JOINT MEETING OF THE BSAI AND GOA GROUNDFISH PLAN TEAMS
May 17, 2011

Members of the Plan Teams present for the meeting included those shown in bold below.

Loh-Lee Low	**AFSC REFM (BSAI chair)**	**Jim Ianelli**	**AFSC REFM (GOA co-chair)**
Mike Sigler	**AFSC (BSAI Vice chair)**	**Diana Stram**	**NPFMC (GOA co-chair)**
Kerim Aydin	AFSC REFM	**Sandra Lowe**	**AFSC REFM**
Lowell Fritz	AFSC NMML	**Chris Lunsford**	**AFSC ABL**
David Carlile	**ADF&G**	**Jon Heifetz**	**AFSC ABL**
Alan Haynie	**AFSC REFM**	**Mike Dalton**	**AFSC REFM**
Jane DiCosimo	**NPFMC (Coordinator)**	**Kristin Green**	**ADF&G**
Henry Cheng	**WDFW**	Tom Pearson	NMFS AKRO Kodiak
Brenda Norcross	UAF	Nick Sagalkin	ADF&G
Mary Furuness	**NMFS AKRO Juneau**	Paul Spencer	AFSC
Grant Thompson	**AFSC REFM**	Leslie Slater	USFWS
Dave Barnard	ADF&G	Nancy Friday	AFSC NMML
Leslie Slater	USFWS	**Henry Cheng**	**WDFW**
Dana Hanselman	**AFSC ABL**	Ken Goldman	ADF&G
Bill Clark	**IPHC**	Bob Foy	AFSC Kodiak
		Sarah Gaichas	**AFSC REFM**
		Steven Hare	IPHC

Others in attendance: Susan Hilber, Teresa A'Mar, Kenny Down, Mark Maunder, Neil Rodriguez, Kari Fenske, Kalei Shotwell, and Obren Davis.

This meeting of the Joint Groundfish Plan Teams constituted one step of the annual process established by the SSC in December of 2009 for the purpose of developing models to be analyzed in the GOA and BSAI Pacific cod stock assessments. Specifically, the Plan Teams' charge during this meeting was to generate a list of models (one for each area) that, upon review and possible modification by the SSC in June, will be analyzed in the preliminary SAFE report which the Plan Teams will review in September. Mike Sigler chaired the meeting.

The process of developing model recommendations this year included a CIE review of the current Pacific cod models in March. The reviewers' reports were made available to individuals who requested them on April 22 and were posted on the Council website on April 26 (with an e-mail to all Plan Team and SSC members alerting them of the reports' availability).

Appendices A, B, and C (attached) comprise the materials distributed to the Plan Teams, SSC, and others on May 3 (with one typographical error corrected and a few minor formatting changes). Appendix A summarizes and systematizes the CIE reviewers' recommendations; Appendix B lists excerpts of recent Plan Team and SSC minutes pertaining to Pacific cod; and Appendix C includes Mark Maunder's recommendations, along with his commentary on the CIE reviewers' reports.

A total of 144 unique proposals were received by the deadline of April 29 (the terms "proposal" and "recommendation" will be used interchangeably in these minutes). These included 128 proposals from the CIE reviewers, 1 from the GOA Plan Team, 10 from the SSC, and 5 from Mark. The complete set of proposals contained in the CIE reviewers' reports is shown in Appendix A, Table A1 (this table includes 19 duplicate proposals; i.e., identical proposals recommended by more than one CIE reviewer, so the total number of proposals shown there is 147 rather than 128). The proposals submitted by the GOA Plan Team, SSC, and Mark are shown here in Table 1.

A spreadsheet that further systematized the proposals was distributed to the Plan Teams, SSC, and others on May 13. The proposals were characterized in the spreadsheet as follows:

1. The stock assessment author would be the most appropriate person to address 99 of the proposals, 11 proposals would involve some sort of programmatic change (e.g., proposals that would necessitate changes to the observer program, survey program, or other program that involves multiple species), and 34 proposals were specific to Pacific cod but were directed to or would be most appropriately addressed by people other than the stock assessment author.
2. 116 of the proposals called for some sort of change (not necessarily to the model, however), while 28 recommended keeping some specific feature of the status quo.
3. 109 of the proposals pertained to features of the model, while 35 pertained to something else.
4. 105 of the proposals could (individually) be accomplished this year, while 39 would take longer.

To keep the list of proposals manageable, Grant Thompson suggested that the Teams consider only those proposals that: 1) were most appropriately addressed by the assessment author, 2) called for some sort of change, 3) pertained to the model, and 4) could be accomplished this year. Use of this filter reduced the list from 144 proposals to 50. These were then organized by subject area (one for each of the 10 CIE terms of reference, one for miscellaneous, and one for Mark's proposals), with closely related proposals listed together in groups. Mark's proposals were listed separately from the others because:

1. three of them address more than one topic,
2. they describe complete models (albeit conditionally—see below) rather than single features, and
3. they are conditional on a model that has not been developed yet (the authors' preferred model).

The meeting proceeded with the following order of events:

1. Grant gave an overview presentation.
2. Mark responded to questions about his proposals.
3. The Teams proceeded with a quick first pass through the proposals, identifying those that could be eliminated without much discussion.
4. The Teams then took a second pass through the remaining proposals, identifying those that were high priority.
5. Finally, the Teams took a third pass through the high priority proposals, allocating them among various new models that the Teams would recommend for development this summer (the Teams used a self-imposed limit of *no more than five* new models for each area).

During the course of events #3-5 above, 5 new proposals were developed by the Teams, bringing the total number of proposals to 149, of which 55 were considered during the meeting. (Note that the Teams decided not to offer proposals during their November 2010 meeting, preferring to wait until the results of the CIE review became available.) The 55 proposals considered during the meeting and the results of their evaluation by the Plan Teams are shown in Table 2 (the codes identifying the 5 new proposals developed during the meeting are shown in bold, red font).

After the first pass, 32 proposals remained (listed as "maybe" in the "1st pass" column of Table 2). After the second and third passes, only eight proposals were ranked as being high priority (listed as "high" in the "2nd pass" column of Table 2). Of these eight, the following two were determined to fall outside the scope of the present exercise (listed as "n/a" in each of the "3rd pass" columns in Table 2):

1. Recommendation JPT3 ("jitter-proof the model") reflected the Teams' belief that the ongoing difficulty in finding the true maximum likelihood was a major concern. The Teams also recognized that this was a major concern of the CIE reviewers. However, the Teams also realized

that this was more of a goal than an identifiable model feature, so instead of allocating this recommendation to one or more new models, the Teams chose simply to list it as a high priority item and use it as a rationale for prioritizing other recommendations.

2. Recommendation JD32 ("see if bad fit to 2010 survey at small sizes is a coding error") was already implemented during the course of the CIE review. During the review, the assessment author demonstrated that the poor fits were not a coding error, but were instead caused by the size-at-age data overwhelming the survey data. This fact motivated the Teams to develop recommendation JPT1 ("omit size-at-age data").

The remaining six high-priority recommendations were allocated among four new models. A condensed version of Table 2, showing only those recommendations ranked as high priority by the Teams, is shown in Table 3. The Teams assumed that last year's model would be included automatically as Model 1.

Model 2 would test two unrelated features: JPT4—One Team member noted that the ability to model selectivity by using splines has very recently been added to Stock Synthesis (SS). The Teams felt that this feature might improve the models' convergence properties significantly. CD33—For many years, inclusion of the pre-1982 survey data in the EBS model was considered to be important because those data helped to monitor the strength of the extremely large 1977 year class as it moved through the population. However, because of a change to the survey gear in 1982, use of the pre-1982 data requires estimation of an additional six selectivity parameters. Given the fact that the 1977 years class left the population many years ago, the Teams felt that testing the effect of removing the pre-1982 survey data would be worthwhile. The Teams viewed both of these recommendations (JPT4 and CD33) as "conditional" changes, meaning that if they resulted in Model 2 being an improvement over Model 1, then they would be used in Models 3-5 also (indicated by these recommendations being listed as "cond." in the columns for Models 3-5 in Tables 2 and 3). Recommendation CD33 would obviously apply only to the EBS models.

Model 3 would be devoted to exploring the possibility of estimating ageing bias inside the model (JD6, SSC6). The ability to model ageing bias in terms of internally estimable parameters was added to SS late last year, and was tested in the EBS Pacific cod model to a small extent prior to and during the CIE review. The Teams felt that internal estimation of ageing bias could potentially be much more efficient and accurate than the manual estimation (i.e., trial and error tuning "by hand") that was used in the 2009 assessments and retained in last year's assessments.

Model 4 would be similar to Model 4 (or perhaps Model 5) from last year's preliminary assessments, in that it would omit age-based data to a very large extent, including elimination of all size-a-age data (JPT1) and all age composition data (JPT2). The only difference between this year's Model 4 and Model 4 from last year's preliminary assessment is that this year's Model 4 describes maturity as a function of age rather than length, as in last year's final models. If the author has time to examine the possibility of estimating length-at-age variance internally and if the results appear reasonable, the Plan Teams would be happy to see this included as a feature of Model 4 (which would make it more like Model *5* from last year's preliminary assessment).

Model 5 would likely result in a reconfiguration of the time blocks currently used to define multiple selectivity schedules for most fisheries, and would likely result in less time variation in the survey selectivity schedules. The approach to be used (JPT5) is very similar to one of Mark's proposals (MM2), except that survey selectivity is included along with fishery selectivity. Recommendation JPT5 is also concordant with SSC4 ("simplifying trawl survey selectivity should be investigated and model fit to data components evaluated"). As with some other recommendations, the desire to simplify the model and achieve improved convergence was a major factor in the Plan Teams' decision to rate JPT5 as high priority.

Table 1. Recommendations received from the GOA Plan Team, SSC, and Mark Maunder. “Sub.” = subsection of Appendix A to which the recommendation most directly relates, “ID”= code used to identify each proposal, “No.” = unique recommendation number (note that SSC6 and SSC8 are identical to other proposals listed in Appendix A), red font indicates that participation by groups or individuals other than the assessment author would be required if that recommendation were to be implemented.

Sub.	ID	No.	Summary of recommendation
D	GPT1	129	it would be useful to have a presentation of the estimates relative to the data, particularly for the most recent survey (and sub-27 cm abundance index)
2a	SSC1	130	evaluate reduced catch season ... structures that are more parsimonious, but do not diminish the information content.
2b	SSC2	131	evaluate reduced … size bin structures that are more parsimonious, but do not diminish the information content.
5	SSC3	132	trawl survey catchability used in the assessment and model sensitivity to model estimates or plausible alternatives should be evaluated
8	SSC4	133	simplifying trawl survey selectivity should be investigated and model fit to data components evaluated
1c	SSC5	134	re-tune ageing bias to try to better match the observed age modes
1c	SSC6	19	explore internal estimation of ageing bias
3	SSC7	135	evaluate Richards growth curve alternative
1a	SSC8	1	continue existing research on age determination/validation
D	SSC9	136	the SSC recommends that an AI assessment be brought forward for evaluation (only) during the next assessment cycle
D	SSC10	137	for the GOA, apply a simple Kalman filter approach, as adopted by the SSC in 2004 for BSAI for estimation of current biomass distrubution
3	SSC11	138	constant growth should be brought forward in future models (run times reduced back to 2-3 minutes)
D	SSC12	139	the ... author and Plan Team should develop a plan of action for how the BSAI cod assessment should evolve vis-à-vis treatment of the BS and AI
1c,1d	MM1	140	authors' preferred model, but with bias and variance of the ageing error matrix estimated inside the stock assessment model
4	MM2	141	authors' preferred model, but with time blocks determined by initially modeling selectivity as a random walk
7	MM3	142	authors' preferred model, but with sample sizes estimated as follows: start with bootstrap estimates, rescale so that average = 300, re-weight iteratively
3,4,8	MM4	143	authors' preferred model, but with time-varying growth and constant
1a,1b,1d	MM5	144	authors' preferred model, but with conditional-age-at-length instead of agecomps, all sizecomps on, mean size off, length-age variance estimated

Table 2 (p. 1 of 4). Proposals evaluated by Plan Teams (see text for details).

			1st pass	2nd pass	3rd pass			
			No/maybe	Priority	M2	M3	M4	M5
1. Use of age data								
1.1	YC6	downweight age and length data if both are used	no	n/a				
1.1	CD6	downweight age and size-at-age data if both are used	no	n/a				
1.1	JD5	investigate appropriate weighting of non-independent data	no	n/a				
1.2	YC7	if length-at-age variance estimated externally, omit size-at-age data	no	n/a				
1.2	**JPT1**	omit size-at-age data	maybe	high			x	
1.2	CD5	if cohort-specific growth not used, omit size-at-age data	no	n/a				
1.3	SSC5	re-tune ageing bias to try to better match the observed age modes	maybe	low				
1.3	JD6,SSC6	explore internal estimation of ageing bias	maybe	high		x		
1.3	CD8	constrain ageing bias to increase with age	maybe	low				
1.3	CD7	consider variable/flexible growth as an alternative to ageing bias	no	n/a				
1.4	JD4	include fishery age composition data	no	n/a				
1.5	**JPT2**	omit all age data	maybe	high			x	
2. Data partitioning/binning								
2.1	CD14	do not aggregate catch across gears if selectivity is held constant	maybe	low				
2.2	SSC1	evaluate reduced catch season ... structures that are more parsimonious, but do not diminish the information content	no	n/a				
2.3	YC23	explore dynamic binning	maybe	low				
2.3	JD10	explore coarser bin structure for large sizes	no	n/a				

Table 2 (p. 2 of 4). Proposals evaluated by Plan Teams (see text for details).

			1st pass	2nd pass	3rd pass			
			No/maybe	Priority	M2	M3	M4	M5
2. Data partitioning/binning (continued)								
2.3	SSC2	evaluate reduced … size bin structures that are more parsimonious, but do not diminish the information content	no	n/a				
2.4	YC21	justify blocks based on analysis of factors that may affect selectivity	maybe	low				
2.4	YC22	omit fleets that have minimal impacts on the assessment	no	n/a				
2.5	JD12	explore possible over-parameterization of GOA sub-27 catchability	maybe	low				
3. Functional form of the length-at-age relationship and estimating the parameters thereof								
3.1	YC28	estimate a_0 externally	maybe	low				
3.1	YC30	consider estimating all growth parameters externally	maybe	low				
3.1	YC29	estimate growth internally only if parameter correlations justified	no	n/a				
3.2	SSC7	evaluate Richards growth curve alternative	maybe	low				
4. Number and functional form of selectivity curves								
4.1	YC31	develop hypotheses to explain derived selectivity curves	maybe	low				
4.1	CD19	if possible, link selectivity to tagging studies or other assessments	maybe	low				
4.1	JD15	force just one selectivity to be asymptotic, justify this assumption	no	n/a				
4.2	**JPT4**	test the use of splines to model selectivity	maybe	high	x	cond.	cond.	cond.
4.3	YC34	evaluate selectivity trend using random walk, one fleet at a time	no	n/a				
4.3	CD20	use random walk selectivity, hold constant where change is small	no	n/a				

Table 2 (p. 3 of 4). Proposals evaluated by Plan Teams (see text for details).

			1st pass	2nd pass	3rd pass			
			No/maybe	Priority	M2	M3	M4	M5
4. Number and functional form of selectivity curves (continued)								
4.3	JD17	use random walk selectivity or justify current blocks statistically	no	n/a				
4.3	JD16	explore bimodal selectivity for GOA survey	no	n/a				
4.3	**JPT5**	authors' preferred model, but with time blocks determined by initially modeling fishery and survey selectivity as a random walk	maybe	high				x
5. Fixing the trawl survey catchability coefficient								
5.1	SSC3	trawl survey catchability used in the assessment and model sensitivity to model estimates or plausible alternatives should be evaluated	maybe	low				
6. Fixing the natural mortality rate								
6.1	YC37	use unbiased age at maturity when applying Jensen's equation	maybe	low				
6.2	YC38	if approach is Bayesian, derive M prior from alternative estimators	no	n/a				
6.3	CD26	change M value only during off-cycle "benchmark" meetings	maybe	low				
7. Input sample sizes and survey sigma								
7.1	JD23	consider setting input N for multinomial equal to number of trips	no	n/a				
8. Annual variability in trawl survey selectivity								
8.1	CD31	force survey selectivity to be constant over time in the model	maybe	low				
8.2	JD25	tie changes in survey selectivity to temperature, not time	no	n/a				
8.3	SSC4	simplifying trawl survey selectivity should be investigated and model fit to data components evaluated	maybe	low				
9. Recruitment sigma								
9.1	YC41	estimate σR iteratively	no	n/a				

Table 2 (p. 4 of 4). Proposals evaluated by Plan Teams (see text for details).

			1st pass	2nd pass	3rd pass			
			No/maybe	Priority	M2	M3	M4	M5
10. Survey and fishery CPUE data								
10.1	YC45	remove fishery CPUE data from the model	maybe	low				
10.2	CD33	remove pre-1982 survey data from the EBS model	maybe	high	x	cond.	cond.	cond.
11. Miscellaneous								
11.1	YC48	justify choices/estimates involving life history, fishery processes	no	n/a				
11.2	YC50	use a fully Bayesian approach	no	n/a				
11.3	CD36,JD31	reduce the number of parameters in the models	maybe	low				
11.3	**JPT3**	jitter-proof the model	maybe	high	n/a	n/a	n/a	n/a
11.3	JD30	investigate parameter redundancy	maybe	low				
11.4	JD32	see if fit to 2010 survey at small size is a coding error	maybe	high	n/a	n/a	n/a	n/a
12. Mark Maunder's scenarios								
12.1	MM1	authors' preferred model, but with bias and variance of the ageing error matrix estimated inside the model	maybe	low				
12.2	MM2	authors' preferred model, but with time blocks determined by initially modeling fishery selectivity as a random walk	maybe	low				
12.3	MM3	authors' preferred model, but with sample sizes estimated as follows: start with bootstrap estimates, rescale so that average = 300, re-weight iteratively	maybe	low				
12.4	MM4	authors' preferred model, but with time-varying growth and constant survey selectivity	maybe	low				
12.5	MM5	authors' preferred model, but with conditional-age-at-length instead of agecomps, all sizecomps on, mean size off, length-age variance estimated inside	maybe	low				

Table 3. Proposals included in the Plan Teams' four recommended new models. "Sub." = subsection of Appendix A to which the proposal most directly relates; "ID"= code used to identify each proposal; "M2," "M3," "M4," and "M5" = Models 2, 3, 4, and 5, respectively. The Teams assumed that last year's model would automatically be included as Model 1.

Sub.	ID	Proposal	M2	M3	M4	M5
4.2	JPT4	test the use of splines to model selectivity	x	cond.	cond.	cond.
10.2	CD33	remove pre-1982 survey data from the EBS model	x	cond.	cond.	cond.
1.3	JD6,SSC6	explore internal estimation of ageing bias		x		
1.2	JPT1	omit size-at-age data			x	
1.5	JPT2	omit all age data			x	
4.3	JPT5	authors' preferred model, but with time blocks determined by initially modeling fishery and survey selectivity as a random walk				x
11.3	JPT3	jitter-proof the model	n/a	n/a	n/a	n/a
11.4	JD32	see if fit to 2010 survey at small size is a coding error	n/a	n/a	n/a	n/a

Appendix A:
Summary of recommendations arising from the review of the EBS and GOA Pacific cod models conducted by the Center for Independent Experts

Table of Contents

I. Overview

The models used to assess the Eastern Bering Sea (EBS) and Gulf of Alaska (GOA) stocks of Pacific cod (*Gadus macrocephalus*) were reviewed during the dates March 14-18, 2011 by three scientists contracted by the Center for Independent Experts (CIE). The reviewers were Drs. Yong Chen, Chris Darby, and Jose DeOliveira. The reviewers' reports were made available on April 22. This document summarizes the recommendations contained in the reviewers' reports. Recommendations cover not only the topics contained in the ten Terms of Reference (Section II), but several other topics as well (Section III).

The procedure used to organize this document was as follows: Recommendations within each topic are listed in alphabetical order by the reviewer's last name and labeled with the reviewer's initials. For each reviewer, recommendations are listed in the order given in the reviewer's report, except in cases where recommendations have been moved between sections or subsections to improve the flow of the document. In cases where a reviewer made exactly the same recommendation multiple times, the recommendation is listed only once; in cases where a reviewer made approximately the same recommendation multiple times, either the recommendations have been merged or only the most specific version of the recommendation has been listed. Each recommendation is listed verbatim with enough accompanying text to make the context clear. For ease of reference, each recommendation is followed by a short, paraphrased summary (shown in italics, surrounded by square brackets). It should be emphasized that these summaries are only "pointers" to the actual recommendations, and are not the recommendations themselves.

Recommendations are color-coded as follows: black = a recommendation that would be handled most appropriately by the senior assessment author (e.g., "estimate ageing error externally"), red = *either* a recommendation that would be handled most appropriately by someone other than the senior assessment author *or* a recommendation that the reviewer explicitly directed to someone other than the senior assessment author (e.g, "continue existing research on age determination/validation"), blue = a recommendation that would require programmatic change (e.g., "change requirements for observer coverage"). Recommendations for additional information to be included in the SAFE report were among those considered to be of this third type.

Table 1A lists the summarized recommendations. A total of 147 recommendations were catalogued, of which 128 were unique (i.e., not duplicated by multiple reviewers). Dr. Chen contributed 61 recommendations, Dr. Darby 49, and Dr. DeOliveira 37.

II. Recommendations on topics contained in the Terms of Reference

Ten terms of reference (the first two of which were divided into four and three parts, respectively) were specified. Reviewers were asked to make recommendations with respect to each of them, for both the EBS and GOA Pacific cod models.

1: Use of age data

1a. Use of age composition data

YC1: Continue exploring various methods … to reduce the likelihood of having ageing errors before ageing data are used in stock assessment. *[Summary: continue existing research on age determination/validation]*

YC2: Estimate age error probability either outside or inside the SS3 (personally I prefer it is estimated outside of the model to reduce confounding of different components in the parameter estimation)…. Ageing errors and variations should be estimated outside the SS3 model. *[Summary: estimate ageing error externally]*

YC3: I believe the age verification process currently employed by the AFSC is scientifically sound and can yield results that can be directly incorporated into stock assessment modeling. *[Summary: retain use of age composition data]*

YC4: Evaluate hypotheses of low catchability of age 2 fish in the survey. *[Summary: explain missing 2-year-olds in GOA survey]*

YC5: However, the on-going and proposed research efforts in validating annulus may be complicated by fish migrations and large temporal/spatial temperature stratifications in the stock areas, resulting in inconclusive results. Other approaches such as using Pacific cod held in aquaculture facilities, evaluating back-calculated size at age for annulus, and conducting more extensive tagging studies should be explored for annuli validation. *[Summary: expand existing research on age determination/validation]*

YC7: Because age composition data were derived from subsamples of length composition data, using both in the same survey is essentially equivalent to up-weighting size composition data. If both sets of data are used in the SS3, they should be down-weighted accordingly so that this set of size (both age and length) composition data has the same weight as other size composition data (e.g., having a weighting factor of 0.5 for both age and length composition data in the survey if they are both used in the SS3). *[Summary: downweight age and length data if both are used]*

CD1: The procedures for collection of otoliths and length samples are considered appropriate. *[Summary: retain current otolith and length sampling procedures]*

CD2: Inclusion of the ageing error is appropriate - given the lack of agreement between readers. *[Summary: retain use of ageing error matrix]*

CD3: Given that:

1. there is information on the error in the reading of the age, based on an agreed standard for determining ages, and
2. there is a known potential bias within the age reading that is being investigated,

then the inclusion of the age composition data in the model fit is considered appropriate. *[Summary: retain use of age composition data]*

CD4: If the research into age reading establishes a new protocol for determining the age of cod that is accepted as the new standard, then one suggestion for reducing the uncertainty inherent in the assessment would be to use otoliths collected from the commercial fishery at regular intervals (e.g. every three years) to augment the survey information. This would require a relatively low increase in sampling levels but would help to stabilize the model estimates from the increased information level. *[Summary: if ageing criteria change, include fishery age composition data]*

JD1,2,3,4: Age composition data are valuable, and their continued use, coupled with an ageing error matrix, is highly recommended. This approach is supported by ongoing research into age determination methods and validation techniques, and this ongoing research is encouraged. The application to fishery data is also encouraged. *[Summaries: retain use of age composition data; retain use of ageing error matrix; continue existing research on age determination/validation; include fishery age composition data]*

<u>1b. Use of mean-size-at-age data</u>

YC7: Use of mean-size-at-age data in the model partially repeats the size composition information already implied in length composition data and age composition data (if both used) in the model. This may subjectively put extra weight on size composition data. If between-individual variability in growth can be estimated outside the model (see my comments below), use of mean-size-at-age data in modeling is not necessary. *[Summary: if length-at-age variance estimated externally, omit size-at-age data]*

CD5: Mean size at age was included within the model to allow the fitting of cohort specific growth. If this model is not used then the data is not required. CIE runs 10 and 11 evaluated the removal of the mean size at age. *[Summary: if cohort-specific growth not used, omit size-at-age data]*

CD6: As mean size-at-age is derived from the same information as the age composition data (age and length frequency samples), the data are not strictly independent and therefore if it is to be included the correlation with the age composition data should be considered carefully (halving the likelihood component contribution?). *[Summary: downweight age and size-at-age data if both are used]*

JD5: The appropriate statistical treatment of non-independent data (e.g. when data based on the same samples are used in two components of the overall likelihood) should be investigated. *[Summary: investigate appropriate weighting of non-independent data]*

<u>1c. Use of ageing bias as an estimated parameter</u>

YC8: Given the complexity of the SS3 model, I believe it is difficult to interpret the estimation results for ageing bias and variation in modeling. Because parameters are, to varying degrees, correlated, ageing bias and variation may not be estimated independently of other parameters. These estimates may not reflect real ageing errors and variations. Rather, they may reflect combined effects of errors and variations of all data sources. An external estimate of ageing errors and variations may be a better way to incorporate the uncertainty of this information in the stock assessment. *[Summary: estimate ageing bias externally]*

CD7: The bias estimated by the model will arise partially from the laying down of false rings, as highlighted by the otolith chemistry studies, but could also result from an inappropriate formulation of the growth curve - in terms of either, the use of a single growth curve when variable growth is more appropriate, or a formulation that is not sufficiently flexible to model the specific seasonal (and regional) characteristics of the length data from the fishery. *[Summary: consider variable/flexible growth as an alternative to ageing bias]*

CD8,9: One area of concern is the modeling of bias as a single value starting at age 2 and which is modeled as a parameter with a symmetric distribution. If the bias results from the formation of false rings then would not bias increase with age as the opportunity to form false rings increases? In addition, the study by Kastelle et al. indicated that many of the otolith ages were read correctly for the remainder age was over-estimated – this would seem to imply an asymmetric bias. *[Summaries: constrain ageing bias to increase with age; revise SS to allow for asymmetric ageing bias]*

JD6: The feasibility of internal estimation of ageing error bias should be explored (the runs considered by the review panel were not focused enough to consider this properly). *[Summary: explore internal estimation of ageing bias]*

<u>1d. External estimation of between-individual variability in size at age</u>

YC9: I suggest back-calculating length-at-age data using otoliths to derive length at each age for each fish with its corresponding otolith sample. A nonlinear random effects model explicitly assumes that an individual's growth parameters are samples taken from a multivariate distribution, which can then be applied to the back-calculated length at age data (Hart 2001; Pilling et al. 2002) to estimate between-individual variability. *[Summary: estimate length-at-age variance by otolith back-calculation]*

CD10,11,12: Presentations to the review established that estimation of between-individual variability in size at age could not be achieved internally…. Models 5 and 6 fitted to the BSAI cod and 5 fitted to the GOA cod both estimate variances for the standard deviation of mean length at age that are significantly larger than the majority of the observations. The method by which the external estimates are obtained and entered as external estimates in the fitted models is considered appropriate at this stage in the model development. However, … there appears to be curvature in the data at increasing size at age. Is this an artifact of temporal changes in the linear relationship such that plotting them together appears curvilinear or is a more complex relationship between the standard deviation and mean length? *[Summaries: estimate length-at-age variance externally; retain current procedure for estimation of length-at-age variance; investigate apparent curvilinearity of length-at-age variance]*

JD7: The provision of external estimates of between-individual variability in size-at-age data should continue as is (efforts to estimate them internally were not successful). *[Summary: retain current procedure for estimation of length-at-age variance]*

2: Data partitioning/binning

2a. Catch data partitioned by year, season, and gear

YC10: Given the strong seasonality in fishing activity and large differences in catchability/selectivity among different gears, I believe the current partition of catch by year, season, and gear is a reasonable and logical approach. *[Summary: retain current partitioning of catch data by year, season, gear]*

YC11: However, the variability of catch quality among years, seasons and gears needs to be carefully evaluated. *[Summary: evaluate variability of catch data quality by year, season, gear]*

YC12: Other sources of fishing mortality that are currently not included in the cod catch estimates also need to be evaluated. These include baits used in crab fisheries, recreational fishing, subsistence fishing, and research surveys. Part of Pacific cod mortality in the halibut fishery is also not included in the cod catch because of lack of observer coverage. *[Summary: include catch from all sources]*

YC13: I suggest that observer coverage should not be determined by vessel size. Rather, it should be determined by data needs, and should have a good representation of gear and vessel size composition in the fishing fleet. [*Summary: change requirements for observer coverage*]

YC14: Because the current (catch accounting) program has some overlaps in catch reporting from different sources, data from different sources can be compared and cross-validated. Such a study can yield some insights about potential errors in catch estimates from different sources. *[Summary: compare and cross-validate catch data from different sources]*

YC15: Given the importance of the catch data in the assessment, I suggest conducting an extensive computer simulation study based on the data collected in the past to evaluate the effectiveness of the current sampling/reporting system in yielding catch estimates, to evaluate potential error sources and levels of catch estimates, and to identify alternative sampling/reporting program designs. *[Summary: evaluate current catch sampling/reporting system via simulation]*

YC16,17: I suggest estimating uncertainty associated with catch estimates to develop a plausible range of catch estimates, which can be used to evaluate impacts of uncertainty associated with catch estimates on stock assessment. *[Summaries: estimate catch uncertainty; once catch uncertainty has been estimated, evaluate its impacts]*

CD13: Following an analysis of the seasonal structure of the amounts of catch landed by month the optimal seasonal structure for the catch model was considered to comprise 5 seasons for BSAI and GOA cod; differing by stock. Three selectivity periods are defined for each gear type which overlap the catch seasons. The reasoning underlying the approach and the analysis to identify the seasonal components is considered appropriate. *[Summary: retain current seasonal structure for catch and selectivity]*

CD14: I would have doubts about the utility of a collapsed model in which length compositions are mixed across gears in proportions that have change markedly and quickly during the time series. *[Summary: do not aggregate catch across gears if selectivity is held constant]*

JD8: Catch estimation for Pacific cod is underpinned by both industry reports and one of the most comprehensive observer programs to be found anywhere (presentation 9 and report 20, Appendix 1). Although variance estimates are not currently available, they are in the pipeline and could be used in future to challenge the assumption of no error in total catch data in current assessment models. The provision of these variance estimates should be encouraged, if practicable, to ensure the models are based on appropriate assumptions regarding the catch data. *[Summary: estimate catch uncertainty]*

2b. Size composition data partitioned by year, season, gear, and 1-cm size intervals

YC18: Given the strong seasonality of fisheries and large differences in selectivity/catchability and fishing seasons among gears, I believe the current partition of fisheries catch size composition by season and gear is necessary and reasonable. The current seasonal partition also yields the best model in the most recent assessment. *[Summary: retain current partitioning of sizecomp data by season and gear]*

YC19,20: Size composition data for fisheries catch are derived from various sources and are likely subject to various errors. However, I did not see the quantification of uncertainty associated with size composition estimates for fisheries data. In-depth analyses should be conducted to evaluate if the quality of size composition data for fisheries catch vary with year, season and gear. Variation or confidence intervals can be estimated for each size bin as indicators for uncertainty associated with size composition data. *[Summaries: quantify uncertainty associated with fishery sizecomp data; evaluate variability in quality of fishery sizecomp data by year, season, gear]*

YC21: Changes in many factors may influence selectivity/catchability in fisheries, which may affect catch size compositions. For example, changes in baits used in longline and pot fisheries among years and seasons may result in annual variations in catchability/selectivity. Squid, which were used in the past as bait, tend to have high catchability, but haven't been used on a large scale in current years because of high prices. Such changes from year to year may influence size composition data and should be considered in determining year block. More in-depth analyses should be conducted to identify factors that may affect selectivity/catchability and evaluate how these factors vary among years and seasons to justify the partitions of catch size composition by year and season. *[Summary: justify blocks based on analysis of factors that may affect selectivity]*

YC22: For a given model configuration, data of different fleets can be deleted one at a time to identify which fleet has had the largest impact on the assessment. Those that have had limited impact can be removed to improve model convergence. *[Summary: omit fleets that have minimal impacts on the assessment]*

YC23: I suggest that more study be done in the future to explore the dynamic binning approach. *[Summary: explore dynamic binning]*

YC24: It also should be noted that the size interval of 1 cm used to group length data implies that measurement errors for fish length should be smaller than 1 cm. This is probably a reasonable assumption, but should be explicitly evaluated and clearly defined to ensure that quality of data collected is adequate for such fine binning. *[Summary: evaluate precision of length measurements]*

YC25: Area closure for Pacific cod fishing in the major Steller sea lion habitats in 2011 may affect effective cod stock areas included in the stock assessment. Because of spatial variability in cod size composition, lack of size composition data in major sea lion habitats from 2011 may introduce extra variations in size composition data. Possible impacts of this closure on size composition data should be evaluated and considered when partitioning size composition data by year. *[Summary: evaluate effects of recent SSL area closures on sizecomp data]*

YC26: For survey catch-size composition data, errors should be relatively small, compared with fisheries catch-size composition data. However, survey stations in EBS and AI are fixed, and more study is needed to evaluate potential impacts of such a design on the quality of size composition data. Uncertainty associated with size composition data should be estimated. *[Summary: evaluate effects of survey sampling design on sizecomp precision]*

CD15: The finer 1-cm bin structure for the size composition data was introduced as a refinement to allow the analysis of length to correspond to the scale at which the data was collected. In the range of lengths for which large amounts of data are collected from the fishery by gear this is considered appropriate. However, at the smallest and largest sizes finer binning introduces large numbers of zeroes in the length distributions. Dynamic binning was examined at the meeting in runs GOA9 and CIE9 and appeared to be the way forward. Questions were raised during the review about how SS3 treats sample sizes when combining bins, and this should be investigated. *[Summary: investigate treatment of sample size in SS when merging bins]*

JD9: Teresa A'mar raised the possibility of a coding error with how SS treats effective sample size when combining bins. *[Summary: investigate treatment of sample size in SS when merging bins]*

JD10: Although the finer bin structure may be justified for smaller sizes, this might not be the case for larger sizes, and a coarser bin structure should be explored for the latter. *[Summary: explore coarser bin structure for large sizes]*

<u>2c. Age composition data partitioned by year, season, and gear</u>

CD16: Commercial fishery age composition data for a single year was used in earlier models for BSAI and GOA cod but not in recent assessments. Use of a single year's data can be problematic in terms of weighting and therefore its omission is considered appropriate. *[Summary: do not use the existing small sample of fishery agecomp data]*

CD17: The trawl survey for GOA cod is separated by length into sub-27 and 27-plus components, which is carried out to help the model resolve a missing mode in the length frequency data for age 2 cod. The way in which the size composition is modeled is an artifact of the restriction to the SS program, this is not ideal; it would be better to have an assessment model that allows allow for this, as the current solution requires extra parameters to fit the model. *[Summary: include a bimodal, parametric selectivity curve in SS]*

JD11: The partitioning of data to deal with data features (e.g. change in gear) and limitations in SS functionality (e.g. lack of bi-modal selection) is sensible. *[Summary: retain partitioning of GOA survey data into sub-27 and 27+ ranges]*

JD12: However, there are problems with the fit to the GOA sub-27 index (exact fits, indicating over-parameterisation) that need looking into. *[Summary: explore possible over-parameterization of GOA sub-27 catchability]*

JD13: The SS developer should be encouraged to include a bi-modal selectivity curve option to avoid the ad-hoc length split, and thereby improve the general functionality of SS. *[Summary: include a bimodal, parametric selectivity curve in SS]*

3: Functional form of the length-at-age relationship and estimating the parameters thereof

YC27: The Richards model, even though more general, provides no better fitting than the von Bertalanffy growth function (VBGF) in one of the test runs conducted during the review. Thus, VBGF is sufficient to describe the length-at-age relationship. *[Summary: retain use of von Bertalanffy growth]*

YC28: Fitting length-at-age data outside the SS3 model to estimate a_0 (age at size of 0) may be an option. Because of the availability of small/young fish in surveys, it is likely that a_0 should have a negative value if this approach is taken. This negative a_0 value can be fixed with the other two parameters being estimated for VBGF in the SS3 model to ensure that the size at age 0 is positive. *[Summary: estimate a_0 externally]*

YC29: Estimating VBGF parameters inside the SS3, although allowing for flexibility in adjusting growth parameters to better fit size composition and data, may create unnecessary correlations between growth and other life history and fishing processes. For a converged run, a close evaluation should be done for the variance-covariance matrix to evaluate possible correlations between growth parameters and other model parameters. High correlations should be biologically justified. If not, spurious correlations may result from tradeoffs of different life history and fisheries processes in model fitting, and the estimates of growth parameters (and other parameters, for this matter) should be questioned. *[Summary: estimate growth internally only if correlations are justifiable]*

YC30: Alternatively, estimating growth parameters outside the SS3 may also be a choice, although this may result in poor fitting of size composition data. *[Summary: consider estimating all growth parameters externally]*

CD18: For the BSAI cod the model fitted with the new growth formulations had a worse fit to the data for the GOA cod (which did not require the initial length to be constrained) there was a marked improvement in the model fit. The Richard's function is more flexible but there are problems in its fitting, potentially implying that it is not flexible enough at the youngest ages/sizes. It would be beneficial, given the potential link to bias estimation, to evaluate other functions if the Stock Synthesis author can be encouraged to code them. *[Summary: include more flexible growth functions in SS]*

JD14: The need to constrain one of the growth parameters to be positive to enable the Richards growth curve to be used leads to poor model fits when this constraint becomes active (e.g. for EBS, but not for GOA). This indicates that the constrained Richards model is actually less flexible than the unconstrained Von Bertalanffy model in some cases, and that more flexible growth models should be considered. *[Summary: include more flexible growth functions in SS]*

4: Number and functional form of selectivity curves estimated, including assumptions regarding which selectivity curves should be forced to exhibit asymptotic behavior

YC31: Current choice of selectivity function tends to have large flexibility to let model fitting decide the selectivity curves, although in some cases selectivity is forced to follow the curves. In many cases, there is lack of justification for the choice of a particular selectivity function for a fishery. I believe relevant hypotheses should be developed to explain the derived selectivity curves. This has not been done explicitly, giving me an impression that the choice of selectivity function was rather ad hoc and even arbitrary. *[Summary: develop hypotheses to explain derived selectivity curves]*

YC32: Forcing a selectivity curve to exhibit asymptotic behavior implies that fish in large sizes/ages are 100% available to and selected by fishing gear. Clearly, this may not be true for longline and pot because they are passive fishing gears and more size selective. Because selectivity here also includes fish availability to fishing gear, it is also hard to imagine that 100% of fish of any size class become available to trawls. However, if fish of certain size classes become unavailable to fishing gears, they are not part of exploitable stock biomass. In this case forcing selectivity to exhibit asymptotic behavior yields the estimates of exploitable stock biomass. This should be considered in interpreting stock assessment results. *[Summary: consider the possible effect of partial availability to the fishery]*

YC33: Seasonal selectivity is biologically justified because fishing activity is likely to vary greatly among seasons and fish distribution and availability to fishing gears tend to have seasonal patterns. Thus, I believe current seasonal selectivity is reasonable. *[Summary: retain current partitioning of selectivity by season]*

YC34: The choice of time block for selectivity is rather arbitrary (BSAI). I believe that a random walk over years may be a better choice. Once a model is run with random-walk selectivity over years, the temporal trend of selectivity plots needs to be examined closely to identify any temporal pattern. The identified temporal pattern can be used in the future to decide the time block for selectivity. For multiple fleets, I believe we need to evaluate one fleet at a time for their temporal trend while holding others constant. *[Summary: evaluate selectivity trend using random walk, one fleet at a time]*

CD19: This is clearly an area for which there is a need for more analysis, as is the case for this constraint in the majority of stock assessments. In general targeted trawl fisheries are assumed to have asymptotic selection, unless there are specific spatial or temporal reasons for assuming otherwise. If possible more information from tagging studies or linkages to assumptions made in other assessments with known selectivity for large fish by the same gears is required. *[Summary: if possible, link selectivity to tagging studies or other assessments]*

CD20: Comparisons with the base model fits indicate improved diagnostics in the models fitted with the block structure - indicating the need for modelling changes in time. However, it is not clear if the transition points between blocks are appropriate and in some cases the variation in the selection, especially at the largest sizes, could result from fitting to noise. Where there is evidence of a drift in selection parameters in time, a time series approach should be considered (similar to that used for the pollock assessment) and for those fleets which do not show significant change in time, a constant selection model should be adopted in order to remove as many selection parameters as possible. *[Summary: use random walk selectivity, hold constant where change is small]*

JD15: The forcing of just one major fishery to have asymptotic selection (e.g. the Jan-Apr trawl fishery for both stocks) should be explored. This is an alternative to the ad hoc approach used to force a number of fisheries to exhibit such behaviour for EBS, but needs to be justifiable, given the additional parameters that may be required. *[Summary: force just one selectivity to be asymptotic, justify this assumption]*

JD16: The inclusion of bi-modal selection may avoid some of the issues surrounding the fit to the sub-27 GOA survey index, and should be explored. *[Summary: explore bimodal selectivity for GOA survey]*

JD17: An alternative to block selectivity is to consider a constrained random walk over time, but if this is not practicable, the current block structure could be justified given model selection criteria (this was not verifiable during the meeting given the runs considered). *[Summary: use random walk selectivity or justify current blocks statistically]*

5: Fixing the trawl survey catchability coefficient for the recent portion of the time series such that the average product of catchability and selectivity across the 60-81 cm size range equals the point estimate obtained by Nichol et al. (2007)

YC35: Given the limitation, this may be the best approach one can take. However, the study by Nichol et al. (2007) was effectively based on 11 fish…. More studies (e.g., tagging, acoustic survey to identify Pacific cod vertical distribution, and comparing catch from varying headlines) are needed to improve our understanding of survey catchability. *[Summary: conduct more studies on survey catchability, including archival tags]*

CD21: Adding to the data base of tags and releases in a larger area will enhance amount of information available for fitting the assessment model. *[Summary: increase the area of release in tagging studies]*

CD22: It was a concern that a large proportion of the tags (released in the initial FIT study) were returned very soon after the study started, which would imply a much higher exploitation rate than that estimated by the assessment. This was discussed with those conducting the experiment who explained that the tags were returned by vessels fishing in the area of the tagging very soon after release. It would be valuable to attempt to guesstimate the mortality rates of the tags in time in order to ensure that localized high exploitation rates are not resulting in problems. If possible, it would be useful to piggy-back tagging studies, using conventional tags, onto the data storage tag studies to enable gear selection to be estimated especially at the largest fish sizes. *[Summary: add conventional tagging studies to future archival tag studies]*

JD18,19: The Nichol et al. study provided valuable insight into survey selectivity, but relied on a few archival tags, resulting in estimates with poor precision. The assessments should continue to use the Nichol et al. estimates, but any further work along these lines should be encouraged. *[Summaries: retain catchability estimates corresponding to Nichol et al. (2007); conduct more studies on survey catchability, including archival tags]*

6: Fixing the natural mortality rate at the value corresponding to Jensen's (1996) Equation 7

YC36: At this point, M, estimated based on Jensen's method, is perhaps the most reasonable choice. *[Summary: retain use of Jensen's equation to estimate M]*

YC37: However, I believe age at maturity used to estimate M should be corrected if any ageing errors were defined either inside or outside the model. *[Summary: use unbiased age at maturity when applying Jensen's equation]*

YC38: In the future, if a Bayesian approach is used in the assessment, I recommend that informative priors be derived for M using M values estimated with different methods. *[Summary: if approach is Bayesian, derive M prior from alternative estimators]*

CD23: Internal estimation of M was attempted in analysis CIE8. The model fit was considerably worse indicating that there is not sufficient information within the current structure to develop alternative values. The comments in this section apply to both the GOA and BSAI cod assessments. *[Summary: estimate M externally]*

CD24: Natural mortality estimates have been estimated in previous assessments and were found to be close to those used currently. Therefore the current fixed values are considered appropriate. *[Summary: retain current estimates of M]*

CD25: It is likely that natural mortality varies (decreases) with age/size as has been estimated using multispecies models for the North Sea by ICES working groups; however until such studies are available for the Pacific cod the single value is considered appropriate to the current state of knowledge for the stocks and the information contributing to their assessments. *[Summary: once data are sufficient, use M-at-age from multispecies models]*

CD26: As more information/studies becomes available, the externally estimated value can be updated; but this should follow a full review of the model fits and consequences for management in a benchmark meeting and not within the annual assessment process that is conducted each year. *[Summary: change M value only during off-cycle "benchmark" meetings]*

JD20: The continued use of the Jensen-based natural mortality estimates is sensible, unless other reliable studies (aimed at estimating natural mortality for Pacific cod) come to light. *[Summary: retain current estimates of M unless studies indicate otherwise]*

7: Input sample sizes for size composition and age composition data, and input log-scale standard deviations for survey abundance data

YC39: The variation calculated from the BS survey may not be correct because the current calculation of standard error implicitly assumes that the survey follows a stratified random design, while the actual survey follows systematic survey design. The standard deviation for the BS survey should be re-calculated using the method consistent with the survey design. *[Summary: adjust survey variances to account for non-random design]*

CD27: Early in the review it was highlighted by one of the panel members that the survey variance calculations for the abundance indices were based on the formulation for random stratified surveys. *[Summary: adjust survey variances to account for non-random design]*

CD28: The rescaling to an average of 300 balances the weighting given to the information from the age and the gear and season size composition sources. This makes the assumption that data collected for ages and size compositions are of equal quality/value in the fitted model. Data collected within a data source, for instance size distributions from a fleet and season, maintain their relative weight within that information set; this is appropriate. *[Summary: retain current method for computing input N for multinomial]*

CD29: If iterative fitting of the model using reweighting according to effective sample size is used, it is possible that multi modal length distributions resulting from incoming recruitment year classes at the smaller sizes could be downweighted at the expense of simpler size composition distributions. Similarly fleets that have a very restricted selection range and simple distribution pattern such as the pot fishery would be given a very high weighting at the expense of those with a broader range that encompasses a number of modes from different year classes. This option was explored between assessments CIE11 and CIE12 – the fit of the model to the simpler age structure of the combined commercial fleets in each season dominated the model fit and the survey size distributions with more modes were considerably down-weighted within the final model. *[Summary: do not iteratively reweight input N for multinomial]*

JD21: The process for deriving estimates of input sample size external to the model appears to be sensible and should continue. *[Summary: retain current method for computing input N for multinomial]*

JD22: In order to investigate the influence of fishery size composition data on model outputs, an additional run was carried out for which the size composition data received very low weight in the model fit…. The fishery size composition data could not be entirely discounted (i.e. allocated zero weight) because the data were still needed to estimate the fishery selectivity parameters. Compared with base run CIE0, there are differences in the model output (e.g. larger Linf and large stock size at the start of the time-series for CIE6), indicating that the fishery size composition data are having an impact, but general stock trends are similar. Importantly, however, inclusion of the fishery size composition data leads to more precise estimates of stock size (compare for example "ts7 Spawning biomass (mt) with 95 asymptotic intervals intervals.png" for the two models), which is important for the provision of management advice. *[Summary: do not downweight fishery sizecomp data]*

JD23: Consideration should be given to a reviewer's alternative suggestion to use number of stations/trips rather than number of samples. *[Summary: consider setting input N for multinomial equal to number of trips]*

JD24: The estimation of input standard deviations for the survey abundance data relies on the assumption of randomness, but the EBS survey has a stratified systematic design, implying these standard deviation estimates are not appropriate, and their estimation should be re-visited. *[Summary: adjust survey variances to account for non-random design]*

8: Allowing for annual variability in trawl survey selectivity

YC40: I recommend that a general linear model (GLM) and/or general additive model (GAM) be developed to include variables that are considered to be important in influencing survey catchability (e.g., temperature, bottom type, location, depth etc.) for developing a standardized survey abundance index. Such indices can remove annual variations in catchability, thus improving the quality of the input data and reducing the complexity of stock assessment model configuration…. Although SS3 has a built-in capacity to accommodate potential temporal trends in selectivity/catchability/availability, I suggest standardizing survey abundance index outside the SS3 to remove the temporal trend in selectivity/catchability/availability. The temporal trend in selectivity/catchability/availability identified in the standardization can also be compared with the temporal trend derived in the SS3 to identify possible differences. *[Summary: standardize survey abundance to remove environmental trends]*

CD30: The surveys design is standardised as far as possible in terms of the trawl gear used, the time and method of deployment, the vessels used to conduct the survey and the sampling procedures. There may be variation in the availability of cod to the survey as a result of environmental change. Studies have established that the spatial distribution of catch rates is related to the distribution of bottom water temperature in the year of the survey. The stratified design should cope with this change but it would provide an interesting PhD to analyse the potential effects of the changes. *[Summary: analyze effects of environmental changes on survey selectivity]*

CD31: Given the standardization of the survey it is surprising that the models are allowing for changes to survey selectivity, at the youngest sizes/ages, which the survey design is attempting to minimize. It may be that the models are fitting to noise. *[Summary: force survey selectivity to be constant over time in the model]*

JD25: Survey catchability is strongly influenced by water temperature, and any attempts to incorporate this knowledge and data into assessment to help quantify year-to-year changes in catchability (rather than modelling annual variability in survey selectivity) should be explored. *[Summary: tie changes in survey selectivity to temperature, not time]*

9: Setting the input standard deviation of log-scale recruitment (σ_R) equal to the standard deviation of the estimated log-scale recruitment deviations

YC41: Fixing the σ_R value in the input data from Myers' database or the standard deviation of log recruitment derived in previous assessments may not be appropriate. In a given assessment year, I believe adjusting the input standard deviation of log-scale recruitment (σ_R) equal to the standard deviation of the estimated log-scale recruitment deviations reflects the current recruitment dynamics and is reasonable. *[Summary: estimate σ_R iteratively]*

CD32: I have little experience of this and other reviewers will comment; however, as with the iterative reweighting using effective sample size, … re-weighting of this form can lead to domination of assessments by particular constraints or model components and if used without caution often leads to misleading model fits. *[Summary: do not estimate σ_R iteratively]*

JD26: Consideration should be given to fixing σR externally to some sensible value (e.g. 0.6) rather than using a time-consuming iterative procedure, which may be difficult to justify on statistical grounds. *[Summary: consider fixing σ_R at an assumed value]*

10: Use of survey abundance data and non-use of fishery CPUE data in model fitting

YC42: A habitat suitability modeling approach (e.g., Chang et al. 2010) can be used to identify suitable habitats for the Pacific cod, based on substrate map and ocean observatory data (or model data), to outline potential habitat maps in the BSAI and GOA and evaluate whether survey sampling stations cover the all effective habitat for cod in different age groups. Such an approach can also be used to project possible changes in cod spatial distribution if key habitat variables (e.g., temperature) change. *[Summary: use habitat suitability to evaluate distribution vis-a-vis survey]*

YC43: Fishery CPUE data are not a reliable abundance index for the Pacific cod stock. *[Summary: do not try to fit fishery CPUE data]*

YC44: I suggest developing standardized fishery CPUE data (Stephens and McCall 2004) outside the SS3 to remove factors that may result in temporal variability in fishery catchability (Punt and Walker 2000; Maunder and Punt 2004). The standardized fishery CPUE for each gear can then be compared to that of each other gear and with the standardized survey abundance index outside the SS3 model to evaluate differences in their temporal trends and develop hypotheses to explain possible differences. Such an analysis outside the stock assessment model can cross check the data that play critical roles in quantifying temporal trends of stock biomass and identify factors that may influence survey catchability and fishery CPUE. Attentions should be paid to those factors identified as important in influencing survey catchability so that caution can be taken in future surveys to minimize impacts of these factors on survey catchability. *[Summary: standardize fishery CPUE data]*

YC45: Current fishery CPUE data are not used in model fitting. However, these data are still included in the model, which may create confusion. I recommend that the fishery CPUE data that are not used in model fitting be removed from the model. *[Summary: remove fishery CPUE data from the model]*

YC46: If any analysis needs to be done between predicted stock biomass and CPUE of a fishery, they can be done outside the model to avoid confusion. *[Summary: compare survey and fishery CPUE externally]*

CD33: The trawl survey for the BSAI cod stock is separated into two periods from 1981 and earlier (three years), and 1982 onwards as a result of a gear change; the data from 1979 - 1981 do not include age structure information. The early period data would not be expected to influence current stock size estimates to any significant degree, the fit of the size composition curves is relatively poor for the survey, and therefore there would seem to be little point in retaining it within the model fit. *[Summary: remove pre-1982 survey data from the EBS model]*

CD34,35: The exclusion of fishery CPUE data from model fits is common practice. Unless standardized the datasets can be:

1. representative of localized concentrations of the stock at particular times of year,
2. affected by gear improvements changing catchability, and
3. altered by management actions, market and fuel prices.

The current assessment fits the commercial CPUE data without using it in the objective function. This provides illustrative trends for comparison with the model results and is considered appropriate. The problem that will be encountered is explaining why the trends may differ if affected by the factors listed. *[Summaries: retain fishery CPUE data in the model; do not try to fit fishery CPUE data]*

JD27: If there is no compelling reason to remove the pre-1982 data for EBS cod, then they should be retained. *[Summary: retain use of pre-1982 survey data in the EBS model]*

JD28,29: Survey data are key to the Pacific cod assessment and should continue to form the basis of the assessments. Continued inclusion of the fishery CPUE data in assessment models (although they are not fitted) is useful for comparative purposes, and allows an independent check on model outputs. *[Summaries: retain fishery CPUE data in the model; do not try to fit fishery CPUE data]*

III. Recommendations on topics other than those contained in the Terms of Reference

A. General modeling approach

YC47: In-depth analysis should be conducted to identify possible sources of uncertainty for a given set of data and relevant analysis should be done to reduce the uncertainty and improve data quality BEFORE the data are used in the stock assessment model. *[Summary: identify/reduce uncertainty, improve quality of all data before use]*

YC48: Given the flexibility and many choices that SS3 provides for functions quantifying life history and fishery processes, one needs to use background information of the collection of fishery and survey data, fish life history theory, and local ecosystem to develop hypotheses to explain choices and resultant estimates. If a result cannot be justified in a reasonable way, the assessment should be evaluated. *[Summary: justify choices/estimates involving life history, fishery processes]*

YC49: The recruitment is currently measured as the number of age 0 fish in the Pacific cod stock assessment. I understand the number of age 0 fish is simply a reflection (discounted for natural mortality) of the number of fish in older ages (say 3) because there is no fishing mortality. However, given that age 0 implies larval stage and that there are no observations in survey and fishery, the biological meaning of the so-called recruitment is inappropriate and not well-defined. As it is defined, the current recruitment is neither representative of fishery recruitment nor the number of fish larvae. Rather, it is an index of the recruitment. Although this may not be an issue to fisheries stock assessment scientists, such a measure of recruitment may be misused by others who are not familiar with the stock assessment. I believe it is more appropriate to measure the fishery recruitment as the number of fish at an age group at which fish are subject to fishing mortality (e.g., number of fish at age 3). *[Summary: report "recruitment" as the number of fish at age 3]*

YC50: A Bayesian approach has not been fully incorporated in the BASI and GOA Pacific cod stock assessment. Thus, uncertainty in the assessment has not been fully incorporated in the assessment and stock projection under different harvest strategies. I would encourage future assessment to fully utilize this function in the SS3. *[Summary: use a fully Bayesian approach]*

CD36,37: The need for such a time consuming process (jittering) results from the model structure pushing the number of estimated parameters to the edge of what is estimable; the models are or are close to being over-parameterised. The problem affects the review, the development time that the assessor can spend on testing and evaluating the model and the quality control and sensitivity analysis that can be applied. There is a trade-off between the number of parameters fitted and the practicality of the fitting in terms of the time available for development, review and reporting to management. The stock assessments and the assessor would benefit from reducing the parameterization, accepting that there will be uncertainty in model estimates and developing management procedures that evaluate and allow for that uncertainty. The management plan evaluations described by Teresa A'mar could form the basis for such a change but they will be extremely difficult for such a complex, slow, model. *[Summaries: reduce the number of parameters in the models; if fewer parameters used, adjust for added uncertainty via MSE]*

JD30,31: The need for a time-consuming process of "jittering" for each new model run to avoid local minima and general problems of lack of convergence point to the data and model configuration being pushed close to the limit in terms of being estimable. This problem affected the effectiveness of the review, because on the whole, jittering was not possible during the meeting due to time constraints, and panel members could not be confident (to the extent jittering gives such confidence) that results presented during the meeting reflected the best fit for a given model configuration. More seriously, however, it raises the possibility that the current models for EBS and GOA cod are too close to being over-parameterised. There are procedures for investigating parameter redundancy (see e.g. Gimenez et al. 2004), and perhaps some of these should be employed for these models, if practicable. The model configuration for CIE11 is one attempt towards simplification that may have some merit, and further attempts along these lines should be encouraged. *[Summaries: investigate parameter redundancy; reduce the number of parameters in the models]*

B. Possible future improvements to SS and R4SS

YC51: Outliers are likely to exist in input data used in the assessment, given that the data are derived from different sources and are subject to different levels of errors. They may bias parameter estimation in stock assessment. Robust likelihood functions can reduce impacts of outliers in size composition and survey abundance index (Chen et al. 2003). *[Summary: include "robust" likelihood functions in SS]*

CD38: It is assumed that once a new SS program has been received it is tested by the assessment authors to the extent that it can reproduce the previous assessments results with the same data. In addition if not already available a test data set with known parameter estimates and uncertainty that would be used to benchmark new versions should be considered. *[Summary: develop test data set to "benchmark" new versions of SS]*

JD32: During the meeting, a couple of potential coding problems in SS were identified. The first has already been mentioned under TOR 2b above. The second relates to the lack of fit to the 2010 trawl survey size composition data at the smallest sizes…. Given that these are mostly age 1, and given that the recruitment deviation has nothing else to fit to, this lack of fit is surprising and may be indicative of a coding error. *[Summary: see whether bad fit to 2010 survey at small sizes is a coding error]*

JD33: Particularly helpful during the meeting was to have the participation of another experienced modeller (Teresa A'mar) who also had experience with using a graphics tool that could convert SS model output into graphical displays (R4SS) – this proved very useful and essential for the review process. Nevertheless, the graphics tool had some features that could be improved (e.g. it was not always clear what some graphs referred to, and there were some problems with duplicated or failed outputs). *[Summary: clarify graphs, reduce redundancy, improve robustness in R4SS]*

JD34: The fit to the GOA 1990 May-Aug Trawl survey size composition data produces enormous residuals at the smaller sizes in "comp_lenfit_residsflt2sex1mkt0.png", but these do not seem to show up in "comp_lenfit_flt2sex1mkt0.png" – this may be easily explained, but needs looking into in case there is a problem. *[Summary: see why sizecomp fits, residuals do not always match in R4SS]*

JD35: The model outputs from SS are not user-friendly, and in particular parameter names are not intuitive or easy to identify (e.g. MGparm[4]?), so one suggestion is that a similar tool be developed for non-graphical output so that model parameters and other useful diagnostics (e.g. likelihood component values and RMSE "scores") are easily identified and interpreted – this would be a huge help for reviewers, and assessment authors may also find it a timesaving device for the own purposes. *[Summary: expand R4SS to summarize non-graphical output]*

C. Future use of non-SS models

YC52: I believe some competitive models at different complexities should be developed for comparison with the SS3. Dr. Teresa A'mar of AFSC is currently developing an operating model for management strategy evaluation (MSE). With some modifications, this model has the potential to be used as a stock assessment model. A comparative study of stock assessment, begot from different models, can help improve understanding of fish population dynamics modeled by the SS3. *[Summary: add non-SS-based models, with varying levels of complexity]*

CD39: The complexity of the SS program makes it difficult to compare the assessment results with runs using other assessment programs, however, this should be attempted particularly with simpler models, e.g. survey based, using alternative assessors to ease the burden on the current one. *[Summary: add simpler, non-SS-based models, using other assessors]*

CD40: Given the data structures available for the assessment there are few if any alternative models for the final assessment. Given the high dependency on the one system, a custom built approach could be developed (as a research project?) to provide an alternative; alternatively a test data set that reproduces the characteristics of the cod stocks should be considered (as is being constructed by Teresa A'mar) as a priority so that evaluation of the current model formulations and changes to them can be examined against known solutions. *[Summary: develop EBS and GOA Pcod test data sets to evaluate models]*

JD36: Another issue is the debate about whether stock assessment should be "custom-built", or whether "off the shelf" modelling frameworks should be used. There are pros and cons on both sides of the argument…. There are a few examples of compromises for the Pacific cod models to enable the SS framework to continue to be used (e.g. lack of bi-modal selection for GOA leading to a split in the survey data, and lack of constrained random walk over time leading to selectivity by time blocks), but given that these models appear to have reached their limit in terms of complexity within SS (a cause of the jitter problem?), perhaps now is the time to revisit this debate? *[Summary: consider replacing SS-based models entirely]*

D. Annual assessment and review processes

YC53: I recommend that retrospective analysis be conducted for all models considered in the stock assessment to evaluate nature (positive or negative) and magnitude of retrospective errors…. Retrospective errors should be carefully evaluated for the estimates of stock biomass, fishing mortality, and recruitment. *[Summary: conduct retrospective analyses of all models]*

YC54: Previous efforts were focused on accommodating many different requests for model configurations. I believe more effort should be spent on model diagnoses to identify if the model assumptions, implicit and explicit, have been violated. This involves evaluating residual patterns for distributional assumptions, CVs of each estimated parameters to identify if an estimated parameter is significant, and the variance-covariance matrix to identify possible correlations between different parameters (and then to see if such a correlation can be justified biologically). *[Summary: increase attention to residual patterns and variances/covariances]*

YC55: The model used in the previous year's assessment model should be included automatically in the next year's assessment as a background check for the model consistency. *[Summary: always include previous year's model in the new assessment]*

YC56: Future assessment should try to keep the stock assessment model relatively stable to avoid among-model variability over years. *[Summary: keep the assessment model relatively stable over time]*

YC57: Many model configurations were used over the time. I recommend analyzing among model variations (for all the final models used different years) to improve understanding of the model performance and possible management implications of making changes to the models over time. *[Summary: examine effects of model changes on performance, management]*

YC58: The Plan Team and SSC need to discuss and recommend a set of criteria that are well defined and measureable for choosing the stock assessment model. *[Summary: determine model selection criteria in advance]*

YC59: The Pacific cod may have a metapopulation structure in the BSAI. This stock spatial structure may call for separate area management for the BS and AI. A separate stock assessment for BS and AI seems to be a logical way to start this process. *[Summary: develop separate stock assessments for BS and AI]*

CD41: There is a heavy reliance on a key stock assessor for the production and presentation of the assessment and output for the two stocks. This reliance on one person could present problems and can result in an excessive workload at key times, especially if the stocks decline towards the SSB threshold at which severe restrictions are imposed. If, as has been suggested, the Bering Sea and Aleutian Islands assessment region is divided into two stocks, … then the workload of the key assessor will become impractical. *[Summary: if BS and AI assessments are separated, use different assessors]*

CD42: Part of the heavy workload results from the requirement for the assessor to run a series of exploratory models as suggested by members of the public, reviewers etc. prior to each annual meeting. This is considered excessive and can place undue pressure on the assessment team whilst also introducing a perception of uncertainty/instability with respect to the assessment process. *[Summary: reduce number of exploratory models]*

CD43: ICES has introduced a system of benchmarking of its assessments in which assessment models are reviewed at a scientific meeting which agrees the best model structure and data sources available at that time. The structure and data sources are then frozen, apart from the addition of new data each year, and the assessment run as an update for a fixed number of years - unless evidence is presented of the need for a new review. At the end of the agreed time frame the process is repeated, the biology of the stock, available data and potential models are investigated, information sources agreed and the cycle restarted…. Such a cycle would allow the stock assessors to concentrate on each stock in alternate years (for instance) so that development can be evaluated in a more relaxed time frame compared the current system which is trying to deliver the best science for two (potentially three) stocks simultaneously. *[Summary: freeze model structure for a pre-determined number of years]*

CD44,45: One way in which the workload could be reduced is to separate the information within the assessment report into two documents; currently the report has a split personality. It tries to present the technical aspects of the collection of the new data available each year from the surveys and observer program, the diagnostics from the model fit to the updated data and also provide a non-technical summary of the output for managers and the SAFE report. The report does not provide the full set of details required for a full and detailed review of the model. This is especially the case when a variety of runs have been evaluated following suggestions from the members of the public and management team. It cannot summarize the build up to the final assessment, sensitivity analysis and consequences for management without being too large to produce each year. An approach that has been used elsewhere is the production of an annual technical report that can be used by reviewers and a summary report for managers that can be updated with new information each year if it is available and relevant. A lot of what is required for the technical report can be automated. *[Summaries: split assessment report into "technical" and "summary" reports; add more detail to the technical assessment report]*

CD46,47: As part of the review process it was very difficult to determine the degree of variation that has occurred in the estimated stock and management metrics between the consecutive assessments. ICES and others produce two forms of quality control diagrams, as part of their annual reporting, that give insight into the variation from year to year in the perception of stock status:

1. Retrospective analysis - the final agreed model structure fitted, stepping backwards in time, removing a year of data each time
2. Quality control diagrams – showing the results of the final agreed assessment from each year

It is suggested that as part of the reporting process such diagrams and their equivalent on a relative scale (e.g. SSB / SSB25% as that is the scale used for management) be considered. *[Summaries: conduct retrospective analysis of final model; add time series of all historical assessment results to SAFE]*

JD37: A related point is that the annual process of coming up with the best assessment seems to have become extremely time-consuming, and raises the question about whether things really are changing that much from year to year (reflected by year-to-year changes in model structure), or whether one is just essentially modelling noise…. An alternative approach would be to settle on a particular model structure for a longer period (say 3-5 years), because real change would probably only be detected on such a time-scale anyway. Of course, detailed work on the next model can continue in the interim period, making use of the latest scientific research, but also keeping an eye on the current model to make sure that assumptions are not violated to the extent that the model leads to poor management decisions. *[Summary: freeze model structure for a pre-determined number of years]*

E. Harvest strategy evaluation

YC60: Although the SS3 has projection capacity, it has no built-in component for MSE. I believe ongoing research efforts to develop an MSE framework for the Pacific cod can provide an important analytical tool to evaluate alternative management strategies and their associated risks. *[Summary: continue existing MSE work]*

YC61: Recent assessments incorporate the model projection. I recommend that the performance of the projection done in the past assessment be evaluated, retrospectively, to evaluate their performance in achieving the management objectives. *[Summary: evaluate performance of last year's projection vis-à-vis objectives]*

CD48: The harvest strategies for the two cod stocks cod (and for other fish stocks in the region) are constructed from sound theoretical reference levels for fisheries systems assumed to be in equilibrium. However, even though the mortality rate has remained well below the target level, following a series of low recruitments to the stock, there was been a decline in SSB to just above B35% for both cod stocks…. This suggests that although a HCR based on the theoretical equilibrium population structure might be expected to perform well, in reality if fishing at the MaxFABC had been permitted the current management plan structure could lead to closure of the fishery with greater frequency than would be expected. The response of the stock at lower levels of exploitation than defined by the HCR, suggests that the HCR may not robust to auto-correlation…. It is suggested that, if they have not already been conducted in the design of the current HCR, evaluations of the HCR of the form described by Teresa A'mar in are conducted. Recruitment autocorrelation should be part of the operating model in order to evaluate the performance of the current HCR with recruitment series that approximate the observed series rather than based on random re-sampling from a fitted equilibrium curve. *[Summary: incorporate recruitment autocorrelation into existing MSE work]*

CD49: The presentation by Teresa A'mar discussed ongoing work to evaluate the management plan used for the cod stocks. This should be fully supported. This recommendation is based on a series of observations from the review process:

1. The first concerns the decrease in stock biomass when the exploitation rate has been low throughout the recent time period in comparison to the potential target levels that could be achieved under the management plan.
2. The second observation is that the cod review raised a number of questions that may not have well defined estimates (e.g. natural mortality levels) but the sensitivity of the model estimates and the outcome of the harvest control rule to their effects could be evaluated and included within modified plans. Some suggestions for the study would be:
    a) The sensitivity of the stock and fishery outcomes to autocorrelation in recruitment rather than based on random re-sampling from a fitted equilibrium curve.
    b) The assumptions concerning natural mortality.
    c) The form of the stock and recruit relationship.
    d) The lack of agreement in ageing cod and the impact of bias.
    e) The frequency of the trawl survey series in the GOA.

Whilst the study would not a definitive answer to all issues, especially as modeling the cap on total catch in the Bering Sea would is problematic, it would highlight key areas of model and HCR sensitivity that could be addressed by modifications to the rule. *[Summary: continue existing MSE work]*

Table A1. Recommendations (p. 1 of 3). "Sec."=section, "Sub."=subsection, "Rec."=recommendation, "Tot." = total recommendation number, "Uni." = unique recommendation no. (duplicates excluded)

Sec.	Sub.	Rec.	Tot.	Uni.	Summary of recommendation
II	1a	YC1	1	1	continue existing research on age determination/validation
II	1a	YC2	2	2	estimate ageing error externally
II	1a	YC3	3	3	retain use of age composition data
II	1a	YC4	4	4	explain missing 2-year-olds in GOA survey
II	1a	YC5	5	5	expand existing research on age determination/validation
II	1a	YC6	6	6	downweight age and length data if both are used
II	1a	CD1	7	7	retain current otolith and length sampling procedures
II	1a	CD2	8	8	retain use of ageing error matrix
II	1a	CD3	9	3	retain use of age composition data
II	1a	CD4	10	9	if ageing criteria change, include fishery age composition data
II	1a	JD1	11	3	retain use of age composition data
II	1a	JD2	12	8	retain use of ageing error matrix
II	1a	JD3	13	1	continue existing research on age determination/validation
II	1a	JD4	14	10	include fishery age composition data
II	1b	YC7	15	11	if length-at-age variance estimated externally, omit size-at-age data
II	1b	CD5	16	12	if cohort-specific growth not used, omit size-at-age data
II	1b	CD6	17	13	downweight age and size-at-age data if both are used
II	1b	JD5	18	14	investigate appropriate weighting of non-independent data
II	1c	YC8	19	15	estimate ageing bias externally
II	1c	CD7	20	16	consider variable/flexible growth as an alternative to ageing bias
II	1c	CD8	21	17	constrain ageing bias to increase with age
II	1c	CD9	22	18	revise SS to allow for asymmetric ageing bias
II	1c	JD6	23	19	explore internal estimation of ageing bias
II	1d	YC9	24	20	estimate length-at-age variance by otolith back-calculation
II	1d	CD10	25	21	estimate length-at-age variance externally
II	1d	CD11	26	22	retain current procedure for estimation of length-at-age variance
II	1d	CD12	27	23	investigate apparent curvilinearity of length-at-age variance
II	1d	JD7	28	22	retain current procedure for estimation of length-at-age variance
II	2a	YC10	29	24	retain current partitioning of catch data by year, season, gear
II	2a	YC11	30	25	evaluate variability of catch data quality by year, season, gear
II	2a	YC12	31	26	include catch from all sources
II	2a	YC13	32	27	change requirements for observer coverage
II	2a	YC14	33	28	compare and cross-validate catch data from different sources
II	2a	YC15	34	29	evaluate current catch sampling/reporting system via simulation
II	2a	YC16	35	30	estimate catch uncertainty
II	2a	YC17	36	31	once catch uncertainty has been estimated, evaluate its impacts
II	2a	CD13	37	32	retain current seasonal structure for catch and selectivity
II	2a	CD14	38	33	do not aggregate catch across gears if selectivity is held constant
II	2a	JD8	39	30	estimate catch uncertainty
II	2b	YC18	40	34	retain current partitioning of sizecomp data by season and gear
II	2b	YC19	41	35	quantify uncertainty associated with fishery sizecomp data
II	2b	YC20	42	36	evaluate variability in quality of sizecomp data by year, season, gear
II	2b	YC21	43	37	justify blocks based on analysis of factors that may affect selectivity
II	2b	YC22	44	38	omit fleets that have minimal impacts on the assessment
II	2b	YC23	45	39	explore dynamic binning
II	2b	YC24	46	40	evaluate precision of length measurements
II	2b	YC25	47	41	evaluate effects of recent SSL area closures on sizecomp data
II	2b	YC26	48	42	evaluate effects of survey sampling design on sizecomp precision
II	2b	CD15	49	43	investigate treatment of sample size in SS when merging bins

Table A1. Recommendations (p. 2 of 3). "Sec."=section, "Sub."=subsection, "Rec."=recommendation, "Tot." = total recommendation number, "Uni." = unique recommendation no. (duplicates excluded)

Sec.	Sub.	Rec.	Tot.	Uni.	Summary of recommendation
II	2b	JD9	50	43	investigate treatment of sample size in SS when merging bins
II	2b	JD10	51	44	explore coarser bin structure for large sizes
II	2c	CD16	52	45	do not use the existing small sample of fishery agecomp data
II	2c	CD17	53	46	include a bimodal, parametric selectivity curve in SS
II	2c	JD11	54	47	retain partitioning of GOA survey data into sub-27 and 27+ ranges
II	2c	JD12	55	48	explore possible over-parameterization of GOA sub-27 catchability
II	2c	JD13	56	46	include a bimodal, parametric selectivity curve in SS
II	3	YC27	57	49	retain use of von Bertalanffy growth
II	3	YC28	58	50	estimate a_0 externally
II	3	YC29	59	51	estimate growth internally only if parameter correlations justified
II	3	YC30	60	52	consider estimating all growth parameters externally
II	3	CD18	61	53	include more flexible growth functions in SS
II	3	JD14	62	53	include more flexible growth functions in SS
II	4	YC31	63	54	develop hypotheses to explain derived selectivity curves
II	4	YC32	64	55	consider the possible effect of partial availability to the fishery
II	4	YC33	65	56	retain current partitioning of selectivity by season
II	4	YC34	66	57	evaluate selectivity trend using random walk, one fleet at a time
II	4	CD19	67	58	if possible, link selectivity to tagging studies or other assessments
II	4	CD20	68	59	use random walk selectivity, hold constant where change is small
II	4	JD15	69	60	force just one selectivity to be asymptotic, justify this assumption
II	4	JD16	70	61	explore bimodal selectivity for GOA survey
II	4	JD17	71	62	use random walk selectivity or justify current blocks statistically
II	5	YC35	72	63	conduct more studies on survey catchability, including archival tags
II	5	CD21	73	64	increase the area of release in tagging studies
II	5	CD22	74	65	add conventional tagging studies to future archival tag studies
II	5	JD18	75	66	retain catchability estimates corresponding to Nichol et al. (2007)
II	5	JD19	76	63	conduct more studies on survey catchability, including archival tags
II	6	YC36	77	67	retain use of Jensen’s equation to estimate M
II	6	YC37	78	68	use unbiased age at maturity when applying Jensen’s equation
II	6	YC38	79	69	if approach is Bayesian, derive M prior from alternative estimators
II	6	CD23	80	70	estimate M externally
II	6	CD24	81	71	retain current estimates of M
II	6	CD25	82	72	once data are sufficient, use M-at-age from multispecies models
II	6	CD26	83	73	change M value only during off-cycle “benchmark” meetings
II	6	JD20	84	74	retain current estimates of M unless studies indicate otherwise
II	7	YC39	85	75	adjust survey variances to account for non-random design
II	7	CD27	86	75	adjust survey variances to account for non-random design
II	7	CD28	87	76	retain current method for computing input N for multinomial
II	7	CD29	88	77	do not iteratively reweight input N for multinomial
II	7	JD21	89	76	retain current method for computing input N for multinomial
II	7	JD22	90	78	do not downweight fishery sizecomp data
II	7	JD23	91	79	consider setting input N for multinomial equal to number of trips
II	7	JD24	92	75	adjust survey variances to account for non-random design
II	8	YC40	93	80	standardize survey abundance to remove environmental trends
II	8	CD30	94	81	analyze effects of environmental changes on survey selectivity
II	8	CD31	95	82	force survey selectivity to be constant over time in the model
II	8	JD25	96	83	tie changes in survey selectivity to temperature, not time
II	9	YC41	97	84	estimate σR iteratively
II	9	CD32	98	85	do not estimate σR iteratively

Table A1. Recommendations (p. 3 of 3). "Sec."=section, "Sub."=subsection, "Rec."=recommendation, "Tot." = total recommendation number, "Uni." = unique recommendation no. (duplicates excluded)

Sec.	Sub.	Rec.	Tot.	Uni.	Summary of recommendation
II	9	JD26	99	86	consider fixing σR at an assumed value
II	10	YC42	100	87	use habitat suitability to evaluate distribution vis-a-vis survey
II	10	YC43	101	88	do not try to fit fishery CPUE data
II	10	YC44	102	89	standardize fishery CPUE data
II	10	YC45	103	90	remove fishery CPUE data from the model
II	10	YC46	104	91	compare survey and fishery CPUE externally
II	10	CD33	105	92	remove pre-1982 survey data from the EBS model
II	10	CD34	106	93	retain fishery CPUE data in the model
II	10	CD35	107	88	do not try to fit fishery CPUE data
II	10	JD27	108	94	retain use of pre-1982 survey data in the EBS model
II	10	JD28	109	93	retain fishery CPUE data in the model
II	10	JD29	110	88	do not try to fit fishery CPUE data
III	A	YC47	111	95	identify/reduce uncertainty, improve quality of all data before use
III	A	YC48	112	96	justify choices/estimates involving life history, fishery processes
III	A	YC49	113	97	report "recruitment" as the number of fish at age 3
III	A	YC50	114	98	use a fully Bayesian approach
III	A	CD36	115	99	reduce the number of parameters in the models
III	A	CD37	116	100	if fewer parameters used, adjust for added uncertainty via MSE
III	A	JD30	117	101	investigate parameter redundancy
III	A	JD31	118	99	reduce the number of parameters in the models
III	B	YC51	119	102	include "robust" likelihood functions in SS
III	B	CD38	120	103	develop test data set to "benchmark" new versions of SS
III	B	JD32	121	104	see whether bad fit to 2010 survey at small sizes is a coding error
III	B	JD33	122	105	clarify graphs, reduce redundancy, improve robustness in R4SS
III	B	JD34	123	106	see why sizecomp fits, residuals do not always match in R4SS
III	B	JD35	124	107	expand R4SS to summarize non-graphical output
III	C	YC52	125	108	add non-SS-based models, with varying levels of complexity
III	C	CD39	126	109	add simpler, non-SS-based models, using other assessors
III	C	CD40	127	110	develop EBS and GOA Pcod test data sets to evaluate models
III	C	JD36	128	111	consider replacing SS-based models entirely
III	D	YC53	129	112	conduct retrospective analyses of all models
III	D	YC54	130	113	increase attention to residual patterns and variances/covariances
III	D	YC55	131	114	always include previous year's model in the new assessment
III	D	YC56	132	115	keep the assessment model relatively stable over time
III	D	YC57	133	116	examine effects of model changes on performance, management
III	D	YC58	134	117	determine model selection criteria in advance
III	D	YC59	135	118	develop separate stock assessments for BS and AI
III	D	CD41	136	119	if BS and AI assessments are separated, use different assessors
III	D	CD42	137	120	reduce number of exploratory models
III	D	CD43	138	121	freeze model structure for a pre-determined number of years
III	D	CD44	139	122	split assessment report into "technical" and "summary" reports
III	D	CD45	140	123	add more detail to the technical assessment report
III	D	CD46	141	124	conduct retrospective analysis of final model
III	D	CD47	142	125	add time series of all historical assessment results to SAFE
III	D	JD37	143	121	freeze model structure for a pre-determined number of years
III	E	YC60	144	126	continue existing MSE work
III	E	YC61	145	127	evaluate performance of last year's projection vis-à-vis objectives
III	E	CD48	146	128	incorporate recruitment autocorrelation into existing MSE work
III	E	CD49	147	126	continue existing MSE work

Appendix B:
Pacific cod excerpts from recent Plan Team and SSC minutes

Table of Contents

Excerpt from the minutes of the BSAI Plan Team (Nov. 2010; no recommendations)

The joint Teams accepted the author's preferred Model B (see Joint Team Minutes). Therefore the remaining issue for the BSAI Team was the OFL and ABC recommendations and ABC area apportionments.

Mike Sigler accepted the model, but suggested that the values of natural mortality and trawl survey catchability were uncertain; he noted that the stock size estimates included a lot of small fish from incoming year classes. Bill Clark observed that the uncertainty of M and q were not very different from other assessments and had been fully discussed in September. Grant Thompson said that small fish were only a small part of the author's recommended ABC for 2011. The Team approved the author's recommended OFL and ABC, set according to the standard control rule for a Tier 3b stock. Still, because of the influence of the incoming 2006 and 2008 year classes on projected biomass, the Team notes that the 2012 estimate may be lower next year than projected this year.

Kerim Aydin observed that in the absence of an area apportionment between the Bering Sea and Aleutian Islands, the exploitation rate of cod in the Aleutian Islands continued to be about twice that in the Bering Sea (based on simple ratios of catch and survey abundance), and biomass continued to decline in the Aleutian Islands. A member of the public commented that for various reasons (including Steller sea lion mitigation measures) cod catches in the Aleutians were unlikely to increase and were very likely to decline in 2011. The Team is nonetheless still concerned about the disproportionate exploitation of cod in the Aleutian Islands and recommends the earliest possible implementation of separate area ABCs.

Applying the Kalman filter approach to the updated (through 2010) time series indicates that the best estimate of the current biomass distribution is 91% EBS and 9% AI, replacing the previous proportions of 84% and 16% respectively.

The author informed the Team of his plans to develop a separate AI Pacific cod assessment in the near future.

Excerpt from the minutes of the GOA Plan Team (Nov. 2010; recommendation highlighted)

The Plan Team accepts model B, and the associated ABC and OFL levels with the caveats and concerns about the discrepancy between the pattern of last years numbers at age and those estimated in this assessment. The Team appreciated the authors' effort in reducing the number of models for presentation.

The Team questioned why the pattern in numbers at age is so different this year compared to last year's assessment given that very little data has been added. In particular, the 2009 survey showed lots of one-year olds but they do not appear to be reflected in the model estimates. This appears to result in a declining trend in the projection model compared to a rapidly increasing trend from last year's version. It was noted that the numbers at age used in last years projection model will be different than the numbers at age for this year's model. The difference may be in the demographic parameters as specified (there were some difficulties converting stock synthesis output to age-specific schedules required for the projection model) but should be explained.

For all models, the recruitment deviation in 2008 appears to go to zero (as reflected in Figure 2.2b) and that appears contrary to the 2009 survey data. The senior author noted that the selected model had survey catchability deviations set to zero in 2009 (along with the recruitment deviation). Also, size at age 1 is really different last couple of years.

The Team noted that it would be useful to have a presentation of the estimates relative to the data, particularly for the most recent survey (and sub-27 cm abundance index). The ABCs in historical perspective indicate that even with a 2012 ABC of 78,200 it would be third highest catch in history (noting that the TAC drops below the ABC due to the state fishery).

Excerpt from the minutes of the Joint Plan Teams (Nov. 2010; non-recommendation highlighted)

Grant Thompson presented the BSAI and GOA assessments, both of which used essentially the same three models. The models were chosen in the course of two rounds of trials and reviews by the Teams and the SSC (in May/June and September/October). Model A was the 2009 preferred model, whose main features were:

(i) Natural mortality M = 0.34 fixed externally.

(ii) Length-specific commercial selectivities, estimated in blocks of years, some forced to be asymptotic. Commercial age compositions fitted where available, length compositions where not. Commercial CPUE not fitted.

(iii) Age-specific trawl survey selectivity with annually varying left limb. Trawl survey age composition and CPUE fitted. The product of catchability and selectivity of 60-80 cm fish required to be 0.47 based on a small set of data from archival tag recoveries.

(iv) IPHC longline survey length compositions (not CPUE) fitted.

(v) Cohort-specific growth parameters, with the standard deviation of length at age estimated externally.

(vi) Aging bias of of +0.4 years at all ages estimated by profiling and accounted for.

(vii) Input standard deviations of a number of parameters estimated iteratively so as to match output standard deviations.

Model B was the same as Model A with some incremental modifications, viz:

(i) Smaller length bins (1 cm instead of 3 and 5) to make full use of the length data.

(ii) Five fishery seasons were modeled instead of 3.

(iii) A single growth schedule was fitted.

(iv) The few fishery length-at-age data were left out.

(v) IPHC survey length data were left out.

(vi) Parameter values estimated iteratively in the 2009 assessment were carried over to Model B.

Model C was the same as Model B but all age composition and length-at-age data were left out because of concern about aging bias.

Recent survey results affected all model fits. GOA survey abundance increased by 200% in 2009 and EBS survey abundance by 100% in 2010.

Convergence was an issue for almost all models. In fitting the models, first a best estimate was located by perturbing ("jittering") the parameter vector at successive local minima. Reproducibility of the best estimate was then tested by jittering the best estimate and refitting many times. The best estimate was seldom relocated. The CV of the present biomass estimate in these trials was about 3% for Model A in the EBS and 10-20% for Models B and C in the EBS and all models in the GOA.

All model fits to EBS survey abundance were good, and to GOA survey abundance similar. All models fitted the catch length compositions well. Models A and B fitted the age compositions well.

Model A approximated the modes in EBS survey length frequencies reasonably well, but Model B less well. Model C matched the modes very closely but at ages that were high by a year because the fitted growth schedule was permitted to be negative at age one. Grant explained that this could happen because there were no age or size-at-age data whatsoever in the model, so the model could fit the data with length-at-age (and survey selectivity at age) shifted relative to Models A and B. This anomaly could easily be fixed.

All models estimated produced similar estimates of EBS trawl survey selectivity. In the GOA the survey selectivity estimates from Models A and B were extremely variable, to the point of being hardly believable. The estimates for Model C were also quite variable but much less so.

Historical abundance estimates for all models were similar in the EBS. In the GOA Models A and B were similar but Model C estimated very high levels of abundance in the 1970s, which Grant thought were impossible.

Grant adopted a number of criteria for choosing a best model, according to which Model B was better than Model A (better bin and season structure, more parsimonious), and Model C was disqualified because of the anomalous length-at-age in the EBS and the impossible abundance estimates in the GOA. Both Teams agreed with Grant's choice of Model B and his rationale.

Grant previewed upcoming developments in the cod assessment: the option in Stock Synthesis of fitting a Richards growth schedule (with positive lengths at age one) instead of the von Bertalanffy, the possibility of estimating aging error internally, a CIE review in March/April, and possibly an Aleutian Islands assessment. In view of the impending CIE review, the Teams did not attempt at this meeting to formulate any requests for modeling work. But we do want the Teams and the SSC to review the CIE recommendations (and any public submissions) in the May/June period before Grant settles on a program of work for the September/October meetings. We would ask REFM to schedule the CIE review accordingly.

Excerpt from the minutes of the SSC (Dec., 2010; recommendations highlighted)

BSAI and GOA Pacific cod

The SSC commends the authors for their thorough and conscientious responses to public, Plan Team, and SSC recommendations. Kenny Down (Freezer Longliner Coalition) provided public testimony on BSAI Pacific cod. He supports the authors preferred model and model estimates and commented that the process was good and many improvements were made such as constant growth. Julie Bonney (Alaska Groundfish Databank) expressed concerns about an increased ABC this year and then declining thereafter.

The Pacific cod assessments and data that went into the assessment have received a great deal of scrutiny over the last few years. There continues to be concern on the accuracy of age readings. Other issues include the natural mortality rate, the trawl survey catchability coefficient, the modeling of commercial selectivity (variable or not, asymptotic or not, fishery by fishery), modeling of survey selectivity, and the modeling of growth (constant, cohort-specific, year-specific).

Since last year, many changes have been considered or made, based on recommendations from the public, the Plan Teams and the SSC. To streamline the model evaluation process, a set of six models were presented in this year's preliminary assessment, as requested by the Plan Teams in May, and reviewed by the SSC in June of this year. Following Plan Team review in September and SSC review in October a final set of three models were requested to be included for final evaluation. The three candidate models (A, B, and C) were considered in developing the 2011 and 2012 OFL/ABC specifications. Model A is identical to the model accepted for use by the BSAI Plan Team and SSC in 2009 and the only model from the preliminary assessment to be carried forward.

Current Models

Model A was the 2009 preferred model. Main features of model A included: 1) natural mortality M = 0.34 fixed externally, 2) length-specific commercial selectivities, estimated in blocks of years, some forced to be asymptotic, 3) age-specific trawl survey selectivity with annually varying left limb, 4) the average product of catchability and selectivity of 60-80 cm fish required to be 0.47, 5) cohort-specific growth parameters, with the standard deviation of length at age estimated externally, 6) Aging bias of +0.4 years at ages 2+ estimated by profiling, 7) Input standard deviations of a number of parameters estimated iteratively so as to match output standard deviations.

Model B was the same as Model A with some incremental modifications including: 1) smaller length bins (1 cm instead of 3 and 5) to make full use of the length data, 2) five fishery seasons were modeled instead of 3, 3) a single growth schedule was fitted, 4) the few fishery length-at-age data and age composition data were left out, 5) IPHC survey length data were left out, 6) values estimated iteratively in the 2009 assessment were carried over to Model B.

Model C was the same as Model B but all age composition and length-at-age data were left out, because of concern about aging bias.

Model Evaluation

The authors used four criteria to evaluate and select the final model. The criteria include: 1) does the model make full use of the information in the size composition data, 2) has the seasonal structure of the model been justified statistically, 3) is the model sufficiently parsimonious, and 4) does the model estimate plausible lengths at age?

SSC Comments and Recommendations

There will be a CIE review of Pacific cod models in early 2011 and information from this review will be used to produce another suite of models that will be considered for PT and SSC review in the spring.

The SSC has a number of model suggestions that may be considered through the next assessment cycle by the author as time permits:

Evaluate reduced catch season and size bin structures that are more parsimonious, but do not diminish the information content.

Trawl survey catchability used in the assessment and model sensitivity to model estimates or plausible alternatives should be evaluated.

Simplifying trawl survey selectivity should be investigated and model fit to data components evaluated.

Re-tune aging bias to try to better match the observed age modes.

Evaluate estimating aging bias within the model.

Evaluate Richards growth curve alternative.

Continued research that would provide information on age-determination errors and potential biases.

Given the divergence in population abundance between the AI and BS the SSC recommends that an AI assessment be brought forward for evaluation (only) during the next assessment cycle. Biomass distribution is currently estimated at 91% EBS and 9% AI compared to previous proportions of 84% and 16%, respectively.

For the GOA, apply a simple Kalman filter approach, as adopted by the SSC in 2004 for BSAI for estimation of current biomass distribution.

Constant growth should be brought forward in future models (run times reduced back to 2-3 minutes).

The SSC offers the following modeling issues that could be considered during the CIE review:

The process of iteratively estimating input standard deviations to match output standard deviations.

Convergence continues to be an issue for most models and this should be examined.

Ways to reduce the number of parameters that may help address issues of convergence.

BSAI Pacific cod

There were a number of data changes and updates in this year's assessment that included; 1) catch data for 2004-2009 were updated, and preliminary catch data for 2010 were incorporated, 2) commercial fishery size composition data for 2009 and 2010 were updated, 3) age and mean length at age data from the 2009, size composition and numeric abundance information from the 2010 EBS shelf bottom trawl survey were incorporated, 4) seasonal catch per unit effort (CPUE) data for the trawl, longline, and pot fisheries from 2009 were updated, as was the 2010 preliminary catch.

The numeric abundance estimate from the 2010 EBS bottom trawl survey was up 24% from 2009. The IPHC survey 2009 estimate was down 35% from 2008 and was the second lowest point in the time series. The 2010 AI biomass estimate, used to compute the current ratio of BSAI biomass to EBS biomass, was down 26% from the 2006 estimate and was the low point of the time series. Applying a simple Kalman filter approach, adopted by the SSC in 2004, the current biomass distribution is 91% EBS and 9% AI compared to previous proportions of 84% and 16%, respectively.

All model fits to EBS survey abundance were good and produced similar estimates of EBS trawl survey selectivity at age, although the estimates from Model C appeared to be shifted by one year relative to Models A and B. Model A produced the most plausible lengths. Model C matched the modes very closely, but at ages that were higher by a year because the fitted growth schedule was unconstrained.

Model B is thought to have a better defined bin and season structure and was more parsimonious than model A. Model C was disqualified partly due to anomalous length-at-age in the EBS. The SSC agrees with author's and Plan Team's rationale, choice of Model B and Tier 3b designation for calculating the ABC and OFL recommendations, shown below in metric tons. The 2006 and 2008 year classes appear to be strong, and stock abundance is expected to increase substantially in the near term.

Stock/		2011		2012	
Assemblage	Area	OFL	ABC	OFL	ABC
Pacific cod	BSAI	272,000	235,000	329,000	281,000

GOA Pacific cod

There were a number of data changes and updates that included; 1) catch data for 2004-2009 were updated, and preliminary catch data for 2010 were incorporated, 2) commercial fishery size composition data for 2009 were updated, and preliminary size composition data from the 2010 commercial fisheries were incorporated, 3) age composition and mean-length-at-age data from the 2009 bottom trawl survey were incorporated into models A and B, 4) age composition and mean length at age data from the 2008 January-May longline fishery were removed from models B and C, 5) seasonal catch per unit effort (CPUE) data for the trawl, longline, and pot fisheries from 2009 were updated, and preliminary catch rates for the trawl, longline, and pot fisheries from 2010 were incorporated, and 6) size composition data from the State-managed Pacific cod fishery for 1997-2009 were updated and 2010 incorporated.

In terms of population numbers and biomass, a record high of 752,651 t was observed by the 2009 bottom trawl survey, when the population was estimated to include over 573 million fish. This followed the lowest observed survey biomass in 2007 of 233,310 t and a 2005 model estimate that was the low point at 140 million fish. The 2009 biomass estimate represented a 223% increase over the 2007 estimate.

All three models fit the GOA survey abundance time series relatively well throughout the time series, with the exception of 2009. In 2009 all model estimates were well below the highest survey abundance in the time series. Models A and B produced similar historical abundance time series; whereas Model C produced a very high historical abundance, implying that spawning biomass was five times B35% for the better part of the first decade. The latter was deemed implausible by the authors. There is little difference in fishery selectivity as estimated by all three models. In general, selectivities that are not forced to be asymptotic tend to show decreasing selectivity at large size.

Model A produces the best fit between observed and expected values for size at age, although the root-mean-squared-errors are about the same for all three models. Model B estimates for age 1 size appears to be about 2 cm high on average (which may be the result of the assumed aging bias) and Model C estimates an age 1 size that is very close to the observed average. Model B is thought to have a better defined bin and season structure and was more parsimonious than model A. Model C was disqualified partly due to impossibly high abundance estimates generated in the GOA model.

Based on Model B results, there is a slight decline in the estimated 2011 spawning biomass of 124,100 t, or 48% of unfished spawning biomass compared to the last assessment. Model B results also indicate a slight decline in subsequent years. This is in contrast to last year's assessment which projected an increase in biomass. Recent year classes (2006 – 2008) are also estimated to be substantially lower than in last year's assessment.

The SSC accepts the Plan Team's and the author's preferred model (Model B), Tier 3a designation, and the 2011/12 ABC and OFLs shown in metric tons below. The probability of the stock being below B20% was estimated to be less than 1% in 2011 and subsequent years.

Stock/		2011		2012	
Assemblage	Area	OFL	ABC	OFL	ABC
Pacific Cod	W		30,380		27,370
	C		53,816		48,484
	E		2,604		2,346
	Total	102,600	86,800	92,300	78,200

Excerpt from the minutes of the SSC (Feb., 2011; recommendation highlighted)

Discussion paper on BSAI Pacific cod split

The SSC received a staff presentation from Jon McCracken (NPFMC). Public testimony was provided by Dave Fraser (Adak Community Development Foundation), Frank Kelty (City of Unalaska), Jon Warrenchuk (Oceana), Kenny Down (Freezer Longliner Coalition), and Brent Paine (United Catcher Boats).

The paper discusses various approaches to sector allocation revisions, should cod BSAI ABC and TAC be separated into BS and AI. A substantial amount of uncertainty remains with respect to these action alternatives, especially in light of the 2010 SSL BiOp and RPAs. We have no empirical experience to understand fishing sector behavioral responses to the RPAs. As the author demonstrated, until these uncertainties can be clarified, it is difficult to arrive at a clear understanding of the "reasonably likely" outcomes that may emerge from each apportionment alternative identified in the paper. The SSC has previously expressed concern when reviewing the Draft RIR/IRFA supporting the 2010 SSL RPA action that conflicting expectations and assertions concerning cod fishing patterns and redeployment in response to recently proposed management actions (e.g., Amend. 90 RIR, 2010 SSL RIR) further confound analysis of impacts of AI and BS sector apportionment splits. The prospect of triggering another ESA consultation on AI Steller sea lions also adds to the difficulty in moving forward with this action.

It is noteworthy that recent cod biomass estimates indicate that the proportion of the combined BSAI biomass that AI represents is smaller than previously estimated (i.e., historical estimate >16%; new estimate ~9%). As AI cod allotments are reduced on the basis of the revised biomass, some sectors' shares may become inaccessible (e.g., NOAA may not be able to open a fishery due to limited TAC). This may have very significant implications for apportioning future AI cod fishing opportunities necessary to sustain patterns of historical dependency (e.g., catch distributions by area, operating mode, and gear type). The split of cod allocation between the BS and AI is likely to reduce the potential for localized depletion of AI cod by the BSAI cod fleet. However, the SSC notes that the potential still remains for localized depletion, given that a large portion of the fishable area may be closed under SSL closures.

The SSC recommends that the stock assessment author and Plan Team develop a plan of action for how the BSAI cod assessment should evolve. The possibilities include maintaining the status quo of a modeling approach in the BS and survey biomass in the AI, having separate models for the BS and AI, or having a single BSAI model (with or without geographic stratification and movement).

The discussion paper cites several aspects of a future AI cod sector apportionment action that may require the Council to revisit its original Problem Statement and 'purpose and need' rationale. Formal clarification of the Council's desire in regards to examining limits on EBS TACs, specifying area-specific allocations, and the disposition of latent permits are identified by the analyst. The interplay between the Federal AI cod fisheries and the State's parallel-waters AI fishery will also require Council examination and guidance, particularly in light of the most recent actions by the Alaska Board of Fisheries and ADF&G regarding SSL mitigation and several pending lawsuits challenging the 2010 BiOp and RPAs.

Depending on the Council's expectations for further analysis of this topic, revisions to this discussion paper could advance the development of the initial documents (e.g., RIR, IRFA) necessary to support formal Council action. If the discussion paper were revised, the SSC recommends expressly incorporating the recently announced State of Alaska AI cod management changes into the analytical baseline.

Appendix C:
Mark Maunder's model requests and commentary on the CIE reviews

Table of Contents

$p(\boldsymbol{\theta} \mid \mathbf{y}) = \begin{Bmatrix} \mathbf{Q} \\ \mathbf{R} \\ \mathbf{A} \end{Bmatrix} \begin{matrix} \textit{uantitative} \\ \textit{esource} \\ \textit{ssessment} \end{matrix}$
LLC

Quantitative Resource Assessment LLC
San Diego, CA
USA.

Bering Sea Pacific Cod Stock Assessment Model Scenarios Requested by FLC/QRA

Introduction

Requesting model scenarios for the Pacific cod stock assessments without knowledge of the assessment authors preferred model and alternatives greatly complicates the process. Therefore, we request that the assessment author use his best judgment when interpreting our requests and contacts us with any questions about the scenarios. We also have provided a commentary of the CIE reviewers' reports that the assessment author, Plan Team, and SSC, can use as a guide in creating their own scenarios or interpreting the scenarios that we request (see Appendix).

Scenarios Requested

<u>Scenario 1</u>

The stock assessment authors preferred model with the following changes

Estimate the bias and variance parameters of the aging error matrix inside the stock assessment model.

<u>Scenario 2</u>

The stock assessment authors preferred model with the following changes

Time blocks determined by initially modeling selectivity as a random walk.

<u>Scenario 3</u>

The stock assessment authors preferred model with the following changes

Use the bootstrap method to estimate the samples sizes, scale them so that the average is 300, then use iterative reweighting to update the samples sizes by fitting the Michaelis–Menton equation to the observed and effective samples sizes.

<u>Scenario 4</u>

The stock assessment authors preferred model with the following changes

Model time invariant survey selectivity, but model temporal changes in growth.

<u>Scenario 5</u>

The stock assessment authors preferred model with the following changes

Include the conditional age at length data and the length composition data, rather than the mean-size-at-age data, and estimate the variation in length at age inside the stock assessment model.

Appendix: Report on the Pacific cod CIE review

Summary

The three reviewers generally agree that the Pacific cod assessment is based on the best available science, but there a few areas that need improving through additional research and data collection. The reviewers did not provide any novel suggestions that would greatly improve the assessment or deal with the remaining issues.

The review process followed a set of questions outlined in the terms of reference. I present my summary below based on these questions. I have also added topics addressed by the reviewers that were not included in the terms of reference. My recommendations are provided at the bottom of each section in italics. I also summarize my recommendations that are relevant to choosing the 2011 model assumptions.

Assumptions for 2011 model

Further investigation is needed to determine the appropriate method to model and estimate the aging error and selectivity parameters.

Include the age composition data and the length composition data (or age conditioned on length and length composition) for all years if an appropriate aging error matrix can be generated, otherwise exclude the age data.

Include the conditional age at length data and the length composition data, rather than the mean-size-at-age data, and estimate the variation in length at age inside the stock assessment model.

Keep the current data partitioning.

Use dynamic binning for composition data

Eliminate the pre 1982 survey data.

Time blocks should be determined by initially modeling selectivity as a random walk.

Fix the catchability at the value estimated by Nichol et al. (2007).

Fix natural mortality at the value from Jensen’s (1996) equation.

Use the bootstrap method to estimate the samples sizes, scale them so that the average is 300, then use iterative reweighting to update the samples sizes by fitting the Michaelis–Menton equation to the observed and effective samples sizes.

Model time invariant survey selectivity, but model temporal changes in growth.

Fix the standard deviation of recruitment the annual residuals at 0.6 and test the sensitivity of management parameters to 0.4 and 0.8.

Terms of reference topics

Use of age composition data

The reviewers acknowledge that there is aging error/bias. They recommend including the age composition data in the assessment model in conjunction with an aging error matrix. They noted that excluding the aging data caused some undesirable model behavior. The reviewers also recommended continuing the research into the sources of the aging bias.

There was some concern that the age composition data used the same information as the length composition data so that the data was used twice. This needs to be clarified. However, double weighting of data is not too concerning since the weights are arbitrary in the current model. If the weights are "estimated" inside the model, then the issue of double weighting needs to be addressed.

Include the age composition data and the length composition data (or age conditioned on length and length composition) for all years if an appropriate aging error matrix can be generated (see below), otherwise exclude the age data.

Use of mean-size-at-age data

The reviewers recommend excluding the mean-size-at-age data, particularly if temporal variation in growth is not modeled. The mean-size-at-age data is the same data as used in the age composition and length composition data so the data sets are not independent.

Include the conditional age at length data and the length composition data rather than the mean-size-at-age data. This data provides information on variation of length at age, mean length at age, and temporal variation in mean length at age. The appropriate data to include also needs to consider the information required to estimate an aging error matrix.

Use of ageing bias as an estimated parameter

The reviewers did not agree on whether estimating the aging bias in the assessment model was the best approach. The models run during the review were not adequate to determine if the aging bias could be estimated appropriately. More research is needed on the form of the aging error and bias and whether it can be estimated within the stock assessment model.

The aging error comes from at least two source: 1) variability in reading the ages as indicated by double reading and 2) bias due to "false" rings being formed or the edge effect. An appropriate functional form for the aging error needs to be developed that can accommodate these two sources of error. We need to obtain the model files and investigate the appropriate method to model and estimate the aging error.

External estimation of between-individual variability in size at age

All three reviewers suggest estimating the variation of length at age outside the stock assessment model. This is partly due to undesirable model behavior when it was estimated inside the model.

The model does not include age conditioned on length data and therefore ignores some of the information available to estimate variation in length at age. Estimating variation in length at age outside the model does not take account of the aging error or selectivity. Variation in length at age should be estimated inside the model while including the age conditioned on length data. The development of a more appropriate growth model should also improve the models estimates of variation in length at age.

Catch data partitioned by year, season, and gear

The reviewers consider that the current catch data partitioning is appropriate. One reviewer noted that there is uncertainty in the catch estimates and this should be investigated.

Keep the current catch partitioning. Consider investigating a model that combines all catch into a single fishery for each season (it might be appropriate to reduce or eliminate the number of seasons) and use time varying selectivity for the fishery (the approach used by Ianelli for assessing pollock). The length composition data would need to be raised to the total catch within each fishery and summed across fisheries.

Size composition data partitioned by year, season, gear, and 1-cm size intervals

The reviewers consider that the current size composition data partitioning is appropriate. They recommended using dynamic binning to reduce the number of zeros in the likelihood function.

Keep the current size composition partitioning and use dynamic binning.

Age composition data partitioned by year, season, and gear

The reviewers consider that the current age composition data partitioning is appropriate. The reviewers were ambivalent about the use of the pre 1982 survey data because it probably does not influence the results.

Keep the current age composition partitioning and eliminate the pre 1982 survey data.

Functional form of the length-at-age relationship and estimating the parameters thereof

The reviewers noted the poor performance of the Richards growth curve due to the need to constrain one of the parameters to be positive.

A new growth curve needs to be developed for the Pacific cod assessment and implemented in Stock Synthesis.

Number and functional form of selectivity curves estimated, including assumptions regarding which selectivity curves should be forced to exhibit asymptotic behavior

The reviewers suggested that at least one selectivity curve should be asymptotic. They also suggested that a random walk should be used to model time varying selectivity to identify changes is selectivity and use this to determine where the time blocks should be applied.

The reviewers did not understand the types of selectivity curves available in Stock Synthesis. A selectivity curve can be created as a random walk over age (or length). This would allow a bimodal selectivity curve. The parameter for each age (the age offset) can also be modeled as a random walk over time, as can the parameters for functional forms.

A more robust approach is needed to model selectivity and determine which selectivity curves are asymptotic. Time blocks should be determined by initially modeling selectivity as a random walk.

Fixing the trawl survey catchability coefficient for the recent portion of the time series such that the average product of catchability and selectivity across the 60-81 cm size range equals the point estimate obtained by Nichol et al. (2007)

The reviewers recommended fixing the catchability at the value estimated by Nichol et al. (2007). They noted that when estimated, the estimate was similar to the Nichol et al. (2007) value. They also recommended collecting more tagging data to improve the estimate.

Fix the catchability at the value estimated by Nichol et al. (2007) and encourage further data collection.

Fixing the natural mortality rate at the value corresponding to Jensen's (1996) Equation 7

The reviewers recommended that the value of natural mortality continue to be fixed at the value from Jensen's (1996) equation. They also noted that it should be updated once the aging bias has been addressed and that age-specific natural mortality should be investigated.

Fix natural mortality at the value from Jensen's (1996) equation until the issues in the stock assessment model have been addressed, then estimate natural mortality within the stock assessment model and consider age specific natural mortality.

Input sample sizes for size composition and age composition data, and input log-scale standard deviations for survey abundance data

The reviewers recommended that the standard errors used for the survey index of abundance likelihood function should be reevaluated based on the survey design. The reviewers generally agreed with the bootstrap method used to estimate sample sizes, but noted that rescaling the averages to 300 caused the samples sizes to be lower than that suggested by the model fit to the composition data (effective sample sizes).

Use the bootstrap method to estimate the samples sizes, scale them so that the average is 300, then use iterative reweighting to update the samples sizes by fitting the Michaelis–Menton equation to the observed and effective samples sizes.

Allowing for annual variability in trawl survey selectivity

The reviewers questioned the need for annual variability in survey selectivity. However, they did recognize that catchability might change over time due to environmental factors such as bottom water temperature.

One reason for allowing the trawl survey selectivity to change over time is to accommodate changes in mean size at age for the young individuals. Temporal changes in catchability could also be due to abundance of different types of prey.

Model time invariant survey selectivity, but model temporal changes in growth.

Setting the input standard deviation of log-scale recruitment (σR) equal to the standard deviation of the estimated log-scale recruitment deviations

The reviewers were not conclusive about how to deal with the standard deviation of recruitment residuals. A value of 0.6 is supported by meta-analysis.

Fix the standard deviation of recruitment residuals at 0.6 and test the sensitivity of management parameters to 0.4 and 0.8.

Use of survey data and non-use of fishery CPUE data in model fitting

The reviewers recommended continuing to exclude the fishery CPUE data from the estimation of model parameters. One reviewer recommended excluding them completely because they might cause confusion. They recommended that the fishery CPUE data should be standardized.

Standardize the fishery CPUE indices and continue to include them in the assessment model, but not contributing to the estimation of parameters.

Other topics

Standardizing the survey

One reviewer suggested that the survey index of abundance be standardized for factors such as vessel, temperature, bottom type, location, and depth using a GLM or GAM. This reviewer also suggested mapping the habitat to improve the survey design.

Standardizing the survey for factors such as vessel, temperature, bottom type, location, and depth is a reasonable approach, but it might be better to post stratify by temperature, bottom type and depth each year rather than simply using a GLM. The habitat mapping could be used in this approach.

Jittering

Jittering the initial starting values of the estimated model parameters came up several times in the reviews. Jittering is a method to make sure that parameter estimates are the best values given the data and model assumptions. This is done because several years ago the model put forward had not converged properly and a better set of parameter values was found prior to the SSC meeting. The sensitivity of results to initial parameter values is probably caused by the selectivity curves. The need to jitter the starting values greatly increases the amount of time needed to do the assessment.

The model needs to be made more stable so it does not need jittering. This might be achieved by developing more robust selectivity curves.

Year to year changes in the model structure

The reviewers questioned the changes in model assumptions from year to year and suggested that the model structure should be fixed for a few years and the assessment only include updated data. In the years between "benchmark assessments" research could be carried out to improve the model.

Fixing the model structure for a few years is a reasonable approach to deal with several practical issues, but it would require the existence of a reasonable model. Unfortunately, and despite the substantial progress made on the Pacific cod assessment, there are still a few major issues that need to be resolved.

Tagging studies

The value of tagging studies came up several times in the reviewers' reports. The obvious need is to determine catchability for the survey using archival tags. However, well designed conventional tagging studies could be used to provide information on selectivity and natural mortality, validate aging, estimate abundance and exploitation rates, and evaluate stock structure.

A well designed and comprehensive tagging study is highly recommended.

<u>Alternative modeling environments</u>

The reviewers noted that alternative modeling environments might be useful to either customize model assumptions or double check model results. Developing a completely new customized assessment model for Pacific cod with all the functionality needed for sensitivity test would be a substantial task. It would be much better to request that the Stock Synthesis code be modified into a form that makes customization easy. Stock Synthesis can be configured to replicate either exactly or approximately many other stock assessment models and it would be better to apply simplifications of Stock Synthesis rather than using other models. The main reason to use another model is to identify programming errors in Stock Synthesis.

Request that Stock Synthesis becomes more user customizable.

<u>Overparameterization</u>

The reviewers mentioned several times that the models are over parameterized or nearly so. I doubt if this is correct. The issue is more likely related to poor model structure and parameterization (i.e. the selectivity curves).

The models are not over parameterized, but work needs to be carried out to make the model more stable.

<u>Aleutian Islands</u>

The reviewers suggest that the Aleutian Islands should be considered a separate stock.

I don't understand this issue well enough to make a recommendation.

<u>Management strategy evaluation</u>

The reviewers recommend continuing the management strategy evaluation work.

Management strategy evaluation (MSE) is very useful, but time consuming. Solving some of the issues in the assessment model are higher priority than the MSE work.

<u>Diagnostics</u>

The reviewers suggested several diagnostics that should be applied to the stock assessment including retrospective analysis, residual analysis, and evaluation of the correlation matrix to identify parameters that are highly correlated.

These are useful diagnostics and could be used to help select which model assumptions are appropriate. Retrospective analysis should not be used to determine the size or direction of the bias, only that some form of model misspecification exists.

<u>Dynamic reference points</u>

One reviewer noted that auto correlated recruitment may cause the abundance to drop below management reference levels even if the fishing mortality is relatively low.

Consider instituting dynamic reference points that take account of variation in recruitment