Center for Independent Experts (CIE) Report

External Independent Peer Review by the Center for Independent Experts - Virtual Panel Review of the Stock Assessment for Pacific Cod in the Eastern Bering Sea April 26-30, 2021

Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration

Dr.Sc. MSc. BSc. Henrik Sparholt
Independent Fisheries Scientist
Member of the Nordic Marine Think Tank (NMTT) and Postdoc at the University of Copenhagen
University of Copenhagen
Fredsvej 8a, 2840 Holte, Denmark
Email: henrik.sparholt@gmail.com

Contents

Executive Summary

The Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration held a stock assessment review panel virtual meeting for Pacific Cod in the Eastern Bering Sea in April 14-18, 2021.

The technical review took place during a formal, public, multiple-day meetings of fishery stock assessment experts. I participated as an external, independent reviewer.

Due to the Coronavirus pandemic, it was held as a virtual meeting with the panel members participating from home via PC. We met each day from 9:00 to 15:00 Seattle time, which was 18:00-24:00 Central European time. This worked out quite well because all the presentations with voice recordings were sent out to the participants in advance of the meeting so that we could all prepare well for the meeting. The assessment documents were tidy and well-structured which also helped. However, most of the normal one-to-one informal interactions between participants were lacking, which under normal circumstances are important for a comprehensive exchange of views, ideas, and opinions.

The first day and a half was used to ask question and discuss the ten presentations made available to the panel in advance of the meeting. They covered the following issues: survey data, ageing, North Pacific Observer Program, catch accounting system and in-season management, tagging, somatic growth variability, movement models, VAST fishery CPUE model, Ecosystem and Socioeconomic Profile assessment background as context for the ToR, information on Bering Sea cod fisheries from the Freezer-Longline Coalition, and transboundary Russia-USA fishery-independent surveys.

The Eastern Bering Sea stock assessment is assessed using Stock Synthesis 3 (SS3). The panel was asked to prioritize the ToRs and decided on the following priorities given in order of the highest one first:

1: Ensemble modeling (originally labeled as “Topic 2”)

2: Movement (originally labeled as “Topic 1”)

3: Fishery CPUE (originally labeled as “Topic 4”)

4: Age data (originally labeled as “Topic 3”)

5: Compositional data (originally labeled as “Topic 5”)

6: Other (originally labeled as “Topic 6”)

Model selection related to ensemble modeling involved a broad discussion of all the models considered. In relation to this, density dependence in growth, maturity, and natural mortality was considered important to include in future assessment models, because missing them can cause a bias in the assessment. The general high level of mathematical and data complexity in the current SS models was a slight concern, because it decreases transparency, increases the risk of coding and input errors, increases the risk of local maxima, hampers the inclusion of important population dynamic mechanisms like density dependence and make high demands on expertise in statistics, mathematics and computer skills. A simpler biomass dynamic model (Surplus Production Model) approach which by design includes all density dependent factors (although not disentangled) was briefly discussed, based on a rough analysis by me at the meeting, using the female SSB and the catch data and various meta-analysis on the shape of the production curve and its height (Fmsy) as input. Such an analysis could in the future be an important supplement to the current approach or might even be an alternative, given the good time series of survey data available for this stock. The panel established five SS3 models and agreed how the models in an ensemble could be weighted. Further, the panel discussed whether model averaging should be applied before or after application of the Harvest Control Rule and tended to the slight preference for calculating the goal parameters, e.g., the ABC, by each model before averaging.

The movement discussion mostly focused on whether cod in the Eastern Bering Sea may move into Russian waters, and there was some emphasis placed on work in progress by Cecilia O'Leary on this topic using data from Russian surveys in Russian waters and pop-up tags which showed several fish moved from US waters to Russian waters. Internal movements within the combined area of Eastern Bering Sea (EBS) and Northern Bering Sea (NBS) were not regarded as one of the most important issues, because the survey now covers the area EBS + NBS for the important recent years and a few earlier years, and the VAST method can fill in the missing years in the time series. That the entire Pacific cod in EBS + NBS + Western Bering Sea (WBS) (Russian part of the Bering Sea) is one stock seems plausible, but it is an exceptionally large area and there might be sub-populations or even genetically distinct population. This seems to be an important future research topic.

Fishery CPUE has been given up in most assessments around the world due to problems getting proper fleet definition. Issues about at what stage a given boat using a given gear is targeting the species in question or not are often difficult to judge from data is a crucial one. Targeting a combination of species and lack of sufficiently detailed data are other ones. For this fishery sufficient detailed data seem to be available and the issue with targeting seems to be less of a problem than usually, at least for the most important fleet component, the hook and line procession vessels. They furthermore fish almost year-round and have observers onboard every trip. Thus, it seems to be a potential option to try to develop a fishery CPUE index for this H&L PV fleet component. It is, however, not something which is done "overnight" and would rather be suitable for, say, a PhD project. Hyper-stability is an issue that needs special attention and the panel speculated that this might be tackled by somehow including "other data" (maybe from the survey or from other fleets and just focusing on the special distribution of the stock) in the approach. Market condition variation from year to year might also influence when a vessel decides to stop fishing at the low catch rates and thus what to assume for not fished cells in the analysis (if it is cell based).

Age data from the commercial fishery have long been a challenge for this stock and age data from the fishery have not been included in the assessment models for many years, as attempts to use those data have proven unsatisfactory for various reasons. Otoliths from commercial fishing operations continue to be collected and the sampling, age readings manuals, test samples and quality control systems now implemented seem good. Research is ongoing on further validation of age readings by chemical analysis of otolith rings, and these seems promising. The panel suggested that growth estimates from tagging studies could also be included in verification of the age readings. The hope is that age data from the fishery can be included in the assessment models at some point. The panel speculated that maybe the used growth curve lacking seasonal variations in growth might contribute to the problem now the assessment model is based on years and not quarters of the year. Of course, this means that the model should go back to how it was some years ago to be based on quarterly time steps or other changes to accommodate this seasonal growth pattern. Most data rich stock assessments in the Northern hemisphere have few problems using age data from the fishery and I cannot see much reason why this stock should be special – for instance, the inter-agreements of age readings between readers seems to be quite good.

"Compositional data" is the issue of effective sample size used in the assessment models for length and age data. The ad hoc approach used in the current models of scaling the hauls numbers from the fishery to the hauls numbers from the survey seems quite sensible although it would not reflect persistent large changes in sampling intensity in either of the two entities.

“Other” issues discussed were dome shape selection curve of survey data and density dependence in growth, maturity, and natural mortality. Regarding the dome shape issue the panel agreed that a model including dome shape selection should be included in the ensemble set of models. It is in general difficult to determine whether a selection is dome shaped or not, but a study referred to from 2016 looking at among other things underwater videos of the behavior of cod in front of the trawl gear during fishing operation indicated that large cod did not avoid the trawl more than young cod indicating that of a flat selection curve. Without going into the study in details it was noted that such studies are notoriously difficult. For instance, the potential spatial distribution of large cod in attractive habitats like rough areas and around shipwrecks where fishing is avoided might still result in a dome shaped selection. Therefore, it was considered prudent to include such a model in the ensemble.

The issue of density dependence in growth, maturity and natural mortality was discussed on the background, that the SS model does not allow for this, and on the fact that missing any one of these will results in the biomass reference points like Bmsy and B100% to be overestimated and the fishing mortality reference points like Fmsy to be underestimated to overestimated. However, the most important density dependent factor, that of the survival of eggs to the recruitment stage, is accounted for the current SS model. This is done via the assumption of the steepness factor (of 1) of the stock-recruitment relationship which means that recruitment is independent of stock size, and this of course must mean that the survival from eggs to recruitment stage is density dependent, i.e., a high survival at low stock size and a low survival at a large stock size. I presented some simple model runs for the panel based on biomass dynamic models (Surplus Production Models – using the SPiCT software and an ad hoc Excel model) which by design includes all four density dependent factors (that in recruitment, in growth, in maturity and in natural mortality - as a combined effect). The input data were catch data and female spawning stock biomass estimates from the ensemble AB model. The SPM models behaved quite well. They revealed, as expected, a somewhat lower B100% than the SS models, and interestingly make it clear that the stock productivity since 1977 have undergone a handful of different productivity periods, low in 1977-1980, high in the 1980s, low in 1990s and 2000s, and high in the 2010s, with a tendency to a reduced productivity in the recent years. These productivity periods were not obvious from the SS models. Thus, the biomass dynamic models can add information to the population dynamics of this stock. It was also mentioned that given the high quality of the survey it might be possible that the biomass dynamic models could be based on this (together of course with the catch data time series) can be used as an alternative assessment model in parallel with the SS model. A retrospective analysis could be conducted to see whether the precision of such an assessment model is good enough to be useful.

The mathematical and statistical modelling of this stock and its assessment have a long history going back to at least the 1980s. In recent decades it has been at the cutting edge of global science in these matters. This science is still in a developing phase, due to the ever-increasing ability of computers and the new opportunities this offers, and to the large uncertainties still apparent in fish stock assessment models. The models are now so complicated that the human brain, and how much it can comprehend, begins to be a bottle neck. There has therefore been an effort in recent years to simplify the models. However, there are still important population dynamic aspects that are not included yet, especially that of density dependence in individual fish growth, maturity, and natural mortality, which are still lacking. Fish stocks in the Northern Hemisphere are generally rebuilding after overfishing have finally been properly addressed and prevented, and this has in many areas resulted in a renewed research interest in density dependence, at least in the North Atlantic area.

Background

The National Marine Fisheries Service (NMFS) is mandated to conserve, protect, and manage the United States' marine living resources based upon the best scientific information available (BSIA). Assessments for this stock will provide the basis for the management of the groundfish fisheries off the U.S. west coast, providing scientific basis for setting Overfishing Limits (OFLs) and Acceptable Biological Catches (ABCs) as mandated by the Magnuson-Stevens Act.

NMFS science products, including scientific advice often require scientific peer reviews that are strictly independent of all outside influences.

The present meeting was an independent peer review process. It took place during a formal, public, multiple-day virtual meeting attended by fishery stock assessment experts, including those selected by the Center for Independent Experts (CIE).

The fishery for Pacific cod in the Eastern Bering Sea is among the most commercially important in the U.S. EEZ. It is also one of the best monitored stocks and it has been exposed to a large research effort on all aspects of its population dynamics and ecosystem relationship. In recent years there has been a substantial northward migration of the stock, to waters outside the area that has been surveyed annually by the NMFS Alaska Fisheries Science Center (AFSC) since 1982. In recent years, this survey has covered this northern area but only on the U.S. side. There is clear evidence that the stock migrates in and out of the Russian zone of the Bering Sea. Interesting co-operation with Russia seems to start up. There are exciting new data from pop-up satellite tagging that in a clear way demonstrate that fish of this stock migrate long distances and far into Russian waters. There have been and still exist conflicts between fishery age composition data and the other data used in the assessment models which has created problems for the assessment. Ensemble modeling has been suggested as a potential solution to the problem of structural uncertainty in the assessment models but attempts to date have been mostly unsuccessful.

The mathematical and statistical modelling of this stock and its assessment have a long history going back to at least the 1980s. In recent decades it has been at the cutting edge of global science in these matters. This science is still in an active developing phase, due to the ever-increasing ability of computers and new opportunities this ability offers, and to the large uncertainties still apparent in fish stock assessment models. The models are now so complicated that the human brain, and how much it can comprehend, begins to comprise a bottleneck. There has therefore been an effort in recent years to simplify the models. However, there are still important population dynamic aspects that are not included yet and that is density dependence in individual fish growth, maturity, and natural mortality, which recent research mainly in the North Atlantic area is beginning to focus on.

CIE Reviewers were appointed to serve as panel members and conduct an impartial and independent peer review. The CIE review panel consisted of Dr. Yan Jiao (USA), and Dr. Arni Magnusson (Iceland), and Dr. Henrik Sparholt (Denmark). The meeting was chaired by Dr. Ingrid Spies, NOAA.

All relevant documentation was made available on a cloud drive two weeks before the meeting. The first two days were spent going through 11 presentations by 10 different presenters, who were all key expert for the issue of their presentation. On Day 4, one more presentation was discussed – one on draft work with Russian scientists about the Pacific cod migration between the U.S. and Russian area of the Bering Sea. The panel was fortunate to also for this issue to have the key scientist involved available for the discussion. The panel recognized the tremendous amount of effort by scientist staff in preparing the assessment and by fishers, observers, managers, and scientists regarding data collection and filtering. Both the documentation and the presentations were of a very high quality. The additional analysis requested by the panel during the meeting were done very competently and expeditiously.

Plenary virtual meetings were held all days between 09:00 and 15:00 Seattle time (equal to 18:00 -24:00 European time). Participants worked solo outside this time window. The two main stock assessment staff, Dr. Grant Thompson and Dr. Steve Barbeaux, participated throughout the plenary and on requests put forward by the panel mainly outside this time window. All answers were presented during the meeting.

The panel discussed the assessment materials in the context of the terms of reference provided for this review.]

Description of the Individual Reviewer's Role in the Review Activities

I read the material posted before the meeting and prepared my key questions to the assessments. I participated in all the plenary meetings from Monday morning 09:00 to Friday afternoon 15:00 (Seattle time). There were good opportunities to discuss the questions as well as the questions from the other panel members. I put forward a few requests to the assessors, as was also done by the other panel members and we agreed a final list of requests each day. The same or next day we got the answers back from the stock assessment staff, and these were then discussed and concluded upon. After the meeting I prepared the present report. I will also participate in drafting the Chairs report of the meeting following the deadline of the present report, 19 May 2021.

Summary of Findings for each ToR for Pacific cod in Eastern Bering Sea.

The ToRs were constructed in a way that the review panel could go quite deep into a few selected issues rather than covering all aspects of this very data rich stock.

TORs:

These were organized into six general topics, with three specific recommendations per topic. After reading the background materials and receiving the initial set of presentations during the review, the reviewers will prioritize the six topics and identify at least one recommendation per topic to be addressed by the review. The reviewers will then address as many of the topics (and the identified recommendation(s)), in priority order, as time allows.

Topic 1: Movement

Recommendation 1a:
Comment on avenues for incorporating spatial dynamics and movement.

Recommendation 1b:
Consider how to inform the dynamics of movement or abundance between the Northern Bering Sea and the Eastern Bering Sea, specifically from additional experiments and analyses, data analyses that include these assumptions (i.e., VAST), and how these can best be used within the different models as indices of abundance.

Recommendation 1c:
Develop movement models.

Topic 2: Ensemble modeling

Recommendation 2a:
Evaluate the use of ensemble modeling in the NPFMC management system, and specifically whether the structural uncertainty and historical challenges in identifying a robust base model make Pacific cod a good application for ensemble modeling.

Recommendation 2b:
Develop the models to include in an ensemble.

Recommendation 2c:
Consider whether to apply the sloping harvest control rule before or after ensemble averaging of SSB and other reference points.

Topic 3: Age data

Recommendation 3a:
Attempt to resolve problems with using fishery age compositions.

Recommendation 3b:
Consider how best to include the fisheries age and size composition data, including consideration of fleet specific age composition data in the model.

Recommendation 3c:
Investigate whether a change in growth contributed to the ageing bias fit for 2008 and onward in the complex models as ageing bias and growth may be confounded.

Topic 4: Fishery CPUE

Recommendation 4a:
Discuss standardization of fishery CPUE using alternative statistical methods, including a discussion of historical changes in the fishery that may affect the relationship of the index to abundance.

Recommendation 4b:
Develop a fishery CPUE index.

Recommendation 4c:
Consider how best to further analyze CPUE, including development of spatio-temporal analyses of fleet specific CPUE indices that may help inform the model or supplement the trawl survey biomass indices.

Topic 5: Compositional data

Recommendation 5a:
Consider methods (e.g., bootstrapping) to estimate uncertainty and variance in the composition data, with the results then used to estimate initial sample sizes for each season, fleet, combination for input into the assessment model.

Recommendation 5b:
Review methods to scale the composition data and include consideration of methods that scale observer samples to the catch by vessel, location, and time of event.

Recommendation 5c:
Consider analyses of the size- and age- composition data to identify if there are specific locations or time periods when a recruitment signal may be apparent to assist in informing the assessment model of the strength of recent recruitment.

Topic 6: Other

Recommendation 6a:
Consider incorporation of dome-shaped survey selectivity.

Recommendation 6b:
Consider the diagnostic plots of fits and residuals (including normalised or Pearson residuals) for the age and size composition data and make recommendations on how the model fits may be improved.

Recommendation 6c:
Consider inclusion of other survey information (e.g., the IPHC and sablefish surveys).

The panel decided on the following priorities given in order of the highest one first:

1: Ensemble modeling (originally labeled as “Topic 2”)
2: Movement (originally labeled as “Topic 1”)
3: Fishery CPUE (originally labeled as “Topic 4”)
4: Age data (originally labeled as “Topic 3”)
5: Compositional data (originally labeled as “Topic 5”)
6: Other (originally labeled as “Topic 6”).

Ad 1. Model selection related to ensemble modeling involved a broad discussion of all the models considered. In relation to this, density dependence in growth, maturity and natural mortality, was considered important to include in assessment models in the future, because it can cause bias especially in the biological reference points estimations missing them. The extremely high level of complexity in the current SS models (in terms of data and mathematics, not so much in biology where maybe it is too simple) was a concern because it decreases transparency, increases the risk of coding and input errors, increases the risk of selecting a local instead of global maximum in the goal function and make high demands on expertise in statistics, mathematics, and computer skills. A simpler biomass dynamic model (Surplus Production Model) approach, which by design includes all density dependent factors (although not disentangled), was briefly discussed. This was based on a rough analysis by me at the meeting, using the female SSB and the catch time series as input data and various assumptions on the shape of the production curve and its height (Fmsy) from recent meta-analysis. Such an analysis could in the future be an important supplement to the current approach or might even be an alternative given the good quality of the time series of survey and catch data available for this stock.

The panel established five SS models (Table 1) by selection some of those presented in the assessment document and by asking for two new models, 20.8a and 21.cie. The models can be characterized as:

19.12a 'base'
19.12 'base + time-varying q'
20.8a 'base + dome-shaped survey'
20.9a 'base + fleet CPUE from vast'
21.cie 'base + estimated survey cv'.

This list retains the historical interest in time-varying q (model 19.12), dome-shaped survey (models 20.8a), and fleet CPUE (model 20.9a), while adding the new interest in estimated survey CV (model 21.cie). The “base” model was the least contested by the panel. All other models had some questions, and they were evaluated in relation to their fit to data and the criterions in Table 1.

The 20.8a model included the dome shape selection. It is in general difficult to determine whether a selection is dome-shaped or not, but a study referred to from 2016 looking at underwater videos of the behavior of cod in front of the trawl gear during fishing operation indicated that large cod did not avoid the trawl more than young cod pointing towards a flat selection curve. Without going into the study in details it was noted that such studies are notoriously difficult to conduct. For instance, the potential spatial distribution of large cod in attractive habitat like rough areas and around shipwrecks, where fishing is difficult, might still result in a dome-shaped selection. The possibly hidden large fish probably come forward at spawning time and if it would be somehow possible to get absolute stock estimates of spawners at that time, maybe this could be used to obtain information about the amount of “hiding” of large fish and then of extent of the dome-shape selection curve. Another possibility might be by the use of pop-up satellite tags and catch rates of these by size of cod, but many tags would probably be needed, and they are expensive. This type of study is probably best conducted separately from the annual stock assessment modelling.

Precisely, because it is difficult to get good data on whether a selection curve is dome-shaped or not, the ensemble approach is especially suitable for this situation because several realistic levels of dome shapes can be included, and the ensemble results might reflect this uncertainty appropriately. In this specific case with EBS-NBS Pacific cod the panel accepted to let the “other” data (probably mainly the length composition data of the survey and of the commercial catch) determine the shape. The panel was aware of the likely problem of the model to estimate the dome shape with a good precision and asked for a retrospective run to see how much the estimated selection curved varied. And sure enough, the retrospective run showed that it was not well estimated (Figure 1). The end point of the selection curve “dome-end” suddenly changed from a stable value of around 0.30 to a stable value around 0.15 from retrospective year 6 to 7 (Figure 1, left panel).

Table 1. Ensemble models and their weights.

Factor 1: Allow *Q* to vary?		no	yes	no	no	no
Factor 2: Allow domed selex?		no	no	yes	no	no
Factor 3: Use fishery CPUE?		no	no	no	yes	no
Factor 4: Estimate survey CV?		no	no	no	no	yes
Criterion	Emph.	19.12a	19.12	20.8a	20.9a	21.cie
General plausibility of the model	3	2	1	0.6667	1	1.3333
Acceptable retrospective bias	3	2	2	1.3333	1	2
Uses properly vetted data	3	2	2	2	0	2
Acceptable residual patterns	3	2	2	2	2	1
Comparable complexity	2	2	1	1	2	2
Fits consistent with variances	2	1	2	1	0	2
Dev sigmas estimated appropriately	0					
Incremental changes	0					
Objective criterion for sample sizes	0					
Change in ageing criteria addressed	0					
Density dependence (other than R) addressed	0					
Regime shifts addressed	0					
Average emphasis:		0.9375	0.8438	0.6875	0.5000	0.8438
Model weight (Ensemble CIE):		0.2459	0.2213	0.1803	0.1311	0.2213

Alternatively, it might be equally good or even better to fix the dome shape to say three different levels and run an SS model for each of them and include all three models in the ensemble. However, for the time being the panel regarded the present dome shape estimate as realistic and the model appropriate for the set of models for the ensemble. Another result of the retrospective analysis was that, despite the strong retrospective pattern, the dome shape selection was consistently estimated to be present.

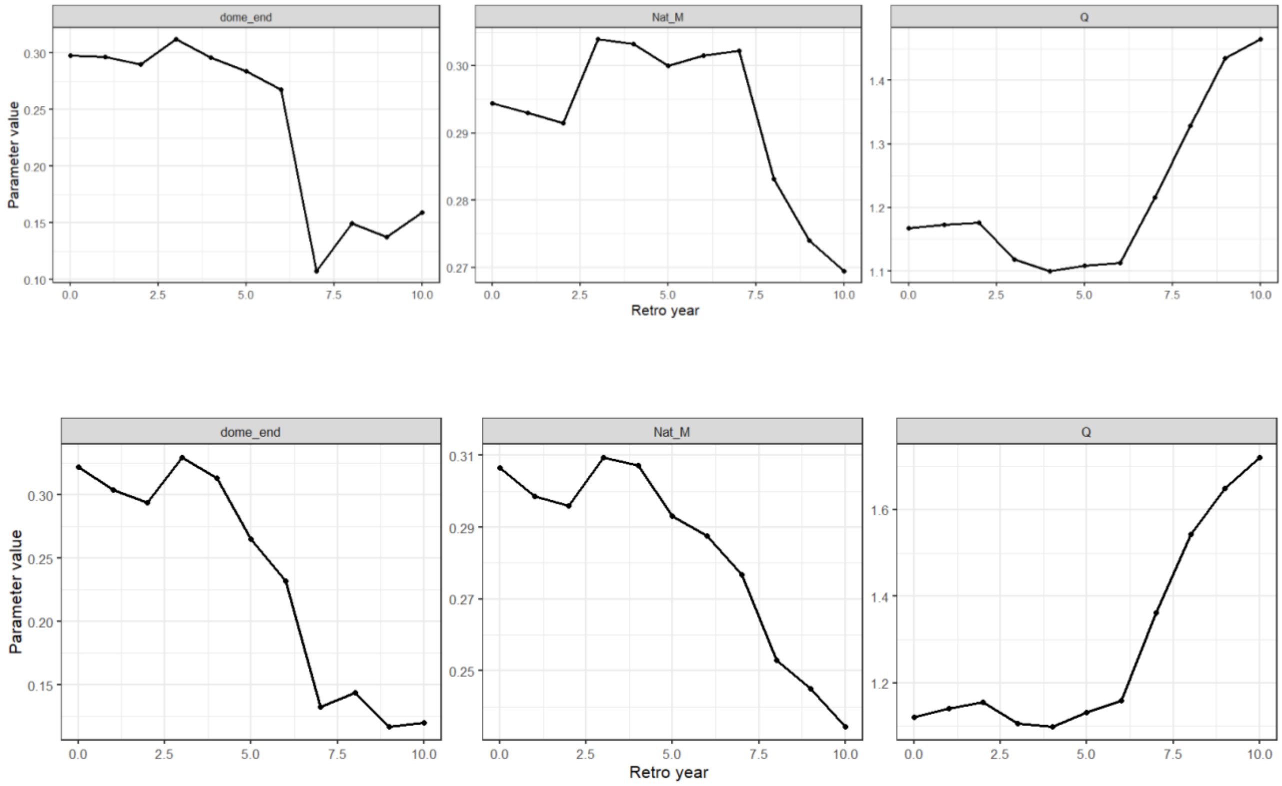


Figure 1. The dome shape parameter estimated by Model 20.8a retrospectively, where the top curve is for varying survey Q and the bottom for fixed survey Q.

Model 20.9 included the fishery CPUE, but a big uncertainty is that the VAST method used to get the input CPUE data is not yet properly vetted.

Model 21.cie was requested by the panel because it was indicated by the other models with fixed q that the precision used for the survey data was too optimistic. A larger CV (0.20 instead of 0.06) of the survey estimates was therefore attached to the survey data in this model 21.cie, consistent with the residuals from the “base” model. However, it was mentioned that the difference between the observed CV in the survey of 0.06 and the modelled CV of 0.20 could be due to process error, e.g., that fish were more pelagic, more concentrated in unfished areas, or other similar phenomenon, in some years than in other years. If this is the case, the most correct modelling would be to allow q to vary by year as model 19.12 did.

All five models in the ensemble had low Mohn’s Rho values (Table 2). The retrospective curves were also produced, presented to the panel, and judged quite acceptable.

Table 2. Mohn's Rho for the ensemble models.

Mohn's Rho						
	19.12A	19.12	20.8a	20.9a	21.CIE	Ensemble = Equal
SSB	-0.019	-0.017	0.053	0.106	-0.029	0.027544
R	-0.132	-0.148	-0.276	0.008	-0.157	-0.13603
F	0.008	0.009	-0.037	-0.090	0.043	-0.00679
Bratio	0.016	0.028	-0.044	0.153	0.008	0.038238

The Panel had extended discussions on how the models in an ensemble could be weighted. The panel agreed on the criterions listed in Table 1. The panel also agreed that each criterion could appropriately be graded 0, 1, or 2, with 2 being the best. Some criteria were judged to be more important than others and therefore an emphasis weighting factor was included. The panel included several criteria which did not distinguish between the five models considered here. This was done to make the table more general, for future use and for use in the assessment of other stocks. The panel was not fully agreeing on all gradings, and some were just the mean of the three reviewer's gradings.

The panel also discussed a suggestion by Grant Thomson on a more objective approach to determine model weights in the ensemble approach. It is called "Cross-conditional model averaging" (CCMA). It is based on a method of model averaging "…that treats each model in a set, one at a time, as though it were the true model, then develops an optimal set of weights for the ensemble, conditional on that assumption, and then finally averages across the set of conditionally optimized ensembles, thus enabling estimation based on an "ensemble of ensembles". The method has an appealing objectivity aspect, which the above approach does not have. Time did not allow the panel to go into a deep analysis of the approach, but noted two other major issues, one is that of increased complexity, and the other that of a challenge of incorporating models structured differently from the SS models like e.g., biomass dynamic models.

The panel discussed whether model averaging should be applied before or after application of the Harvest Control Rule and tended to the slight preference for calculating the goal parameter, e.g., the ABC, by each model, before averaging. However, based on a presentation by Grant Thompson at the meeting where all the pros and cons were listed, it was not easy to judge. It was not even easy to say which approach was the simplest one as at least the one where the averaging is done before the HCR is applied, can be conducted in many alternative ways. In terms of the often-suggested strategy, that of a module build approach where each element in the scientific advice is done separately, the philosophy of having the averaging done to reflect the best estimate of the current stock size and reference points estimates, even though they might not be completely consistent (understood as could be derived by one model), the averaging done before would be better. The panel did not reach a conclusion. I am inclined to favor the option of averaging the ABCs after the HCR has been applied to each model, because it seems simpler and because it can accommodate different model structures like cohort-based models mixed with biomass dynamic models.

Table 3 gives the result of the ensemble modelling for both the "before" and the "after" approach. It should be noted that the two new models 20.8a and 21.cie were not fully tuned or exposed to a jittering and therefore are only preliminary. Model 21.cie seems especially questionable. It deviated quite a bit from the survey trend in the recent years and the panel reflected on the possible reasons for that, but could not come up with any points specifically, except that it probably was due to the influence of the size and age composition data. This model also deviated substantially from the other models in both the B2021 estimate and the ABC calculated. Before model 21.cie is used in management a proper verification of the run is probably needed.

Table 3. Applying the model averaging before or after application of the Harvest Control Rule on the selected models for the ensemble.

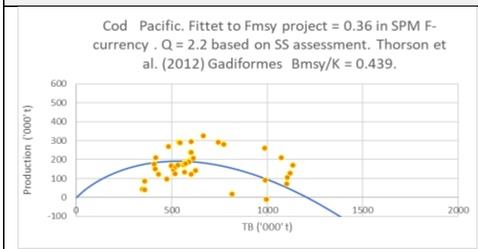
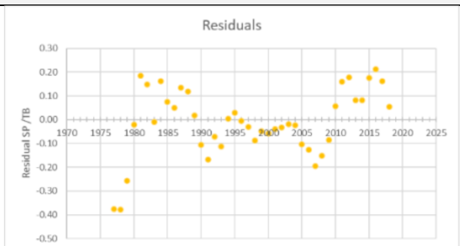
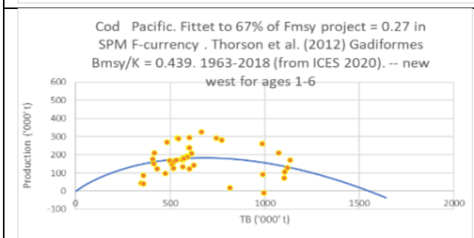
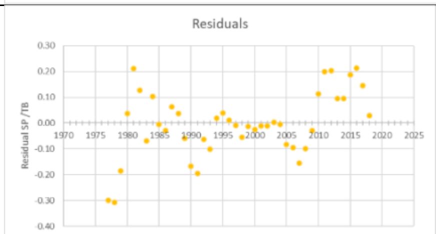
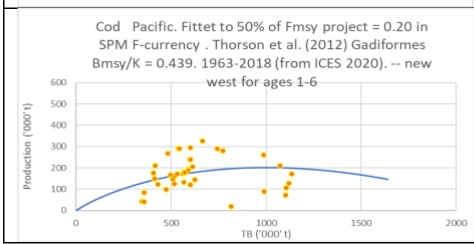
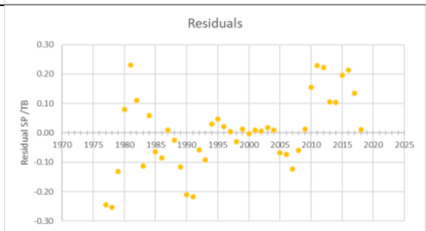
Results from running each model with F2021 set by applying the HCR with model-specific parameters											
		Model:	19_12a	19_12	20_8a	20_9a	21_cie	Ens.			
Quantity	SS_label	Weight:	0.2459	0.2213	0.1803	0.1311	0.2213	1.0000			
B2021	SSB_2021	Mean:	228219	210551	285785	212363	143142	213781			
		Sdev:	18820	23753	30200	17851	33557	52541			
B40%	SSB_SPR	Mean:	260965	265460	310114	264200	259533	270930			
		Sdev:	6135	7727	14599	5624	9505	20680			
B/B40%	Bratio_2021	Mean:	0.8651	0.8651	0.9226	0.7941	0.5475	**0.7959**			
		Sdev:	0.0791	0.0791	0.0919	0.0727	0.1228	0.1651			
F40%	annF_SPR	Mean:	0.3494	0.3285	0.2856	0.3524	0.3254	**0.3284**			
		Sdev:	0.0175	0.0185	0.0164	0.0178	0.0200	0.0291			
F2021	F_2021	Mean:	0.3029	0.2570	0.2620	0.2796	0.1718	0.2533			
		Sdev:	0.0396	0.0421	0.0366	0.0378	0.0469	0.0622			
ABC	ForeCatch_2021	Mean:	123453	99310	128966	107922	47195	**100190**	(This is the ABC from the "before" approach.)		
		Sdev:	22621	24099	26092	20771	22096	**38127**		CV:	0.3805
Results from running each model with F2021 set by applying the HCR with "average" parameters											
		B/B40%:	**0.7959**	F40%:	**0.3284**	F2021:	**0.2578**				
		Model:	19_12a	19_12	20_8a	20_9a	21_cie	Ens.			
Quantity	SS_label	Weight:	0.2540	0.2222	0.1640	0.1058	0.2540	1.0000			
F2021	F_2021	Mean:	**0.2546**	**0.2546**	**0.2546**	**0.2546**	**0.2546**	**0.2546**			
		Sdev:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
ABC	ForeCatch_2021	Mean:	105716	98469	125643	99132	67891	**98471**	(This is the ABC from the "after" approach.)		
		Sdev:	8070	10480	11469	7808	15254	**21831**		CV:	0.2217

SPM models

I presented a rough run using biomass dynamic models (Surplus Production Models - SPM). The SPMs by design includes all density dependent mechanisms, although not in a disentangled way. Such a disentangling is not needed for ABC advise, but of course it would be useful to = understand the population dynamics of the stock. The runs were based on the software SPiCT (Pedersen *et al.* 2016 and [GitHub - DTUAqua/spict: Surplus Production model in Continuous Time]), a software used extensively by ICES expert groups in recent years and on an ad hoc Excel software. I ran a suite of alternative SPM models – not all equally realistic (Figure 2). Input data were the second-best estimate of the stock biomass, namely female SSB from the assessment ("Ensemble AB" run April 2021 – see assessment report) and catches. It would probably have been better to use total stock biomass, but this time series was not readily available from the assessment report. The annual production is catch plus the change in total stock biomass. Meta-analysis of 141 stocks by Thorson *et al.* (2012) determined the mean shape of the production curve for Gadoid fish and the shape was fixed into that value in some models. Based on a meta-analysis of 53 stocks by Sparholt *et al.* (2020) relating Fmsy (i.e., MSY/Bmsy) to the growth parameters Loo, K, and age-at-50% maturity, Fmsy was fixed in some runs. In one run both the shape and the hight (i.e., MSY/Bmsy) of the curve were freely estimated. In most runs the catchability q was set to 2.2 because we need to take account of the biomass of the males and of the immature fish. Interestingly, when q was freely estimated by the data it came out much higher, as 3.6 – this might be a support for the dome shaped selections curve in one of the SS models as discussed above because it would be consistent with a quite strong dome-shaped selection curve of the survey. As expected, (due to the inclusion of all density dependent factors and not just that in egg survival to the recruitment stage), the B100% (and thus B40% and B35%) was generally estimated to be lower than by SS and Fmsy higher. MSY was generally estimated to be about the same. If this model is closer to the true population dynamic of the stock, it has of course implications for the annual advice. Therefore, it might be fruitful to analyse this approach in much more detail than done here for the current assessment. An interesting side effect coming out of this analysis is the consistent picture of productivity variations over time for all the SPM models: low productivity in the late-1970s, high in the 1980s, low in the 1990s and 2000s, and high again in the 2010s.

Alternative SPMs could be run based on the same catch data, but the female SSB replaced by the survey data. This would give an alternative assessment to the SS ones, rather than build on them, as the ones presented above are, because they use the female SSB estimated by the SS model (Ensemble AB).

As SPMs by construction include all density dependent parameters, it might supplement the knowledge from the current assessment model in a nice way I think. They are furthermore very simple and transparent, maybe even elegant as Steven Hawkins thinks models should be [models: "…1)… must be elegant, 2) … contains few arbitrary or adjustable elements, 3) …must agree with — and explain — all existing observations, and 4) … must make detailed predictions about future observations that can disprove or falsify the model …"], and at least simple as Einstein liked (a model should be "…as simple as possible…"). Retrospective analysis could reveal whether they are precise enough to be useful for setting ABCs.

Model	residuals	Parameters estimated '000't (SSB, B100, B40 and B53 are females only)
		SSBmsy 146 TBmsy 530 MSY 192 B100 332 B40 133 B35 116
		SSBmsy 187 TBmsy 681 MSY 184 B100 426 B40 170 B35 149
		SSBmsy 274 TBmsy 999 MSY 201 B100 624 B40 250 B35 219

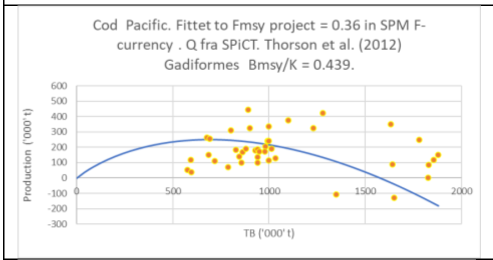
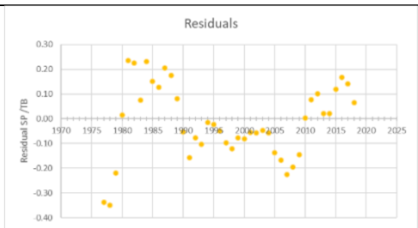
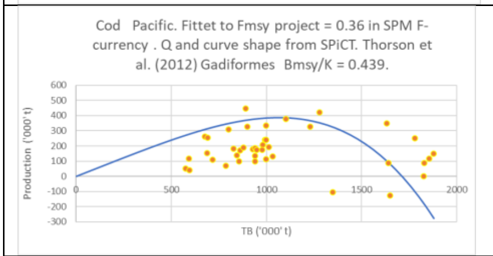
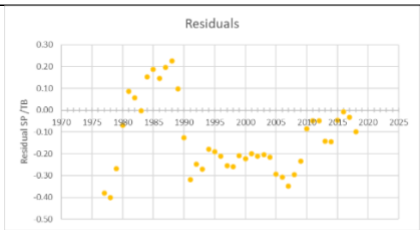
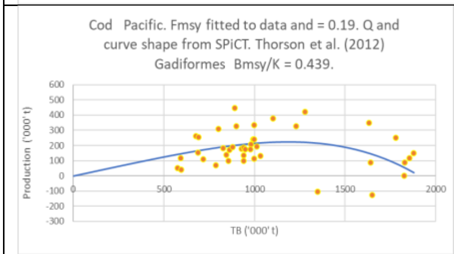
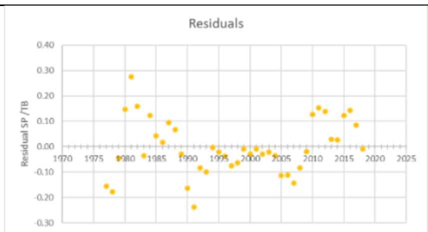
<table>
<tr><td></td><td></td><td></td></tr>
<tr><td>
</td><td>
</td><td>SSBmsy 190
TBmsy 693
MSY 251
B100 433
B40 173
B35 152</td></tr>
<tr><td>
</td><td>
</td><td>SSBmsy 292
TBmsy 1063
MSY 385
B100 470
B40 188
B35 164</td></tr>
<tr><td>
</td><td>
</td><td>SSBmsy 326
TBmsy 1187
MSY 222
B100 525
B40 210
B35 184</td></tr>
</table>

Figure 2. Based on female SSB (“Ensemble AB” run April 2021) and catches. Pacific cod East Bering Sea.

Finally, it could be mentioned that in recent years in both the western and eastern north-Atlantic area there have been an increasing focus on density dependent (DD) factors in parameters other than recruitment. This is probably because overfishing as a general phenomenon in these areas finally after almost a century, has ended and stocks are building up. Then DD becomes more important. There are several meta-analyses in recent years, that has revealed that indeed DD is observable in growth, maturity and natural mortality of individual fish, and sometimes even more important than DD in recruitment for estimation of biological reference points (e.g., Kovalev and Bogstad 2005, Lorenzen 2016, Morgan 2016, Zimmermann et al. 2018, and ICES 2021 – these references were not given at the meeting where only the general phenomenon were mentioned, but are provided here in case others are interested in the issue).

<u>Ad 2</u>. The movement discussion mostly focused on whether cod in the Eastern Bering Sea may move into Russian waters, and there was a large emphasis placed on preliminary work by Cecilia O’Leary on this topic using data from Russian surveys in Russian waters and pop-up tags which showed that several fish moved from U.S. waters to Russian waters. Internal movements within the Eastern Bering Sea, including the Northern Bering Sea, were not regarded as one of the most important issues, because the survey now covers the area EBS + NBS and the VAST method can fill in the missing years in the past time series. Whether the entire EBS + NBS + WBS (Russian part of the Bering Sea) area has one stock only seems plausible, but it is an exceptionally large area and there might by sub-populations or even genetically distinct population. This seems to be an important future research topic to try to find out.

Even if biologically distinct, P.cod is one stock spanning the entire U.S. and Russian area, in terms of management it might be practical to keep the US part separate from the Russian part. Because the area is so huge and both countries are now running a sensible management of the fisheries (Russia got its fishery

MSC certified a few years ago) it is unlikely that one part could severely impact the total stock and that way damage the fishery for the other part by its management or lack of management.

Furthermore, it is not unlikely that there are in fact genetically separate sub-stocks (which then would be real stocks) in this huge area. In the North Sea, a similar sized area in the eastern Atlantic, it has recently been discovered by use of the now easily available genetics techniques that Atlantic cod (a remarkably similar species) is in fact made up of at least two genetically separate stocks (that mix outside spawning time) (ICES 2020).

Ad 3. Fishery CPUE has been given up in most data rich assessments in the northern hemisphere due to problems getting proper fleet definition. An issue often encountered is: when is a given boat using a given gear targeting the species in question and no other species or a combination of species? A huge amount of data is often needed to address this issue properly. For the fishery on P.cod in the EBS+NBS sufficient detailed data seems to be available and the issue with targeting seems to be less of a problem than usually, at least for the most important fleet component, the hook and line procession vessels. This fleet furthermore fishes almost year-round and have observers onboard every trip. Thus, it seems to be a potential option to try to develop a fishery CPUE index for this H&L PV fleet component. There is so much data collected on the fishery operation that it would a pity not to try to explore it. It has failed in most other areas and stocks, but if it cannot be done in the fisheries with so much good data, it probably never will be possible to use fleet CPUE data for any stock. Developing a quality CPUE index is, however, not something which is done “overnight” and would rather be suitable for say, a PhD project.

Hyper-stability is an issue that needs special attention and the panel speculated that this might be tackled by somehow including “other data” (maybe from the survey or from other fleets and where the focus should be on the special distribution of the stock by season) in the approach.

Variation in market conditions from year to year might also influence when a vessel decides to stop fishing at low catch rates, and this influences what to assume for not-fished space-time cells in the analysis (if it is cell based), and this is an important further complication.

Technological development and improvements in fishing gears and thus in catchability are extra challenges. Usually, many aspects of the fishing operation are changing just in a single year, and of course even more so over a long time series and often these cannot be easily revealed and quantified. The panel speculated that the ambition could be to aim for a quite short CPUE timeseries – say 10 years – so that not too much technological creeping is going on in the time window considered. This could mean that for each future year the time series should be truncated by discarding data more than 11 years old.

Ad 4. Age data from the commercial fishery has always been a challenge for the assessment of this stock. In the past several years, age data from the commercial fishery have not been included in P. cod assessment models, as attempts to use those data have proven unsatisfactory for various reasons. Otoliths from commercial fishing operations continue to be collected and the sampling, age readings manuals, test samples and quality control systems now implemented seem good. Research is ongoing on validation of age readings by chemical analysis of otolith rings, and this seems promising.

The panel suggested that growth estimates from tagging studies could also be included in verification of the age readings. This has proven useful for other fish stocks.

The hope is that age data from the fishery can be included in the assessment models at some point in time. The panel speculated that maybe the used growth curve lacking seasonal variations in growth might contribute to the problem, and now the assessment model is based on years and not quarters of the year. Of course, this means that the model should go back to how it was some years ago, to be based on quarterly time steps or other changes to accommodate this seasonal growth pattern. In order not to go against the aim for reducing model complexity one could think of stop including length data and use only age data. These should then be worked up by fleet and season before merged into annual data and entering the assessment model. This would also avoid the complexity of having to estimate growth. Maturity by age could then also be given by year, which is an unfortunate lack of biological complexity at the moment of the SS models. In that way density dependence in both growth and maturity would automatically be included in the part of the model assessing the historical stock development.

Most data rich stock assessment in the Northern hemisphere have few problems using age data from the fishery and I cannot see much reason why this stock should be special – for instance, the inter-agreements of age readings between readers seems to be quite good.

Ad 5. “Compositional data” is the issue of effective sample size used in the assessment models for length and age data. The ad hoc approach used in the current models of scaling the hauls numbers from the fishery to the hauls numbers from the survey seems quite sensible although it would not reflect large persistent changes in sampling intensity in either of the two entities.

Ad 6. “Other” issues discussed were dome shape selection curve of survey data and the panel added an extra issue: the importance of density dependence in growth, maturity, and natural mortality for especially B100, FOFL and similar biological reference points.

The “dome shape” selection is dealt with in section **Ad 1** above.

The issue of density dependence in growth, maturity and natural mortality was discussed in the background, that the current SS model does not allow for this, and on the fact that missing any one of these, will result in the biomass reference points like Bmsy and B100% to be overestimated and the fishing mortality reference points like Fmsy to be underestimated. However, the most important density dependent factor, that in survival of egg to the recruitment stage is accounted for in the current SS model. This is done via the assumption of the steepness factor (of 1) of the stock-recruitment relationship which mean that recruitment is independent of stock size, and this of course must mean that the survival from eggs to recruitment stage is density dependent, i.e., a high survival at low stock size and a low survival at a large stock size. I presented some simple model runs for the panel based on biomass dynamic models (Surplus Production Models – and the SPiCT software and some done in Excel) which by design includes all four density dependent factors (that in recruitment, in growth, in maturity and in natural mortality - although as a combined effect). The input data were catch data and female spawning stock biomass estimates from the ensemble AB model and the SPM models behaved quite well. They revealed as expected, a somewhat lower B100% than the SS models, and interestingly make it clear that the stock productivity since 1977 have undergone a handful of different productivity periods, low in 1977-1980, high in the 1980s, low in 1990s and 2000s, and high in the 2010s, with a tendency to a reduced productivity in the recent years. These productivity periods were not obvious from the SS models. Thus, the biomass dynamic models can add information to the population dynamics of this stock. It was also mentioned that given the high quality of the survey it might be possible that the biomass dynamic models could be based on this (together of course with the catch data timeseries) and in that way might be used as an alternative assessment model to and in parallel with the SS model. A retrospective analysis could be conducted to see whether the precision of such an assessment model is good enough to be useful.

The mathematical and statistical modelling of this stock and its assessment have a long history going back to the 1980s at least. In recent decades it has been at the cutting edge of global science in these matters. This science is still in a developing phase, due to the ever-increasing ability of computers and new opportunities this offers, and to the large uncertainties still apparent in fish stock assessment models. The models are now so complicated that the human brain, and how much it can comprehend, begins to be the bottleneck. There has therefore been an effort in recent years to simplify the models. However, there are still important population dynamic aspects that are not included. Probably the most important is that of density dependence in individual fish growth, maturity, and natural mortality. Research mainly in the North Atlantic area has started focusing on these aspects.

Conclusions and Recommendations

The panel did not attempt to reach conclusions and recommendations. However, for each of the points discussed I have tried to condense the outcome of the panel below.

1: Ensemble modeling (originally labeled as “Topic 2”). It was regarded as a sensible way forward to use an ensemble approach. The panel selected a set of five models which – after the two new models 20.8a and 21.cie have been probably vetted – could be used by management. The weighting scheme agreed by the panel could be used.

2: Movement (originally labeled as “Topic 1”). There seems not to be an urgent need for a movement model inside the area EBS and NBS. However, WBS could be included in cooperation with Russia.

3: Fishery CPUE (originally labeled as “Topic 4”). The CPUE series presented at the meeting were not fully vetted. There seems to be potential for arriving at a useful CPUE time series via a comprehensive analysis of the rich amount of quality data available in this fishery for especially the hook and line processor vessels fleet.

4: Age data (originally labeled as “Topic 3”). There seem to be potential in continuing the good work on aging and its verification. Maybe a quarterly model or an annual model based on cod age-distribution rather than on cod length-distribution by fleet could be developed and used in the future.

5: Compositional data (originally labeled as “Topic 5”). The present ad hoc method seems sensible, and it is difficult to come up with a better one.

6: Other (originally labeled as “Topic 6”). The issue of dome shape selection is difficult, but it seems prudent to include a model allowing for that in the ensemble set of models. Density dependence in growth, maturity and natural mortality are important to include in the forecast part on the population dynamic modelling to obtain unbiased biological reference points. If this cannot be modelled individually, a way forward could be to use Surplus Production Models, where all density dependence is included by design.

The NMFS review process

The review process was influenced by the Covid-19 pandemic, so a physical meeting could not take place. Instead, a virtual meeting was conducted with individual participants working from home via their PC. This worked quite well. It was important that almost all presentations were produced two weeks before the start of the meeting so that it was possible to prepare well for the meeting. However, most of the normal one-to-one informal interactions between participants were lacking, which under normal circumstances are important for a comprehensive exchange of views, ideas, and opinions. Daily sessions from 09:00 to 15:00 Seattle time were appropriate and meant that participants from central Europe could end the day at midnight.

The documentation and presentation were of a remarkably high quality. Documentation was sent out two weeks before the meeting using a cloud drive. The meeting was conducted in an efficient, engaged, and positive atmosphere.

The guidelines to the reviewers from the CIE lead coordinator were clear and to the point.

The exchange of knowledge between the reviewers and the scientific staff was very fruitful, it seemed for both parties.

The panel put forward some requests to the two assessors. These were very efficiently answered, although the assessors had to work hard and had long days during the meeting.

The presentations of all the important aspects relevant for the review were very much appreciated by the panel.

I tried hard to think of possible improvements to suggest, but the only minor issue I could come up with is that the presentation of the hook and line process vessel fishery was very welcomed, and it would have been good if there were a similar presentation of the other fleets participating in the fishery – these are smaller fleets and one presentation covering all would suffice. The NMFS review process has evolved over time and seems now to have reached a very high standard in my opinion.

All in all, a particularly good process seen from the reviewer’s perspective, for doing a comprehensive and in-depth review on selected topics.

References

ICES. 2020. Workshop on Stock Identification of North Sea Cod (WKNSCodID). ICES Scientific Reports. 2:89. 82 pp. http://doi.org/10.17895/ices.pub.7499.

ICES. 2021. Workshop of Fisheries Management Reference Points in a Changing Environment (WKRPChange, outputs from 2020 meeting). ICES Scientific Reports. 3:6. 39 pp. https://doi.org/10.17895/ices.pub.7660.

Kovalev, Y. A., and Bogstad, B. 2005. Evaluation of maximum long-term yield for Northeast Arctic cod. In Ecosystem Dynamics and Optimal Long-Term Harvest in the Barents Sea Fisheries. Proceedings of the 11th Russian–Norwegian Symposium, Murmansk, Russia, 15–17 August 2005, pp. 138– 157. IMR-PINRO Joint Report Series 2/2005.

Lorenzen, K. 2016. Toward a new paradigm for growth modeling in fisheries stock assessments: Embracing plasticity and its consequences. Fisheries Research, 180: 4–22. https://doi.org/10.1016/j.fishres.2016.01.006.

Morgan, M. J., Shelton, P. A., González-Costas, F., and González-Troncoso, D. 2016. Compensation potential in six depleted groundfish stocks from the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences, 73(2): 257–269 https://doi.org/10.1139/cjfas-2015-0056.

Pedersen, Martin W., and Casper W. Berg. 2017. “A stochastic surplus production model in continuous time.” Fish and Fisheries 18 (2): 226–43. https://doi.org/10.1111/faf.12174.

Sparholt, Henrik, Bjarte Bogstad, Villy Christensen, Jeremy Collie, Rob van Gemert, Ray Hilborn, Jan Horbowy, Daniel Howell, Michael C. Melnychuk, Søren Anker Pedersen, Claus Reedtz Sparrevohn, Gunnar Stefansson and Petur Steingrund. 2020. Estimating Fmsy from an ensemble of data sources to account for density-dependence in Northeast Atlantic fish stocks. ICES Journal of Marine Science. ICES Journal of Marine Science, doi:10.1093/icesjms/fsaa175.

Thorson, J. T., J. M. Cope, T. A. Branch, O. P. Jensen. 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Canadian Journal of Fisheries and Aquatic Sciences, 69: 1556-1568.

Zimmermann, F., Ricard, D., and Heino, M. 2018. Density regulation in Northeast Atlantic fish populations: Density dependence is stronger in recruitment than in somatic growth. Journal of Animal Ecology, 87(3): 672–681. doi: 10.1111/1365-2656.12800.

Appendix 1. List of material provided.

Thompson, G. 2021 Cross-conditional model averaging: A potential tool for improving stock assessment estimates. Draft manuscript.

Thompson, G., Conner, J., Shotwell, K., Fissel, B., Hurst, T., Laurel, B., Rogers, L., Siddon, E. 2020. Assessment of the Pacific cod stock in the Eastern Bering Sea.

Spies, I., Gruenthal, K.M., Drinan, D.P., Hollowed, A.B., Stevenson, D.E., Tarpey, C.M. and Hauser, L., 2020. Genetic evidence of a northward range expansion in the eastern Bering Sea stock of Pacific cod. Evolutionary applications, 13(2), pp.362-375.

O'Leary, C., Kotwicki, S., Hoff, G., Thorson, J., Kulik, V., Ianelli, J., Lauth, R., Nichol, D., Conner, J., Punt, A.

Estimating spatiotemporal availability of transboundary fishes to fishery-independent surveys DRAFT NOT FOR CIRCULATION.

The following presentations were pre-recorded and posted on the website:

1. Conner—survey data. Author: Jason Conner works in the Groundfish Assessment Program (GAP) of the AFSC.
2. Stone and Anderl—ageing. Authors: Kali Stone and Delsa Anderl work in the Age and Growth Unit of the AFSC.
3. Kraski—North Pacific Observer Program. Author: Joel Kraski works in the Fishery Monitoring and Analysis Division of the AFSC.
4. Furuness—catch accounting system and inseason management. Author: Mary Furuness works in the NMFS Alaska Region office.
5. Nielsen—tagging. Author: Julie Nielsen (Kingfisher Marine Research) is a contractor with GAP/AFSC.
6. Correa—somatic growth variability. Author: Giancarlo M. Correa is a PhD student of Lorenzo Ciannelli at Oregon State University.
7. Thorson—ADT movement models. Author: Jim Thorson leads the Habitat and Ecosystem Processes Research Program of the AFSC.
8. Thorson—VAST fishery CPUE model. Author: See above.
9. Shotwell—Ecosystem and Socioeconomic Profile (ESP). Author: Kalei Shotwell works in SSMA/AFSC.
10. Thompson—assessment background as context for the ToR. Author: See above.

Additional presentations:

11. Merrigan, G. Additional information on Bering Sea p-cod fisheries from the Freezer-Longline Coalition (FLC = Catcher-processor hook-and-line vessels).

12 O'Leary, C. Estimating spatiotemporal availability of transboundary fishes to fishery-independent surveys.

Appendix 2. Statement of work.

Performance Work Statement (PWS)

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Virtual Panel Review of the Stock Assessment for Pacific Cod in the Eastern Bering Sea

April 26-30, 2021

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The fishery for Pacific cod in the Eastern Bering Sea is among the most commercially important in the U.S. EEZ. Recent developments of note include a substantial northward migration of the stock, to waters outside the area that has been surveyed annually by the NMFS Alaska Fisheries Science Center (AFSC) since 1982. Efforts at modeling this movement have been hampered by the scarcity of both survey data from the northern region and tagging data in general. Conflicts between fishery age composition data and the other data used in the assessment models also pose problems for the assessment. Ensemble modeling has been advocated as a potential solution to the problem of structural uncertainty in the assessment models, but attempts to date have been mostly unsuccessful.

The goal of this review will be to ensure that the stock assessment represents the best available science to date and that any deficiencies are identified and addressed. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (TORs) of the peer review are listed in **Annex 2**. Lastly, the tentative agenda of the virtual panel review meeting is attached in **Annex 3**.

Requirements

NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the PWS, OMB guidelines, and the TORs below. The reviewers shall have a working knowledge of, and recent experience in, the following areas:

- The Stock Synthesis modeling framework;
- Movement (migration) models;
- Ensemble modeling (model averaging); and
- Federal fisheries science requirements under the Magnuson-Stevens Fishery Conservation and Management Act.

The chair, who is in addition to the three reviewers, will be provided by the AFSC; although the chair will be participating in this review, the chair's participation is not covered by this contract.

Tasks for Reviewers

1) Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site to the CIE reviewer all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review.

2) Additionally, two weeks prior to the peer review, the CIE reviewers will participate in a test to confirm that they have the necessary technical (hardware, software, etc.) capabilities to participate in the virtual panel in advance of the review meeting. The AFSC NMFS Project Contact will provide the information for the arrangements for this test.

3) Attend and participate in the virtual panel review meeting. The meeting will consist of presentations by NMFS scientists, review of model runs conducted during the course of the evening, and discussion among the reviewers, assessment scientists, other scientists involved in the assessment or management process, and members of the public.

4) After the virtual panel review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

5) Each reviewer should assist the Chair of the meeting with contributions to the summary report.

6) Deliver their reports to the Government according to the specified milestones dates.

Place of Performance
The place of performance will be held remotely, via Google Meets video conferencing.

Period of Performance
The period of performance shall be from the time of award through June 2021. The CIE reviewers' duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables
The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Schedule	Milestones and Deliverables
Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
April 26-30, 2021	Virtual panel review meeting
Approximately 3 weeks later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
No travel is necessary, as this meeting is being held remotely.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:
Grant Thompson
grant.thompson@noaa.gov

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Performance Work Statement
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

The Terms of Reference were compiled from recommendations submitted by the Groundfish Plan Team for the Bering Sea and Aleutian Islands, the Scientific and Statistical Committee, and Alistair Dunn (a consultant contracted by the Freezer Longline Coalition). These were organized into six general topics, with three specific recommendations per topic. After reading the background materials and receiving the initial set of presentations during the review, the reviewers will prioritize the six topics and identify at least one recommendation per topic to be addressed by the review. The reviewers will then address as many of the topics (and the identified recommendation(s)), in priority order, as time allows.

Topic 1: Movement

<u>Recommendation 1a:</u>
Comment on avenues for incorporating spatial dynamics and movement.

<u>Recommendation 1b:</u>
Consider how to inform the dynamics of movement or abundance between the Northern Bering Sea and the Eastern Bering Sea, specifically from additional experiments and analyses, data analyses that include these assumptions (i.e., VAST), and how these can best be used within the different models as indices of abundance.

<u>Recommendation 1c:</u>
Develop movement models.

Topic 2: Ensemble modeling

<u>Recommendation 2a:</u>
Evaluate the use of ensemble modeling in the NPFMC management system, and specifically whether the structural uncertainty and historical challenges in identifying a robust base model make Pacific cod a good application for ensemble modeling.

<u>Recommendation 2b:</u>
Develop the models to include in an ensemble.

<u>Recommendation 2c:</u>
Consider whether to apply the sloping harvest control rule before or after ensemble averaging of SSB and other reference points.

Topic 3: Age data

<u>Recommendation 3a:</u>
Attempt to resolve problems with using fishery age compositions.

<u>Recommendation 3b:</u>
Consider how best to include the fisheries age and size composition data, including consideration of fleet specific age composition data in the model.

<u>Recommendation 3c:</u>
Investigate whether a change in growth contributed to the ageing bias fit for 2008 and onward in the complex models as ageing bias and growth may be confounded.

Topic 4: Fishery CPUE

Recommendation 4a:
Discuss standardization of fishery CPUE using alternative statistical methods, including a discussion of historical changes in the fishery that may affect the relationship of the index to abundance.

Recommendation 4b:
Develop a fishery CPUE index.

Recommendation 4c:
Consider how best to further analyze CPUE, including development of spatio-temporal analyses of fleet specific CPUE indices that may help inform the model or supplement the trawl survey biomass indices.

Topic 5: Compositional data

Recommendation 5a:
Consider methods (e.g., bootstrapping) to estimate uncertainty and variance in the composition data, with the results then used to estimate initial sample sizes for each season, fleet, combination for input into the assessment model.

Recommendation 5b:
Review methods to scale the composition data and include consideration of methods that scale observer samples to the catch by vessel, location, and time of event.

Recommendation 5c:
Consider analyses of the size- and age- composition data to identify if there are specific locations or time periods when a recruitment signal may be apparent to assist in informing the assessment model of the strength of recent recruitment.

Topic 6: Other

Recommendation 6a:
Consider incorporation of dome-shaped survey selectivity.

Recommendation 6b:
Consider the diagnostic plots of fits and residuals (including normalised or Pearson residuals) for the age and size composition data and make recommendations on how the model fits may be improved.

Recommendation 6c:
Consider inclusion of other survey information (e.g., the IPHC and sablefish surveys).

Annex 3: Tentative Agenda

Google Meet link: TBD
Phone: TBD

DRAFT AGENDA TBD

Virtual Panel Review of the Stock Assessment for Pacific Cod in the Eastern Bering Sea

Virtual Panel

April 26-30, 2021

Point of contact: Grant Thompson (grant.thompson@noaa.gov)

Appendix 3. List of participants.

Ingrid Spies ……………..……………….Chair, NOAA Fisheries, Alaska Fisheries Science Center

Grant Thompson……..Assessment Author, NOAA Fisheries, Alaska Fisheries Science Center

Steve Barbeaux………Assessment Author, NOAA Fisheries, Alaska Fisheries Science Center

Henrik Sparholt……..………………………………………...Reviewer, University of Copenhagen

Yan Jiao……..………………..…………………………..…..…………..…Reviewer, Virginia Tech

Arni Magnusson……………………………………………….………………………………Reviewer

Thomas Helser……..……...…………………NOAA Fisheries, Alaska Fisheries Science Center

Jason Conner……..……...…………………..NOAA Fisheries, Alaska Fisheries Science Center

Delsa Anderl……..……...……………………NOAA Fisheries, Alaska Fisheries Science Center

Joel Kraski……..……...………………………NOAA Fisheries, Alaska Fisheries Science Center

Chad See……..……………………….…..……………………......……Freezer Longline Coalition

Kalei Shotwell……..……...………………..….... NOAA Fisheries, Alaska Fisheries Science Center

Tim Loher…………………..……………………….....……International Pacific Halibut Commission

Craig Kastelle……..……...…………………... NOAA Fisheries, Alaska Fisheries Science Center

Kali Stone……..……...………………………..…NOAA Fisheries, Alaska Fisheries Science Center

Suzanne Mcdermott……..……...……………… NOAA Fisheries, Alaska Fisheries Science Center

Julie Neilsen…………………………………………..……………University of Alaska Fairbanks

Mary Furuness……..……...………………..……..………………………………… NOAA Fisheries

Gerry Merrigan……..……………………………..…………..……......……Freezer Longline Coalition

Giancarlo Correa…………………………….....……...…………………………University of Oregon