of maturity is the same as the age of recruitment to the fishery. The F_{max} figures will be updated to quantile plots (instead of only the maximum), and additional explanation will be required in the text to explain the meaning of F_{max} in Tables 2 and 3.

- In the North, high *F* values are only in some years and some MCMC iterations; 95% confidence intervals (Cis) show a more reasonable picture, but some models have quite high median *F*.
- Note that the upper bound on *F* is 20 in the model.
  - Swain noted that *F*=2 would be really high for this species; it would require an unrealistic amount of effort to achieve *F*=2 given WAP life history.
- In the South, *F*s are a bit lower and more plausible. Some models only have high *F*s in one or two years (e.g., S01, S04). Here, most models with high *F* also have *k*=4 or *k*=5, which means that the early age classes are not vulnerable to the fishery.

- Swain indicated that these high *F*s are implausible and that these sensitivity runs should be discarded.
  - High *F* suggests a mismatch between what the model is assuming, and what harvesters are catching. Thus the model is misspecified: high *F*s indicate that all recruited fish are caught, and this model assumes that recruitment is equal to maturity. Models that give non-credible *F*s probably also give other non-credible estimates. Only keep models with k=3 for the model averaging process.
  - Authors note that they could proceed using only models with k=3 for the model averaging process.
- Suggestion to only retain models with $F_{max} \leq 2$ because F_{max} >2 is implausible. This would only keep models with k=3, although models with k=4 and k=5 tend to have a better fit to mean weight data for both the North and South.
- Meeting participants agreed to include models where F_{max}>2 in only 1 year, on the assumption that growth or recruitment will vary between years and that fewer age three fish were available.

GENERAL DISCUSSION

General discussion was focused on topics related to input data informing the stock assessment modeling, stock assessment assumptions and analyses, assessment results and output, uncertainties and information gaps, ecosystem considerations and developing conclusions and future work recommendations in association with developing a Science Advisory Report (SAR).

MEAN WEIGHT

Discussion about why fish in the North are larger than fish in the South.

- Observations in the 1970s indicated the mean weight between the two regions was similar.
  - The data from the 1970s might not be representative because there were very few samples from that period. In addition, samples from the 1970s were collected using different gear types and different locations and tended to be samples taken from landings, which may not reflect any on-deck sorting.
  - Harvesters noticed size differences between North and South stocks in the 1970s.
- Currently, insufficient data are available to determine the cause of this observed difference, but three hypotheses are suggested:
  - Difference in growth rates,
  - Migration from South to North, and/or
  - Higher exploitation in the South.
- There was agreement that borrowing growth data from Alaska and the Asian Sea of Okhotsk was required for this BC WAP stock assessment, given the lack of available BC data. However there were questions regarding the suitability of these data for the BC stock.
  - Collecting BC length-age data was identified as the best way to avoid using age data from other WAP stocks in future assessments. These data might also give insight into why fish in the North are larger than fish in the South.

CHOICE OF K

- The best fit for mean weight data was achieved when k=5 in the South, and k=4 in the North, but high Fs in these models suggested only keeping runs with k=3. These differences indicate that there is a conflict between the simplified model assumption of knife-edge recruitment at one age only and the assumed growth models.
  - Unfortunately, strong simplifying assumptions are required to model data deficient stock like WAP.
  - Might be able to include k=4 in future assessments if we put an upper bound on F (e.g., $F \le 2$).
  - High F values when k>3 could imply that the stock isn't productive enough. Or more likely that the assumption of constant knife-edge selectivity is violated.
  - Or, natural mortality M might be too low. Note that this assessment used M values from Alaska WAP models; BC could have higher M because of different environmental conditions. Does M reflect changes in predator abundance, such as sea lions?
  - Authors were reluctant to run more model runs at this point, but could explore higher M values in future assessments.
  - Option to only include models with k=3 for model averaging.
- Unanswered question about why the year 2003 has consistently high F values.

DECLINE IN PROJECTED BIOMASS

Model projections consistently show a decline in biomass at current average harvest rates. This could be due to various factors, such as the lack of occasional large recruitments in model projections. These occasional large recruitments may be an important factor for BC WAP.

- These occasional large recruitments don't show up in the short-term projections because the projections are based on average recruitment with added stochasticity. Thus, the projections may only be simulating part of the recruitment process.
- It was noted that a different recruitment error assumption might allow projections to have occasional large recruitments, and be more realistic. Alternatively, model projections could sample/bootstrap historic recruitment. This was suggested to be explored in future assessments.
- General agreement in the meeting that stock projections were unreliable and uncertain, and that there is low confidence in the 3-year projections because of the poor quality of the available data and the uncertainty associated with WAP growth and life history (e.g., WAP generally live about 12 years).
- Large recruitments are estimated to be more common in the North stock.
- Model recruitment is largely driven by the mean weight data and the indices of abundance. Recruitments for projections are based on a mean and assumed CV which may not adequately model this process.
- The meeting participants agreed to only provide 2-year projections in the final research document, and 1-year projections in the SAR.

BIOMASS TRAJECTORIES

- Agreement to include biomass trajectories for all model runs used in the model averaging process because the average does not show the range of potential trajectories for each model run. This will help convey the underlying uncertainty in the stock assessment.
- The axes labelled ‘biomass depletion’ in, for example, figure 7 is a misnomer – this is more properly “biomass relative to a reference point”. References to “depletion” need to be updated throughout the document.

NORTH STOCK IS A SUBSET OF THE ALASKA STOCK

If the North stock is part of a larger SE Alaskan stock, is it reasonable to do an assessment for the North under the assumption that it is an independent stock?

- There are no data to support this hypothesis, but harvesters believe this to be the case. There are also no resources for performing joint stock assessments with the United States agencies.

BURN-IN

Burn-in allows model parameters to converge to plausible values perhaps because they were initialized with implausible values. Sensitivity analysis on initial values would potentially clarify this.

- Agreement to increase the burn-in by discarding another 200 samples (of 1,000) to reduce autocorrelation, and then re-do the model averaging process.
- The authors did this overnight between meeting days, but the results did not change. The meeting participants agreed to keep the versions with 1,000 samples.

DFO’S PRECAUTIONARY APPROACH

The first part of this discussion was general in scope (i.e., not directly related to WAP), and was intended to identify components of DFO’s Precautionary Approach (PA) Framework (DFO 2009) that require review and updating. Generally, it was suggested that the wording of the PA framework does not align with the harvest control rule diagram presented in the PA.

- The PA diagram is outdated in that it doesn’t distinguish between operational control points (OCPs; Cox et al. 2013) and biological reference points.
- OCPs represent triggers for management action (e.g., the stock biomass that triggers cessation of fishing in the harvest control rule).
- The limit reference point (LRP) is defined in the PA framework and in international best practice as a stock biomass below which serious harm can occur to the stock, and which should be avoided with high probability.
- The LRP also delineates the critical zone from the cautious zone in the PA diagram.
- However, as presented, the PA framework shows the LRP as both the stock biomass to be avoided with high probability, and also the trigger for management action to cease fishing. This leads to the question about whether fishing all the way down to the LRP is consistent with the objective of avoiding the LRP with high probability.

- There was also discussion about whether the LRP and upper stock reference (USR), which are defined as management triggers, are also appropriate reference points for delineating the critical and cautions zones, and the cautious and healthy zones, respectively. The critical zone should be defined by biology of the stock rather than management objectives, although there was not complete agreement on this point at the meeting.
- The USR represents management objectives rather than the biology of the stock.

This part of the discussion is specific to the current WAP assessment.

- This fishery should comply with DFO's PA framework.
- In this paper, the LRP is B_{min}, and the USR is $2B_{min}$.
- It is problematic to express the LRP as B_{min}/B_{avg}, because B_{avg} is used both as a proxy for B_0 and for B_{MSY}. However, in other models, B_{MSY} can be roughly $0.5B_0$.In this document: B_{avg} should be a proxy for B_{MSY} only.
- Therefore it does not make sense to express the LRP as the ratio B_{min}/B_{avg}. The authors should express the LRP as the biomass B_{min} instead.
- Authors need to make it clear that the LRP and USR do not define the boundary between the critical, cautious, and healthy zones. Instead, update the figures and text to state that the stock is above/below the USR/LRP. In addition, authors need to clarify the language around the LRP and USR with respect to terminology for B_{min}, B_{avg}, and proxies.
- Authors could define the terms (e.g., LRP, USR, HCR, B_{min}, B_{avg}) in the paper, and explain their significance.
- Authors should include in the document the reason why they chose $2B_{min}$ instead of B_{avg} as the USR.
    - B_{avg} is uncoupled from B_{min}, so it could theoretically be lower – this issue needs to be addressed in the text.

AGE DATA

There was discussion regarding the need to start collecting otoliths to model growth of BC WAP. Currently, research surveys collect pectoral fins for ageing, but fins are not useful for ageing WAP because they do not provide good estimates of maximum age because accretion to the fin slows down as fish approach maximum growth. This issue has been noted in the past, but has not been incorporated into research survey protocol.

- BC WAP are difficult to age (i.e., more difficult than WAP stocks in Alaska), even when sufficient suitable ageing structures and personnel are available. One solution could be to collect both otoliths and fin rays, as was done for Pacific Cod, which is also a species that is difficult to age.
- DFO may have WAP otoliths from previous surveys that could be aged. A request from Fisheries Management would be required to start this process, and could take a year to implement.
- Collecting WAP otoliths may require omitting the collection of another type of data, since surveys are currently fully committed in terms of resources. In addition, the ageing lab would have to be notified so that they can accommodate additional otolith samples.

- The groundfish at-sea observer program (ASOP) is a potential source of WAP otolith data, but this might also require omitting or reducing sampling effort for other species. The ASOP collects WAP otoliths, but WAP is a low priority compared to other species.
- There are challenges with capacity in the groundfish survey group and the ageing lab. There is also a disconnect between SARs and research survey development, in that recommendations that are included in previous SARs have not influenced research surveys.
- Currently, it would be difficult to determine how many WAP otolith samples are required for a growth model, but there is a need to determine whether WAP can be aged reliably using otoliths.
- It is unlikely that sufficient age data could be collected to do an age-structured model for the next stock assessment, especially considering the difficulty of ageing this species in BC waters. However, it would be useful to have sufficient data to model a BC WAP growth curve.

STOCK DELINEATION

There seems to be a mismatch between the biological and management delineations of the North and South stocks.

- For example, there is no TAC for areas 3C and 3D, and TAC for major area 5A includes minor area 12.

NATURAL MORTALITY

The model seems to work better when *M* is higher. For Pacific Cod, *M* was estimated to be 0.4.

- This should be considered in future assessments.
- Also, future assessments might consider an increasing *M* over time, such as on the West Coast of Vancouver Island where the number of predators has increased.

MODELS TO INCLUDE IN THE MODEL AVERAGING PROCESS

Generally, including more models increases the uncertainty associated with the advice. However, including non-credible models will result in poor management advice. Therefore, a decision was made to only include models where the median $F_{max} \leq 2$ or $F_{max} > 2$ in one year only. All included models were required to have suitable autocorrelation and trace plots.

- The following models were selected for inclusion in the model averaging process:
    - BC North: S00, S03, and S10
    - BC South: S00, S01, S03, S04, S06, and S07
- Regarding figures 10 and 11 (horizontal bar plots showing model averages):
    - The x-axis is misleading. The authors need to be clear that B_{avg} is a proxy for B_{MSY}.
    - Remove vertical lines, and put points for each included model.
    - Update the text: replace ‘critical’, ‘cautious’ and ‘healthy’, with above/below the LRP or USR.

- Consensus to remove the ‘Ranking’ column from tables (e.g., table 1). It was decided to omit this table from the SAR, and include instead (if possible, depending on page limits) a small table that shows which model runs were included in the model averaging process.

DECISION TABLES

- Remove some rows from decision tables: keep row with TAC=0, and remove rows with TAC>2500 (or above the highest historical catch), keep rows with TAC around the average catch, and the past TAC.
- Change the word TAC to either Catch or Mortality to be more accurate.
- Decision to format probabilities to two decimal places: more decimals gives a false sense of certainty.

DELAY-DIFFERENCE MODEL

Add a sentence on why a delay-difference model was used. This would probably reflect it has intermediate complexity, and does not require age-structured data.

SOURCES OF UNCERTAINTY

- North stock may be a part of the larger Southeast Alaska stock. Currently this is not known, and is a combination of speculation and harvester knowledge. If it’s true, then a stock assessment for the BC North stock has limited utility.
- This assessment assumes that there are two WAP stocks: the assessment results won’t be reliable if this is not the case. However, there is uncertainty with respect to stock structure. For example, there could be migration between the two stocks. The authors will clarify this issue in the document.
- Discussion on why model projections are uncertain. Agreement to update the text “delay-difference model used in this stock assessment is not capable of making reliable multi-year projections” from “is not capable” to “is less able” (i.e., page iii). This reflects the feeling that the lack of a latent age structure doesn’t prelude these models from making projections

ECOSYSTEM CONSIDERATIONS

Consensus to omit this section because it is not required, and is not covered in the working paper.

TERMS OF REFERENCE

The group reached consensus that the authors had met each TOR objective (Appendix A).

- Objective 5 required some explanation: “identify additional information needed to enhance appropriate stock assessment advice consistent with goal of implementing ecosystem-based fisheries management, as articulated in the Sustainable Fisheries Framework.”
    - Discussion about the intent of this objective, and agreement that it did not include doing ecosystem-based fisheries management (EBFM) for WAP at this point.
    - The Minister’s mandate specifies that EBFM should be a part of Science advice, but Science has not received direction on this from the Ecosystems group.

- It was noted that age data would help improve the WAP assessment, and be consistent with EBFM.
- Items that relate to EBFM include closed areas, natural mortality, sea lion predation, and this stock assessment, which is the first quantitative assessment for BC WAP.
- Discussion on whether ‘bycatch’ fits into EBFM, which could alternatively be under ‘Ecosystem considerations.’
- Consensus that the authors did the best they could with the available data.

CONCLUSIONS

The working paper is accepted with the aforementioned revisions, as well as minor and editorial comments by the reviewers and participants.

- The Terms of Reference objectives (Appendix A) were achieved.
- Three model runs for the North stock and six model runs for the South stock will be used in the model averaging process, which will provide management advice.
- This assessment considers WAP from two BC stocks: North and South. Management advice is provided at this spatial scale. This assessment does not include Strait of Georgia WAP.
- Given uncertainty in growth, natural mortality M, and selectivity k, improved biological sampling and data is recommended before the next assessment.

RECOMMENDATIONS & ADVICE

There is a need for ageing data from BC WAP. Ideally this would be sufficient data for an age-structured assessment model, but even acquiring enough data for a growth model would be an improvement. Acquiring ageing data requires several steps, including Fisheries Management requesting the data, providing advance warning to the ageing lab of additional work and possibly training, updating survey index sampling protocols, and determining the required ageing structures (i.e., otoliths and/or fin rays).

FUTURE WORK

- Biological data is needed for the North and South stocks: length, weight, and age. This includes the collection and analysis of more age structures than are currently collected and analysed (i.e., otoliths), and would improve future WAP stock assessments.
- BC WAP could be a good candidate for management strategy evaluation (MSE) using the Data-Limited Modelling tool (DLMtool).
- Estimate standard deviations for process and observation error instead of using fixed values.
- Do retrospective analyses to determine the reliability of biomass projections.
- Add a fourth hypothesis regarding size differences between North and South stocks: The smaller South stock could be outmigration from the Strait of Georgia.
- Improve biomass projections by, among other things, updating the assessment model.

ACKNOWLEDGEMENTS

Thanks to Sarah Hawkshaw for taking notes during the meeting, and thanks to the reviewers Doug Swain and Sean Anderson for providing written and oral reviews. Thanks also to the participants for their time and contributions.

REFERENCES CITED

Cox, S.P., Kronlund, A.R., and Benson, A.J. 2013. The roles of biological reference points and operational control points in management procedures for the sablefish (*Anoplopoma fimbria*) fishery in British Columbia, Canada. Environmental Conservation 40, 318-328.

DFO 2009. A fishery decision-making approach incorporating the Precautionary Approach. Fisheries and Oceans Canada.

APPENDIX A: TERMS OF REFERENCE

WALLEYE POLLOCK (*THERAGRA CHALCOGRAMMA*) STOCK ASSESSMENT FOR BRITISH COLUMBIA IN 2016

Regional Peer Review Process – Pacific Region

November 14-15, 2017
Nanaimo, BC

Chairperson: Dominique Bureau

Context

The last assessment of Walleye Pollock, conducted in 1997, did not estimate stock status relative to reference points that are consistent with the DFO's Fishery Decision-Making Framework Incorporating the Precautionary Approach (DFO Precautionary Approach) (DFO 2009). Harvest advice is required for this species to determine if current harvest levels are sustainable and are compliant with the DFO Precautionary Approach. A 2013 request to reassess Walleye Pollock was considered unachievable at the time due to data limitations.

The Walleye Pollock fishery in Alaska is substantially larger than the Walleye Pollock fishery in BC, and supports the largest landings of any single species in the USA. The highest concentrations of Walleye Pollock occur in the Eastern Bering Sea (annual catches > 1 million tonnes), with lesser amounts caught in the Gulf of Alaska (annual catches > 100,000 tonnes). These Alaskan fisheries are supported by complex, age-structured models; with a strong investment in acoustic and trawl surveys, as well as catch sampling and an intensive ageing program. Data available from the Alaskan fisheries can be utilized for the BC Walleye Pollock assessment (Dorn et al. 2012, Dorn et al. 2015, Ianelli et al. 2015).

Fisheries and Oceans Canada (DFO) Fisheries Management has requested DFO Science provide advice regarding the assessment of the coastwide Walleye Pollock stock relative to reference points that are consistent with the DFO's Precautionary Approach, and implications of varying harvest rates on expected stock status. The advice arising from this Canadian Science Advisory Secretariat (CSAS) Regional Peer Review (RPR) will be used to inform fisheries management decisions. This work may also inform and supplement decisions external to DFO; including those on Marine Stewardship Council (MSC) certification and trans-boundary fisheries management.

Objectives

The following working paper will be reviewed and provide the basis for discussion and advice on the specific objectives outlined below.

Starr, P.J., Haigh, R. Stock assessment for Walleye Pollock (*Theragra chalcogramma*) in British Columbia in 2016. CSAP Working Paper 2013GRF03.

Guided by the DFO Sustainable Fisheries Framework, the following objectives for this assessment have been established:

1. assess the current biomass and status of Walleye Pollock (*Theragra chalcogramma*) for BC waters (the number of stock assessments will depend on the stock structure determination and the availability/suitability of data);

2. suggest candidate reference points (either biomass-based or fishing mortality-based) consistent with the DFO Precautionary Approach, including alternatives to model-based reference points;
3. provide the rationale used to select recommended candidate reference points, including reasons if the candidate points differ from the PA framework default reference points;
4. provide decision tables forecasting the predicted status of Walleye Pollock relative to reference points across a range of management actions;
5. identify additional information needed to enhance appropriate stock assessment advice consistent with goal of implementing ecosystem-based fisheries management, as articulated in the Sustainable Fisheries Framework;
6. if possible, propose an appropriate time interval between assessments, and a trigger mechanism that may affect the assessment schedule; or, provide rationale why this is not possible.

Expected Publications

- Science Advisory Report
- Proceedings
- Research Document

Participation

- Fisheries and Oceans Canada (DFO) (Science and Groundfish Fisheries Management)
- Commercial and Recreational Fishing Representatives
- Environmental Non-government Organizations
- Academia

References

DFO 2009. A Fishery Decision-Making Framework Incorporating the Precautionary Approach. (Accessed March 31, 2017)

Dorn, M., Aydin K., Barbeaux, S., Jones, D., McCarthy, A., Spalinger, K., and Palsson, W. 2012. Assessment of the Walleye Pollock stock in the Gulf of Alaska. In: Stock Assessment and Fishery Evaluation Report for Groundfish Resources of the Gulf of Alaska. Prepared by the Gulf of Alaska Groundfish Plan Team, North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. North Pacific Fisheries Management Council, Anchorage, AK. 95 p. + 5 Appendices.

Dorn, M., Aydin, K., Jones, D., McCarthy, A., Palsson, W., and Spalinger, K. 2015. Assessment of the Walleye Pollock stock in the Gulf of Alaska. In: Stock Assessment and Fishery Evaluation Report for Groundfish Resources of the Gulf of Alaska. Prepared by the Gulf of Alaska Groundfish Plan Team, North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. North Pacific Fisheries Management Council, Anchorage, AK. 158 p. + 4 Appendices.

Ianelli, J.N., Honkalehto, T., Barbeaux, S., and Kotwicki, S. 2015. Chapter 1. Assessment of the Walleye Pollock Stock in the Eastern Bering Sea. NPFMC Bering Sea and Aleutian Islands SAFE, North Pacific Fisheries Management Council, 53-152.

APPENDIX B: WORKING PAPER ABSTRACT

A new stock assessment is presented for two British Columbia (BC) stocks of Walleye Pollock (WAP, *Theragra chalcogramma*), with the BC North stock encompassing the three most northerly Pacific Marine Fisheries Commission (PMFC) major areas (5C, 5D, 5E) and the BC South including the remaining four outside PMFC major areas (3C, 3D, 5A, 5B plus minor areas 12 & 20). These stock definitions were selected on the basis of a difference in observed mean weights, with the BC North mean weights estimated near 1.0 kg/fish while the equivalent BC South mean weights averaged near 0.5 kg/fish. A delay-difference production model was used to assess each stock, using data from fishery-independent surveys, a CPUE series derived from commercial catch rates, and an annual mean weight series derived from unsorted commercial catch samples. Because there are no useable BC ageing data, we used survey age samples from the Gulf of Alaska (GoA) to specify growth for the BC North stock. The BC South proved more problematic, with the GoA growth model unable to fit the BC South observed mean weights, eventually requiring us to use a published WAP growth model from the Asian Sea of Okhotsk. Each stock assessment explored a range of plausible natural mortality values as well as a range of ages for the knife-edge selectivity assumption because the biomass indices and the mean weight data used in the delay-difference model were not informative for these parameters. The stock assessment was conducted in a Bayesian framework, where the best fit to the data was used as the starting point for a search across the joint posterior parameter distributions using the Monte Carlo Markov Chain (MCMC) procedure. Twelve runs were made for the BC North stock and 11 for the BC South stock, with each run consisting of 60 million MCMC iterations, sampling every 50,000th iteration, discarding the first 200 draws for burn-in, leaving 1,000 draws to comprise the posterior. Composite reference (model averaged) scenarios were used to represent each stock, with the model average for both stocks consisting of eight model runs selected on the basis of a subjective evaluation of the quality of the MCMC posterior. Each composite reference scenario included three values for instantaneous natural mortality (M=0.25, 0.30, 0.35) and covered two or three ages at which knife-edge recruitment (k) to the fishery occurred (k=3, 4 in BC North and including k=5 in BC South). The MCMC posteriors for the two composite scenarios were constructed by pooling the 1000 MCMC samples from each of the selected runs to give a posterior of 8,000 samples, thus giving equal weight to each run. The composite reference scenario was evaluated against historical reference points (HRPs) based on the reconstructed spawning biomass trajectory due to concerns about the stability of estimating B_0 and B_{2017}. The HRP B_{avg}, the average spawning biomass from 1967-2016, was used in place of B_0, and B_{min}, the minimum spawning biomass from which it subsequently recovered to B_{avg}, was used in place of 0.4B_{MSY}. Twice B_{min} was used in place of 0.8B_{MSY}. The average exploitation rate over the period 1967-2016 (u_{avg}) was used in place of u_{MSY}. The model average BC North stock was evaluated as being primarily in the "healthy zone" (>0.8B_{MSY}) while the BC South stock was evaluated as being entirely in the "healthy zone". For each stock, the assessment provides a decision table which evaluates the probability of the model average case staying above five reference points across a wide range of 22 constant catches. However, the paper warns that the probabilities in these decision tables should be viewed cautiously as the delay-difference model used in this stock assessment is not capable of making reliable multi-year projections because it has no latent age structure to inform predictions and the stock-recruitment function is poorly determined.

APPENDIX C: AGENDA

Canadian Science Advisory Secretariat

Centre for Science Advice Pacific

Regional Peer Review Meeting (RPR)

Walleye Pollock (*Theragra chalcogramma*) stock assessment for British Columbia in 2016.

November 14-15, 2017

Pacific Biological Station, Seminar Room

Chair: Dominique Bureau

DAY 1 – Tuesday, November 14, 2017

Time	Subject	Presenter
09:00	Introductions Review Agenda & Housekeeping CSAS Overview and Procedures	Chair
09:15	Review Terms of Reference	Chair
09:30	Presentation of Working Paper	Rowan Haigh, Paul Starr
11:00	***Break***	
11:15	Overview Written Reviews	Chair + Reviewers & Authors
12:00	***Lunch Break***	
13:00	Overview Written Reviews	Chair + Reviewers & Authors
13:30	Discussion & Resolution of Technical Issues	RPR Participants
14:45	***Break***	
15:00	Discussion & Resolution of Technical Issues	RPR Participants
16:15	Check in on progress and confirmation of topics for discussion on Day 2	RPR Participants
16:30	Adjourn for the Day	

DAY 2 – Wednesday, November 15, 2017

Time	Subject	Presenter
09:00	Introductions Review Agenda & Housekeeping Review Status of Day 1	Chair
09:15	Discussion & Resolution of Technical Issues (Continued from Day 1)	RPR Participants
10:30	***Break***	
10:45	Discussion and Resolution of Working Paper Conclusions	
11:30	Develop Consensus on Paper Acceptability & Agreed-upon Revisions	RPR Participants
12:00	***Lunch Break***	
13:00	*Science Advisory Report (SAR)* Develop consensus on the following for inclusion: • Sources of Uncertainty • Results & Conclusions • Additional advice to Management (as warranted)	RPR Participants
14:30	***Break***	
14:45	*Science Advisory Report (SAR)* (Continued)	RPR Participants
16:15	Next Steps – Chair to review • SAR review/approval process and timelines • Research Document & Proceedings timelines • Other follow-up or commitments (*as necessary)*	Chair
16:30	***Adjourn meeting***	

APPENDIX D: PARTICIPANTS

Last Name	First Name	Affiliation
Anderson	Sean	DFO Science
Bureau	Dominique	DFO Science
Christensen	Lisa	DFO Science
Forrest	Robyn	DFO Science
Fu	Caihong	DFO Science
Grandin	Chris	DFO Science
Grinnell	Matthew	DFO Science
Haigh	Rowan	DFO Science
Hawkshaw	Sarah	DFO Science
Lochead	Janet	DFO Science
MacDougall	Lesley	DFO Science
Mose	Brian	CIC Trawl
Obradovich	Shannon	DFO Science
Olsen	Norm	DFO Science
Starr	Paul	Groundfish Conservation Society
Surry	Maria	DFO Science
Swain	Doug	DFO Science
Tadey	Rob	DFO Fisheries Management
Turris	Bruce	Groundfish Conservation Society

APPENDIX E: WRITTEN REVIEW BY DOUG SWAIN

Date: November 7, 2017

Reviewer: Doug Swain, Fisheries and Oceans Canada, Gulf Fisheries Centre, Moncton, NB

Stock structure

BC walleye pollock (WAP) were treated as two stocks for this assessment, a North stock distributed around Haida Gwaii (5CDE) and a South stock in Queen Charlotte Sound and off the west coast of Vancouver Island (3CD5AB + minor areas 12 and 20). This decision was based on a large difference in mean weight between fishery catches in the North and the South, with northern fish averaging twice the mean weight of southern fish (about 1.0 versus 0.5 kg/fish).

The authors offer three hypotheses for the cause of this large difference: A) true growth differences between discrete northern and southern populations, B) northern migration of large old fish, and C) higher exploitation rates in the south. The authors prefer hypothesis A, though they cannot rule out hypothesis B. As evidence against hypothesis B, they note that many small fish occur in GoA survey data. Is this also the case in survey data from 5CDE? Another hypothesis could be that this difference reflects a difference in selectivity between the northern and southern fisheries. Is a similar difference observed in survey catches?

The evidence for this difference in size appears to be largely restricted to the period since the late 1980s (Table A.6). There is no clear difference in mean weight between these 2 areas in 1972-1979, when mean weight in both areas was > 1 kg. However, mean weights in the south have been much smaller since 1988 than they were in the 1970s. Can the authors offer an explanation for the large difference in mean weight between the two areas since the late 1990s but not in the 1970s? The authors have made a reasonable choice for the stock structure used in this assessment, but further work is needed to understand the causes of these striking differences in mean size between and within areas.

The authors note that the North stock likely belongs to a larger stock extending into the waters off southeast Alaska. Is there a concern that the assessed stock is not a closed population, so that biomass indices may be biased due to movement of pollock between the assessed area and the portion of the population further north?

Assessment Approach

The authors indicate that there is very little reliable ageing data for BC Pollock. This precludes using a fully age-structured assessment model. Instead, the authors chose to use a delay-difference model. This model incorporates more realistic dynamics than the simpler alternative, a surplus production model. Unlike surplus production models which aggregate the three components of productivity (rates of recruitment, individual growth and adult natural mortality) into a single parameter, delay-difference models treat each of these processes separately. This provides the model with greater flexibility to fit biomass changes in the indices, and also potentially provides greater scope for understanding the causes of biomass changes. However, this advantage comes at the cost of greater requirements in terms of model inputs. In addition to fishery catch and biomass indices, delay-difference models require inputs needed to develop a recruitment model, a model for individual growth and a value for adult natural mortality. In addition the models make the strong (and unrealistic) assumption that selectivity to the fishery and the surveys is knife-edged with selectivity increasing from 0 to 1 at age k. Age k is also assumed to be the age at maturity. Thus, all mature fish (all fish aged k and older) are assumed to be fully vulnerable to the fishery and surveys, and all juvenile fish (all fish aged less than k) are assumed to be invulnerable. Model success depends on an appropriate choice for k. Lacking this information for BC pollock, life history information was borrowed from other WAP stocks. In my view, the choices made generally seem appropriate, as follows.

Based on the value used in Alaskan assessments of WAP, the instantaneous rate of adult natural mortality, M, was assumed to be 0.3 for both stocks and was a fixed input to models. Sensitivity analyses were also conducted using values of 0.25 and 0.35.

Recruitment was modelled as follows. R_0, equilibrium unfished age-0 recruits, was a parameter, estimated on the $\log_e$ scale. Average recruitment was assumed to equal R_0, and R_t, recruitment in year t, was given by R_0 times a lognormally-distributed deviate with a mean on the log scale of 0 and σ_R=0.6. Predicted values for R_t, i.e.$\hat{R}_t$, were obtained using a Beverton-Holt recruitment model. Differences between R_t and its predicted value from the BH model contributed a component of the objective function, constraining recruitment to be consistent with Beverton-Holt dynamics. Steepness *h* was an estimated parameter with an informative prior based on previous work by Forrest et al. (2010 and 2015).

A von Bertalanffy growth model developed for eastern Gulf of Alaska (GoA) WAP was used for the North stock. The authors concluded that this model adequately fit the mean weight data for the North stock assuming k equalled 3, 4 or 5, though I would argue that the fit was good with k=4 and less good with the alternate values. This model was inappropriate for the smaller fish in the South stock. A model developed for WAP in the Sea of Okhotsk was used for the South stock. It fit better than the GoA model, but was unable to fit the high mean weights estimated for the South stock around 1980 and in the late 1980s.

The age k of knife-edged recruitment to the fisheries and surveys was selected based on estimated selectivity ogives for GoA fisheries and surveys. Ages 3 and 4 were chosen as the most likely ages for k, apparently based on the ages near 50% selectivity for these ogives. Sensitivity analyses with k=5 were also conducted given the slower growth of the South stock.

Survey Biomass Indices

Four survey indices were used for the North stock, abbreviated here as the GIG, HS assemblage, HS synoptic and WCHG synoptic surveys, and three for the South stock – the GIG, QCS synoptic and the WCVI synoptic surveys. Little information is provided on consistency of catchability within these surveys. Four vessels contributed to the GIG survey, 5 vessels to the HS assemblage survey, 2 vessels to the HS synoptic survey, 4 vessels to the WCHG synoptic survey, 3 vessels to the QCS synoptic survey, and 1 vessel to the WCVI synoptic survey. Was the same trawl used in all years in all surveys? (What was/were the trawl(s) used and did they use a liner?). Were the different vessels within a survey calibrated using paired tows at the same locations? If not, were they similar in size and power?

For both stocks, none of the surveys cover the entire stock area, yet they are each used as an index of biomass changes over the entire area. This requires the assumption that there have been no important changes in WAP distribution within the stock area over the time span of these surveys. Is there evidence to support this critical assumption?

For the South stock, the QCS and WCVI surveys overlap in time. The two surveys show similar trends in time, with low values prior to 2010 and high values since then. Thus, in this case, there is support for the assumption that the surveys in separate portions of the stock area reflect biomass trends over the whole stock area. On the other hand, the GIG survey covers only a small portion of the South stock area and I found no evidence in the assessment document for or against the assumption that the biomass trends in this small area reflect the trends over the whole stock area.

For the North stock, this assumption cannot be tested among the GIG, HS assemblage and HS synoptic surveys, which overlap little or not at all in time. The HS and WCHG synoptic surveys do overlap in time. The WCHG index tends to increase from 2007 to 2012. Given the high uncertainty in the HS indices, it is difficult to identify any temporal trend in this index. So there is no strong evidence either against or in support of the assumption that biomass trends within the survey areas reflect biomass trends within the stock area as a whole. Furthermore, from my point of view, it is difficult to defend the use of the GIG index of the South stock in the analysis of the North stock.

In the modelling, the biomass indices are assumed to consist entirely of fish k years of age and older. However, in the assessment document there is no indication that the survey catches have been subsampled to make this a reasonable assumption. There is no reliable age data for these surveys but presumably the length composition of the survey catches is well known. For example, are length frequency distributions obtained for a sample of the WAP catch in each tow? If so, the growth model for each stock could be used to determine an appropriate length cut-off for age k and the indices could be restricted to the portion of the survey catch above that cut-off. Even given the uncertainties in the growth models, I would argue that this approach would be more appropriate than including all fish caught (including many fish smaller than age k fish) in the index.

Fishery CPUE Indices

For each stock, separate CPUE indices were developed for mid-water and bottom trawls. Catches were standardized to control for variation associated with fishing location (DFO locality, latitude), depth, season (month), hours fished and, for bottom trawls, vessel. A year term was incorporated first in the model and provided the annual index. For bottom trawls, presence/absence of WAP was modelled using a binomial logit model and, where WAP were caught, catch was modelled using a lognormal model. Results of the two models were combined into a single index using the delta distribution. For mid-water trawls, analyses presented in the document did not proceed beyond the lognormal model. Results of this model were not considered reliable because residuals were not well fit by a lognormal distribution and there were strong abrupt annual shifts in the data. In contrast, fit and performance of the standardization models for bottom trawls were considered adequate for bottom trawls, and indices were developed for both the North and South regions.

As noted by the authors, fishery CPUE indices need to be treated with caution as indices of biomass even when standardized. This is because many of the factors affecting fishing success are not amenable to statistical standardization. These factors include technological improvements such as improved fish locating equipment and gear positioning equipment. Catchability can also change due to fish behaviour such as increased concentration in fish distribution as fish abundance declines. Thus, it is important to examine the consistency between fishery cpue and research survey indices.

There is some evidence for consistency between these two sources of biomass indices for both the South and North stocks. In the South, both the QCS and the WCVI indices were at low levels in about 2003 to 2009 and at a higher level since 2010. The fishery cpue index for bottom-trawls was likewise at a low level from 2000 to 2009 and then exhibited a an increasing trend between 2010 and 2015. This gradual increase appears to contrast the sudden increase in the survey indices. However, the sparse survey data and its high uncertainty do not preclude a more gradual increase. Inclusion of the CPUE index has an obvious effect on the fit to the both the QCS and WCVI survey indices. With the CPUE index excluded, there is a sharp jump in the predicted values for these indices in 2010 or 2011, matching the possible year effects in these years. Including the CPUE index, the increase in the predicted survey indices is more gradual (which seems more plausible). In the North, the bottom-trawl cpue index declined in the late 1990s. This is not inconsistent with the HS assemblage survey data. The cpue index gradually increased from about 2001 to 2015, again not inconsistent with the indices from the HS and WCHG synoptic surveys.

Model Results

A large number of models were fit for each stock. These models compared results for different plausible values for M (0.25, 0.30 or 0.35) and k (age 3, 4 or 5) and compared results including or excluding the fishery CPUE index and, for the North stock, the GIG survey. For each stock, a subset of models was determined to be acceptable based on a subjective evaluation of their fit to the indices and mean weight data. The MCMC samples from all acceptable models were then included in analyses to estimate the probabilities of various outcomes (e.g., $B_{2020} > B_{2017}$) given different levels of future catch. This seems to me to be an excellent way to incorporate model uncertainty in management advice.

The fit of most of the models was deemed acceptable and most were thus included in the "model averaging". However, I think that the credibility of model estimates also needs to be considered when selecting acceptable models. For most models, the maximum F estimates were impossibly high. For the North stock, 9 of 12 models had maximum F estimates in the 8 to 20 range. I regard only two or three of these models as plausible (Table 2): S00 (M.30+K3), S03 (M.35+K3) and possibly S01 (M.25+K3). I also question the use of the "southern" GIG survey in models for the North stock. Unfortunately, no North model omitting the GIG survey was plausible. For the South stock, I regard only three models as plausible (Table 4): S00 (M.30+K3), S03 (M.25+K3) and S06 (M.35+K3). All the remaining models had impossibly high maximum F estimates ranging from 14 to 20. In my view, these models with impossibly high maximum F estimates are not credible and need to be excluded from the "model averaging".

Management Advice

Clear management advice is provided based on the population modelling. Uncertainties are well described. A suite of models were conducted, examining the sensitivity of results to a range of plausible values for key model parameters (M,k) and the inclusion or exclusion of particular indices. Based on adequate fit to the data, most of these alternate models were incorporated in the advice by including their MCMC samples in the calculation of decision tables describing the estimated probabilities of future outcomes at different levels of future catch. Including the results of a range of plausible alternate models in these calculations is a good way of including model uncertainty in advice. However, I think that the credibility of model results also needs to be considered in deciding on the models to incorporate in the advice. In this case, I think that many of the models incorporated in the advice are not credible and should be excluded from the calculation of the probabilities in the decision tables.

Minor comments

1. . I found the procedure for partitioning total variance between observation and process error a bit confusing. Observation error consists of σ_O, σ_C and σ_W. Process error = σ_R. σ_O is defined as the overall standard deviation of observation residuals. But actually it appears to be a component of the σ's of the indices (E.29). In Table E1, all the σ's appear to be assigned fixed values (except σ_C which is presumably assigned a small value). So why is there a need for ϕ, ρ, E.27 and E.28? Since the process error σ is assigned a fixed

value, as are σ_W and presumably σ_C, why not estimate σ_O? After weighting σ_O (0.2) by the CV of each index observation, the index σ's are effectively quite small. It might be a good idea to let the model estimate at least this component of error to see whether such low values of $\sigma_O CV_{jt}$ are justifiable.

2. . Why are the log q's included as parameters with uniform priors from -10 to 0? The log q's are calculated in E.31: $\bar{z}_j = \ln(q_j)$

References

Forrest, R.E., McAllister, M.K., Dorn, M.W., Martell, S.J.D., and Stanley, R.D. 2010. Hierarchical Bayesian estimation of recruitment parameters and reference points for Pacific rockfishes (*Sebastes* spp.) under alternative assumptions about the stock-recruit function. Can. J. Fish. Aquat. Sci. 67: 1611-1634.

Forrest, R.E., Rutherford, K.L., Lacko, L., Kronlund, A.R., Starr, P.J., and McClelland, E.K. 2015. Assessment of Pacific Cod (*Gadus macrocephalus*) for Hecate Strait (5CD) and Queen Charlotte Sound (5AB) in 2013. DFO Can. Sci. Advis. Sec. Res. Doc. 2015/052. xii +197 pp.

APPENDIX F: WRITTEN REVIEW BY SEAN ANDERSON

Date: November 7 2017

Reviewer: Sean C. Anderson, Offshore Section, Pacific Biological Station, DFO CSAS
Working Paper: 2013GRF03

Working Paper Title: Walleye Pollock (*Theragra chalcogramma*) stock assessment for British Columbia in 2017

Overall, the Working Paper is well written and decisions relating to data selection and modelling choices are generally well reasoned. The authors have attempted to make the best of the available data and make up for the lack of reliable ageing data for these stocks by borrowing ageing data from the Gulf of Alaska and a growth model from a population in the Asian Sea of Okhotsk. The authors fit a delay-difference model and deal with uncertainties related to the age of assumed knife-edged selectivity, natural mortality, and the inclusion or exclusion of various data sources by fitting multiple models and considering an averaged result. I could suggest additional sensitivity analyses but the authors have already explored many models, the working paper is already extensive and quite complex, and I don't think there would be substantial added value in further complicating the Working Paper. Below I provide a number of more specific comments for discussion and consideration.

Comments for discussion, consideration, or acknowledgement at the CSAP review meeting

1. Section 6.4: The authors raise the hypothesis that older and larger fish may migrate into the north population and acknowledge that we cannot rule this out. Could the authors include discussion of how this hypothesis might affect the model interpretation and decision tables?
2. Page 12: subjective ranking of MCMC convergence. I agree that there is always benefit to graphically inspecting diagnostic plots such as trace plots, but why not also consider well-established metrics of chain convergence? For example, effective sample size (neff) and the potential scale reduction factor ($\hat{R}$) (Gelman et al. 2014, Bayesian Data Analysis V3, p 286–287). The thresholds described in the Gelman et al. book are commonly used in the literature: $\hat{R} < 1.1$ and neff > 100 for reasonable inference about the mean. In particular, it would be useful to know about the sampling stabilty of derived parameters of interest such as B2017/Bavg. If the MCMC sampling for these models can be performed in a reasonable amount of time, then perhaps longer runs would result in reasonable convergence for all or more models.
3. One of my overall concerns is whether there is any bias in the models that were used in the averaging due to the selection according to subjective MCMC convergence. The use of quantitative well-established metrics of chain convergence would help. Also, perhaps including the MPD estimates in Figure 10 and 11 (including for those for which MCMC results shouldn't be shown) would help convince the reader that there isn't substantial bias due to excluding the models rejected due to insufficient sampling.
4. Table 6 (and other tables): Providing probabilities up to 6 decimal places greatly over-states the level of precision available here. There are numerous parameters that have been fixed at arbitrary values, uncertainty that is not carried through from fitted models (e.g. growth parameters), uncertainty from the finite MCMC sampling, and above all that, structural uncertainty about the degree to which the fitted model represents the true system dynamics. This is, of course, inherent to nearly all stock assessments, but

presenting results to so many decimal places gives a false sense of precision to managers. I would rather see two — three at most.

5. I am having trouble following Eq. C.4, which combines the lognormal and binomial index. I am familiar with combining the proportion estimate for a given year from a binomial model with the mean estimate for a given year from a lognormal/Gamma model by multiplying the two. Vignaux (1994, cited in the Working Paper) combines the two indices as the following, where POS_index = the positive component index on the response scale and ODDS_index = the binomial index on the odds scale:
6. POS_index / (1 - P0(1 - ODDS_index))
7. but C.4 would have 1/B_Y_y = odds index or exp(log odds index). Perhaps the authors could clarify what scale the binomial indices are on.
8. 9.3 Assessment schedule: “Although advice for the interim years is explicitly included in this assessment in the form of decision tables. . . ” Advice is only provided for 3 of the 5 suggested years until the next assessment. I assume 3 years was chosen instead of 5 because of lack of confidence in long projections given the lack of age structure (which is fine), but I just wanted to make sure this was acknowledged and OK with CSAP given that the suggested plan is to have 2 years without science advice on harvest levels.
9. The authors note that many models estimate implausibly high maximum rates of fishing mortality (Fmax) for most models with k > 3 and we can see this in Tables F.4 (north) and F.23 (south) with some hitting the boundary at 20 and the MCMC chains undoubtedly hitting the boundary. This is in spite of a normal(log(0.2), 4) penalty function on Ft in log space. Does it make sense to include these models that are not credible in the averaged result? Additionally, if reference point estimates from these models remain in the main document, then perhaps caveats related to this should be more strongly emphasized.I would have expected uncertainty in the projected biomass estimates (e.g. Figs F.8, F.31) to become wider each year, since as I understand it, the projections include process error and the error should compound each year. This reflects our natural intuition about uncertainty — longer projections are more uncertain. Why is this is not the case here?
10. I think there is some danger in presenting the model-averaged results as the main result, since the truth may very well be at one of the extremes and not be well represented by averaging the various hypotheses. This assessment does retain decision tables for the multiple hypotheses in the supplementary tables in Appendix F, but perhaps including the range of probabilities that B2020 > LRP and B2020 > USR across the various hypotheses in the main body of the assessment would help emphasize the range of possibilities across the various models. Alternatively, if the authors narrow the selection of models to average (e.g. excluding models if Fmax is too high), and the estimates are more similar across models, this might become less of an issue.
11. Dropping the CPUE index for the north stock results in a very different estimate of B2017/Bavg (e.g., 0.57 vs. 1.30 in Table F.4), which then gets included as 2/8 of the models for the north stock. The explanation given is “because there is uncertainty as to whether fishery-dependent data track abundance.” It seems like this is a fairly influential decision for the composite results. Besides the usual general concerns about CPUE, are there specific reasons for expecting standardized CPUE to be a biased estimate of abundance through time in this case? To my eye, the north bottom-trawl CPUE trend in Fig. C.40 looks plausibly consistent with the Hecate Strait surveys (Fig. B.24 and B.33 except for the high but uncertain survey index in 1998) and the West Coast Haida Gwaii index (Fig. B.63). Even with CPUE excluded, the model is not able to account for those

high survey index values in the late 1990s in the Hecate Strait (Fig. F.14). Perhaps some of the strong CPUE influence comes from the chosen CV of 0.3, which looks relatively tight compared to the CVs on recent survey index trends in HS and WCHG.

More minor comments

1. I may have missed it, but is there a reason some combinations of M and k are not explored? M.25+k4+north, M.35+k3-GIG-north?
2. Page 15: The paragraph at the bottom of this page seems to conflate model fit (e.g. negative log likelihood) with MCMC sampling properties. The sampling performance is the product of many factors including the choice of sampling algorithm (of which the Metropolis Hastings is among the least efficient), the model parameterization, prior choice, sampling tuning parameters, and length of warm up.
3. Page 97: “It is always preferable to standardise for as many factors as possible when using CPUE as a proxy for abundance”. Theoretically, yes, but in reality this is not the case. For example, the analysis within this working paper chooses not to standardize for as many factors as possible in an attempt to achieve some level of model parsimony through model selection.
4. Section C.2.4 Binomial model: “Such a model provides an alternative series of standardised coefficients of relative annual changes that is analogous to the series estimated from the lognormal regression.” Perhaps the authors could be explicit here that the model estimates the proportion of fishing events encountering pollock in a given year. Or are the indices reported as log odds or odds? And then how are they scaled for plotting?
5. Table F.4, F.5, F.9 etc.: if the MCMC sampling is deemed to not have converged, then I would not include parameter estimates or decision tables for these models. The values we are seeing are dependent on this specific random model run. Hopefully longer chains and/or longer warmups will eliminate or reduce this issue, maybe combined with not MCMC sampling for models with implausibly high MPD Fmax values. Some trace plots, such as for log(R0) in M.25+k3 North (Fig. F.18) suggest that a longer warmup/burn-in would be helpful to reduce the MPD starting point influence.
6. In Figure 10 and 11, the LRP and USR are actually unique for each model, correct? I assume Figure 10 and 11 show the reference points for the average scenario. I wonder if it would be worth showing the model-specific reference points, since these may be quite different. For instance, the example south case should have a LRP of roughly 16655/89549 = 0.19 (Table F.20), which is double that shown in Figure 11.

Responses to the five general questions:

1. Is the purpose of the working paper clearly stated?

   Yes.

2. Are the data and methods adequate to support the conclusions?

   Yes, for the most part, with the exception of minor comments made above.

3. Are the data and methods explained in sufficient detail to properly evaluate the conclusions?

   Yes, for the most part, with the exception of minor comments made above.

4. If the document presents advice to decision-makers, are the recommendations provided in a usable form, and does the advice reflect the uncertainty in the data, analysis or process?

   Yes, for the most part, with the exception of minor comments made above.

5. Can you suggest additional areas of research that are needed to improve our assessment abilities?

   As the authors note, reliable ageing data is a high priority to improve assessment ability for the stock. Additionally, genetic testing may help establish north-south population structure and the degree of migration.

Closed-loop simulation, or even just fitting models to simulated data with mismatches between the data-generating process and the model, could be used to explore the impact of various plausible structural and parameter mismatches between reality and the model. For example, the effects of knife-edged maturity assumptions, assumed M values, migration from the south to the north, the impact of possible ‘rescue’ from the Alaskan stock, and the value that various quantities of ageing data would provide could all be explored through simulation.

Minor copy-editing comments are attached separately.