External Independent Peer Review by the Center for Independent Experts (CIE)

Assessment of the Pacific cod stocks in the Eastern Bering Sea and Aleutian Islands

By

Robin Cook

April 2016

Contents

Executive Summary

1. Three CIE reviewers conducted a peer review of Eastern Bering Sea and Aleutian Islands Pacific cod assessments during a meeting at the AFSC, Seattle from the 16th-19th February 2016.
2. The various models investigated to date lie on line where ending biomass (on a log scale) is inversely related to fishing mortality. The spread of model results is large for both stocks indicating that there is potentially a wide range of interpretation of the data.

<u>Eastern Bering Sea assessment</u>

3. Overall, the change from parametric descriptions of selectivity to random effects varying over time are likely to be important improvements to the Bering Sea model that reflect recent trends in stock assessment modelling and should be pursued. There may be advantages in modelling time varying parameters as a random walk to exploit "memory" in model.
4. Estimating age specific selectivity with annual variability for older fish is likely to be affected by large errors associated with small samples. I would suggest collapsing fish older than 10-12 into a plus group and setting the selectivity of the plus group to be equal the oldest true age.
5. While it may be possible to estimate fixed M and fixed Q given adequate contrast in the data, allowing Q to vary over time could mean that model mis-specification may arbitrarily emerge as variations in annual changes to Q. It may be better to fix M as an age dependent vector. Given that survey sampling protocols seek to minimise random changes, it is probably better to model annual changes in Q as a random walk to avoid over-fitting the data.
6. It is probably more useful to estimate Q within the model and regard it as a value that reconciles the assessment scale to the survey scale. Fixing Q within the model will add a degree of rigidity that may lead to severely biased estimates of fishing mortality, especially where the catch is treated as an exact constant.
7. A number of the exploratory runs performed at the meeting considered the inclusion of the IPHC longline, NMFS longline, and NMFS slope trawl survey in addition to the conventional use of the NMFS shelf trawl survey. Unlike the shelf survey, the three other surveys do not sample smaller fish and the length frequencies of their samples resembles more closely the commercial fishery length compositions. Trial runs for the EBS when these additional surveys were included tended to result in higher ending F and lower ending biomass so their inclusion is pertinent to management decisions. There is an issue about the appropriate weighting to give these surveys and this needs further exploration. Weighting the data by their estimated precision would seem appropriate.
8. An exploratory run that used Francis weighting for length and age compositions resulted in much lower estimates of F and much higher values of ending biomass. The sensitivity to an alternative weighting method is a cause for concern. An alternative error structure might be explored as a sensitivity test by fitting the model to the index of number-at-age in the survey which assumes that the observed number at age is lognormally distributed with age specific error distributions.

The Aleutian Islands assessment

9. This assessment differs from the Bering Sea assessment in that recorded catch biomass shows very large inter-annual changes and the trawl survey data series is not continuous. There are many missing years and there have been changes to the survey protocol over time. The base model used in exploratory runs (15.7) showed poor retrospective properties and a high sensitivity to the data used in the trawl survey.
10. At present it seems preferable to use parametric functions for the fishery and survey selectivity and to fix these over time. However, the very large range of outcomes from exploratory runs is a concern in the assessment and further analysis is required.
11. The very large annual variations in estimated F which may change by a factor of 2 or more in a single year, does not appear to produce a response in the stock biomass and implies recruitment is matching the change in biomass. This seems highly unlikely and suggests a problem with the assessment. Either the catches are less variable than they appear or the survey index does not adequately reflect real changes to stock biomass.
12. In the past a Tier 5 model has been used for the assessment of AI Pacific cod and in view of the problems with the Tier 3 model a simpler modelling approach is desirable as a fallback assessment. It might be possible to go further than simply smoothing the biomass indices by using a model similar to that outlined in Annex A, Table A2. Whether a simpler modelling approach is useful should be investigated if only to understand better which data contain useful information.

Other comments

13. Modifying the SS3 projection equation so that fishing mortality is explicitly modelled as the product of an age (or size) effect and a year effect offers scope for removing errors in the catches (provided these are treated as observations rather than constants), as well as exploring alternative models of fishing mortality (e.g. fixed effects, random walk) or even using fishing effort as a covariate.
14. Estimating the random effects in a conventional likelihood framework within the core ADMB requires external intervention to tune the analysis. Where complex random effects models are assumed, as in the case of some selectivity models, there remains some doubt as to the reliability of the estimates. An obvious way forward would be to redevelop Synthesis within the RE version of ADMB so that the random effects can be estimated within a conventional and tested mathematical framework.
15. The distinction between “Bayesian priors” and “penalty functions” as used by SS3 needs to be made clearer particularly where “uniform priors” are used. Where the latter represent bounds on the parameters, this should be made explicit especially where the converged model solution lies on a bound.
16. Where random effects are used, it was difficult to judge model performance in relation to goodness of fit versus the number of parameters. It would be desirable to calculate the Deviance Information Criterion (DIC) to try to overcome this problem so that different models can be compared.
17. It is worth investigating whether the assumption of normal distributions of length at age in SS3 is consistent with strictly incremental growth. It is possible that the model structure implies some fish much decline in length to be consistent with the size distribution model.

18. While Synthesis is an important and effective tool in the assessor's kit, it would be worth devoting effort to a more thorough investigation of the various data components before applying Synthesis as well as giving thought to alternative models that make different assumptions about errors in the data.

Background

The combined Pacific cod fisheries in the Eastern Bering Sea (EBS) and Aleutian Islands (AI) are economically important. Although there is a long history of assessments for these stocks in which a large number of alternative models have been explored, the annual assessments of the Pacific cod stocks in the EBS and AI continue to be controversial. Of particular concern currently is the estimation of catchability and selectivity for the bottom trawl survey in each area. An independent review of updated assessments of EBS and AI pacific cod was organised to assist in the development of final assessments that will go forward to inform the management process. The review considered preliminary assessments for each stock that comprised updates of models used in previous years and new models configured to address weaknesses or develop new approaches to the assessments. The review was intended to help guide further development of the assessment models and was not intended to peer review final models to be used for management advice.

Three CIE reviewers conducted the peer review during a meeting at the Alaska Fisheries Science Center, Seattle, Washington, from the 16th-19th February 2016.

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

Access to the pre-meeting review documents was available from the 2nd February, and comprised assessments of EBS and AI stocks as well as comments on earlier assessments from the SSC and BSAI Groundfish Plan Team. In addition, at the request of one of the reviewers, SS3 data and assessment files were made available on the Web. These documents were reviewed prior to the meeting. By way of data exploration, the reviewer also ran a number of models on some of the data and presented the results at the review meeting to inform discussion. These analyses are reported in Annex A. During the meeting, the reviewer participated actively in the discussion and suggested a number of exploratory runs to inform potential improvements in the assessment model configuration. Following the meeting, results of exploratory runs requested at the meeting but not available at the time were reviewed and used to prepare a final review report.

Summary of Findings for each ToR

The review meeting focused mainly on configurations of Stock Synthesis models (SS3) in which fleet selectivity and survey catchability were treated as random effects. With this configuration, model runs were undertaken to investigate the inclusion of the IPHC and NMFS longline surveys and the NMFS slope trawl survey. Other factors such as data weighting and recruitment assumptions were also considered. For convenience in this report the model results are summarized as a phase plot of ending biomass versus end year fishing mortality in Figures 1 and 2 for the Eastern Bering Sea (EBS) and Aleutian Islands (AI) respectively. As might be expected the models lie on line where ending biomass (on a log scale) is inversely related to fishing mortality. The spread of model results is large for both assessments indicating that there is potentially a wide range of interpretation of the data.

The following section on the Bering sea assessment discusses ToRs 1 and 2a-h. Given the nature of the current status of the Aleutian Islands assessment, the ToRs are dealt with in a more general context, and for AI no specific recommendations are made with respect to the configuration of SS3 models.

Bering Sea Assessments

In previous EBS assessments fishery selectivity has been broken down by fleet and season with parametric functions used to describe fleet and season selectivity. In many respects modelling at this scale can more realistically capture the true nature of the fishery, but is vulnerable to model mis-specification. The combination of fixed seasons and fairly rigid functional forms for selectivity may force a degree of inflexibility in the model, as well as a large number of parameters to estimate. These factors may have contributed to problematic retrospective patterns. At the review, the principal innovations proposed for a new SS3 configuration were:

1. Each year consisted of a single season instead of five.
2. A single fishery was defined instead of nine season-and-gear-specific fisheries.
3. Selectivity for both the fishery and survey was modeled using a random walk with respect to age (SS selectivity-at-age pattern #17) instead of the usual double normal.
4. Selectivity for both the fishery and the survey were allowed to vary annually.
5. Survey catchability was allowed to vary annually.
6. Initial abundances were estimated for the first ten age groups instead of the first three.
7. The natural mortality rate was estimated internally.
8. The base value of survey catchability was estimated internally.

Points 1-2 greatly simplify the model and avoid the need to estimate a large number of selectivity parameters. Point 3 is an important change that allows the data to determine the shape of fishery and survey selectivity. Allowing these to vary over time (point 4) then allows selectivity to evolve as the fishery changes. This is a natural way to accommodate changes to the activity and developments in different fleets that target different age and size components of the stock. Overall, these are likely to be important improvements to the model that reflect recent trends in stock assessment modelling and should be pursued.

While the use of random effects models has advantages in terms of the number of parameters to be estimated and model flexibility, there is danger of allowing too much flexibility. The annual changes to catchability (points 4 and 5) need to be considered carefully. My understanding of model 15.6, which formed the basis of most of the CIE requested runs, is that annual changes to selectivity and survey catchability were independent with respect to time. The danger of such an approach is that annual changes may simply reflect noise in the data rather than any true signal, because the model has no “memory” of what happened in the previous year. In many fisheries fleet behavior does not change substantially from year to year but evolves gradually over time. Thus, unless there is a “shock” to the system, one would expect selectivity in successive years to be correlated. Such correlation should be exploited in the model by, for example, modelling selectivity (sel) as:

$$sel(age,year)=sel(age, year-1)*exp(e_{year})$$

where e_{year} is a random innovation drawn from a normal distribution.

Using a random walk with respect to age to model selectivity is a fair enough assumption but is perhaps not strictly necessary if selectivity is modelled as a random walk over time. I would suggest estimating age dependent selectivities in the initial year as free parameters and then allow the base selectivity pattern to follow a random walk over time. Because the model remembers the previous year's selectivity, it means that all the data inform the estimates of fleet selectivity. This is an increasingly common assumption in current stock assessment models (e.g. Nielsen and Berg, 2014; Cook et al., 2015).

As regards the age range for estimating selectivity, it would seem desirable to consider only those older age groups that are adequately sampled. Unfortunately, misspecification of selectivity on older age groups that are poorly sampled can have a major effect on the assessment. If the estimates of fishing mortality obtained in the runs shown in Figure 1 are approximately correct, and if M is around 0.34, this implies values of total mortality in the region of 0.8. With such an exploitation rate, fish age 10 or older are likely to be very rare in samples. Estimating age specific selectivity with annual variability for older fish is therefore likely to be affected by sampling error. I would suggest collapsing fish older than 10-12 into a plus group and setting the selectivity of the plus group to be equal the oldest true age.

It is noteworthy that model 15.6 is configured to estimate both natural mortality (M) and base survey catchability (Q), which are typically highly correlated and difficult to estimate jointly. Estimating both these parameters and allowing survey catchability to change with little constraint over time seems imprudent. While it may be possible to estimate fixed M and fixed Q given adequate contrast in the data, allowing Q to vary over time demands a great deal of the data, especially when the assumption of a fixed M over time and age is clearly very unrealistic. It means that model mis-specification may arbitrarily emerge as variations in annual changes to Q. It may be better to fix M as an age dependent vector, determined for example by the Lorenzen relationship as shown in Annex A (Fig A.1), which used mean weight at age to estimate M. In this case, the estimated value of M over the mid-upper age classes is not very dissimilar to the value of 0.34 often used for this stock. Adopting an age or size dependent M value may be relevant to the assumptions used in these assessments for dome shaped selectivity. With M determined, it is then possible to estimate Q albeit conditioned on the assumed value of natural mortality. Given that survey sampling protocols seek to minimise random changes, it is probably better to model annual changes in Q as a random walk to avoid over-fitting the data.

There is much discussion in the Region of the value of survey catchability. There is clearly a desire for swept area estimates of Q to be seen to be close to values estimated from stock assessment models. It is undoubtedly of interest to compare such estimates and to try to understand the causes of any differences. However, there should be no surprise if such estimates differ and it is a mistake to force a given value of Q into the assessment since the assumptions on which the calculations are based are quite different. Where Q is estimated experimentally there is an assumption that the survey is sampling the same population as the fishery, and that the scaling factors used to raise trawl survey samples to absolute abundance are both accurate and unbiased. None of these assumptions is completely correct and there will be considerable uncertainty surrounding them. It is probably more useful to estimate Q within the model and regard it as a value that reconciles the assessment scale to the survey scale. Fixing Q within the model will add a degree of rigidity that may lead to severely biased estimates of fishing mortality, especially where the catch is treated as a known constant.

A number of the exploratory runs performed at the meeting considered the inclusion of the IPHC longline, NMFS longline and NMFS slope trawl survey in addition to the conventional use of the NMFS shelf trawl survey. It is generally considered good practice to include all the available data unless there are strong reasons to omit it. On that basis, all the surveys should be included. Unlike the shelf survey, the three other surveys do not sample smaller fish and the length frequencies of their samples resembles more closely the commercial fishery length compositions. It would appear therefore that these other surveys sample a part of the population that is not so well sampled by the shelf survey. Trial runs for the EBS when these additional surveys were included tended to result in higher ending F and lower ending biomass so their inclusion is pertinent to management decisions. There is an issue about the appropriate weighting to give these surveys. One run (BS_Model_15pt6_C_extraSD) weighted the indices by their respective standard deviation estimated internally, and this reduced the higher estimates of F seen in the other runs. Weighting the data by their estimated precision would seem appropriate and may prove the best way forward. There is, however, a somewhat different issue which is that these additional surveys all appear to sample a similar size range of the population and adding three similar surveys may bias the assessment toward the population seen by these surveys. Some exploration of this issue is required, but as mentioned, weighting by the precision of the data may be the appropriate solution.

A central feature of SS3 is that length and age compositions are fit as proportions rather than numbers at length or age. This requires an estimate of the effective sample size which is generally much lower than the actual number of fish sampled. In model runs carried out before and during the review sample sizes were constrained to be in the region of 300. An exploratory run (15.6_Francis, Figure 1) used an alternative weighting that resulted in much lower estimates of F and much higher values of ending biomass. The sensitivity to an alternative weighting method is a cause for concern though it should be noted that in the guidance notes for the use of Francis weighting there is a caution that “The large number of options available in SS makes it very difficult to be sure that what this function does is appropriate for all combinations of options”. There is no simple answer to this issue but something that might be explored as a sensitivity test would be to fit the model to the index of number-at-age in the survey as described in Annex A, Table A1. This assumes that the observed number at age is lognormally distributed with age specific error distributions. As an alternative error structure, it is not without its own problems (e.g. correlated errors), but if the data are rich in abundance information it would offer an insight into the robustness of the assessment.

Term of reference 2h requests advice on models that have apparently converged yet with a large gradient at the minimum. Where the parameter covariance matrix is calculable, there is some reassurance that a meaningful minimum has been reached. However, this is a technical problem and its resolution will depend on the algorithm used to minimize the negative log-likelihood. If SS3 offers a choice of minimization routines these could be explored to try to diagnose the problem.

Aleutian Islands Assessments

In principle much of the discussion relating to the EBS assessments should apply to the Aleutian Islands. However, there are at least two important differences that need to be considered. Firstly, the recorded catch biomass shows very large inter-annual changes that are apparently related to major changes in fishery management from year to year. Secondly, the trawl survey data series is not continuous. There are many missing years and there have been changes to the survey protocol over time. The latter point is of particular relevance since surveys are an important point of reference for an assessment and with missing or inadequately standardized data the use of random effects in the model may be unwise. The base model used in exploratory runs (15.7) showed poor retrospective properties and when the earliest two years of survey data were omitted, a radically different estimate of ending biomass and fishing mortality was observed (Figure 2, models 15.7 and 15.8), suggesting a high sensitivity to the trawl survey. While there was a significant change in the survey between 1994 and 1997 in terms of tow duration this does not seem sufficient to justify the removal of earlier years from the analysis. It also shortens the time series of available data substantially for an assessment already lacking in calibration data.

When selectivity was modelled with parametric functions fixed over time, the retrospective pattern improved which suggests a more rigid model may be better when calibration data are scarce or unreliable. At present, therefore, it seems preferable to use parametric functions for the fishery and survey and to fix these over time, but the very large range of outcomes shown in Figure 2 does not inspire confidence in the assessment and further analysis is required.

One issue that needs investigation is the significance of the very large changes in the annual catch biomass. Since the catch is treated by SS3 as a known parameter, the variability in the catches is translated directly into variability in estimates of fishing mortality because the assessment suggests stock biomass only shows very gradual change (see for example Annex A, Figure A3 that show F from model 15.7). Given the very large annual variations in estimated F which may change by a factor of 2 or more in a single year, one might expect to see a response in the stock biomass, but this is not apparent and implies recruitment is matching the change in biomass. This seems highly unlikely and suggests a problem with the assessment. Either the catches are less variable than they appear or the survey index does not adequately reflect real changes to stock biomass.

For the reasons above, I did not feel that the SS3 models were currently in a state to form the basis of an assessment through further model exploration, especially those using additional surveys may yet prove adequate.

In the past a Tier 5 model has been used for the assessment of AI Pacific cod. This model simply smooths the IPHC longline survey and trawl survey indices using a random walk. Such methods can be of use especially if the indices are a true reflection of the biomass trend, but inevitably offer little insight into the stock dynamics. It might be possible to go one step further using a model similar to that outlined in Annex A, Table A2. That model assumed that fishing mortality follows a random walk which may be too strong an assumption if management intervention has introduced a series of shocks to the fishery. The random walk assumption could be relaxed so that fishing mortality was modelled as a purely random effect with a large standard deviation.

Whether a simpler modelling approach is useful should be investigated if only to understand better which data contain useful information.

Other Comments

The population projection model within Synthesis appears to treat the observed catches as parameters rather than as observations. Where catches are very precise this approach may work well, but it is a strong assumption and for many stocks for which Synthesis is used the assumption of exact catches is hard to justify. In the case of EBS Pacific cod there are good reasons to suppose that recorded catches are precisely known for recent years, but historically this is probably not the case. Furthermore, little appears to be known about catches outside the US EEZ, which even in recent years may have an impact on the stock. A model that avoids the need to treat catches as known is therefore highly desirable. Modifying the projection equation so that fishing mortality is explicitly modelled as the product of an age (or size) effect and a year effect offers much more scope for removing errors in the catches by treating these as observations rather than constants. It also allows exploring alternative models of fishing mortality (e.g. fixed effects, random walk) or even using fishing effort as a covariate.

Until fairly recently, selectivity in the EBS assessment was modelled using parametric functions. Time varying selectivity was handled by dividing the time series into blocks where each block has its own selectivity values. In the models discussed at the review selectivity was sometimes modelled with random effects both over age and year. Estimating the random effects in a conventional likelihood framework within the core ADMB requires external intervention to tune the analysis. Where complex random effects models are assumed as in the case of some selectivity models there remains some doubt as to the reliability of the estimates. An obvious way forward would be to redevelop Synthesis within the RE version of ADMB so that the random effects can be estimated within a conventional and tested mathematical framework. This would avoid the need for *ad hoc* tuning and potentially would speed up the assessment process.

In configuring Synthesis some model parameters are constrained by penalty functions that are added to the likelihood and informally referred to as “priors”. These of course are not priors in a Bayesian sense and some care is required in their interpretation as a result. Some model parameters are described as having “uniform” priors implying no constraints on the parameters when in practice bounds are set to prevent estimates reaching values considered unrealistic and are therefore highly informative. This differs substantially from a Bayesian uniform prior where bounds are set primarily to avoid the MCMC chain sampling values outside the posterior distribution rather than setting limits on acceptable parameter values. Such a prior is uninformative and will give true unconstrained maximum likelihood estimates of the parameters. Given that the penalty function may be influential in parameter estimates in Synthesis much more thought needs to be given to the choice of these functions and more attention paid to their influence in the estimates especially where bounds are reached. Hitting a bound would tend to suggest insufficient information in the data to estimate the parameters.

SS3 allows a very large range of models to be fit to the data. This often means that the number of parameters being fit varies greatly as the assessment is developed. As is well known, more parameters usually mean a better fit to the data, but not necessarily a better model. In a likelihood approach model performance can be evaluated using the AIC which trades model fit against the number of parameters. In the review carried out at this meeting, it was difficult to compare models using AIC as the number of effective parameters is not clear in random effects models. As a result, it was difficult to judge model performance in relation to goodness of fit versus the number of parameters. It would be desirable to calculate the Deviance Information Criterion (DIC) to try to overcome this problem so that it is clear when a model is over-parameterised.

A particularly important feature of SS3 is that it can make use of both length and age data. The underlying model, however, is age-based and the population length composition is reconstructed from the dispersion around the mean length at age. This may well be an adequate assumption but one issue that perhaps merits investigation is whether this approach implies non-incremental growth since the assumption of strictly normal length distributions may not be compatible with the requirement that individual fish cannot get smaller as they age (except in exceptional circumstances). The question is whether, given a normal distribution at age a, the distribution of the same year class at age a+1 is simply a normal distribution centred on the mean length at age or some other distribution? Where length data drive the assessment inconsistency with incremental growth may lead to bias.

Conclusions and recommendations

For the Eastern Bering Sea SS3 models that use random effects to model selectivity by year, and perhaps age, are the preferred configurations at this stage of assessment development. Time varying parameters may be better modeled using a random walk to prevent over-fitting the data. The longline surveys and slope trawl survey should be included provided an appropriate way of weighting the data can be found. This needs to take into account the survey sampling precision as well as weighting relative to the shelf survey to avoid over-emphasis on the deeper water component of the stock. It is probably better to fix natural mortality externally if survey catchability is estimated internally, and especially if survey catchability is treated as a time varying parameter.

Much more model exploration is required for the Aleutian Islands assessment. It is especially desirable to try to explain on the basis of objective criteria why the historical catch shows such large inter-annual variability. The relative scarcity and lower reliability of fishery independent data to calibrate the assessment also makes the current Tier 3 models rather uncertain. While developing the Tier 3 model, consideration should also be given to enhancing the Tier 5 model to include a simple population model in order to obtain a little more information from the data as opposed to simply smoothing the time series.

SS3 is a well established and powerful tool that can be used both for data exploration and full assessments. A number of aspects of the tool deserve consideration for the future development of the model. This includes a more formal way of estimating random effects through ADMB RE, treating the catch data as observations as opposed to parameters and providing a statistic such as the DIC to compare best models when random effects are being used.

Stock Synthesis appears to be the only modelling tool considered when a full population dynamic model is fitted. While it is an important and effective tool in the assessor's kit, it would be worth devoting effort to a more thorough investigation of the various data components before applying Synthesis, as well as giving thought to alternative models that make different assumptions about the data. For example, given the major presence of length frequency data, a model that used a length based projection matrix might offer useful insights into the information contained in the data and treat growth in a more realistic fashion. Similarly models such as SAM (Nielsen and Berg, 2014), used by ICES, might provide a contrast to the multinomial assumption implemented in SS3. Truly Bayesian approaches that provided true estimates of the parameter posterior distributions may be more informative about the data than the application of penalty functions.

References

Cook, R., Holmes, S. and Fryer, R. 2015. Grey seal predation impairs recovery of an over-exploited fish stock. Journal of Applied Ecology 52, 969–979. doi: 10.1111/1365-2664.12439.

Nielsen, A. and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research 158: 96–10.

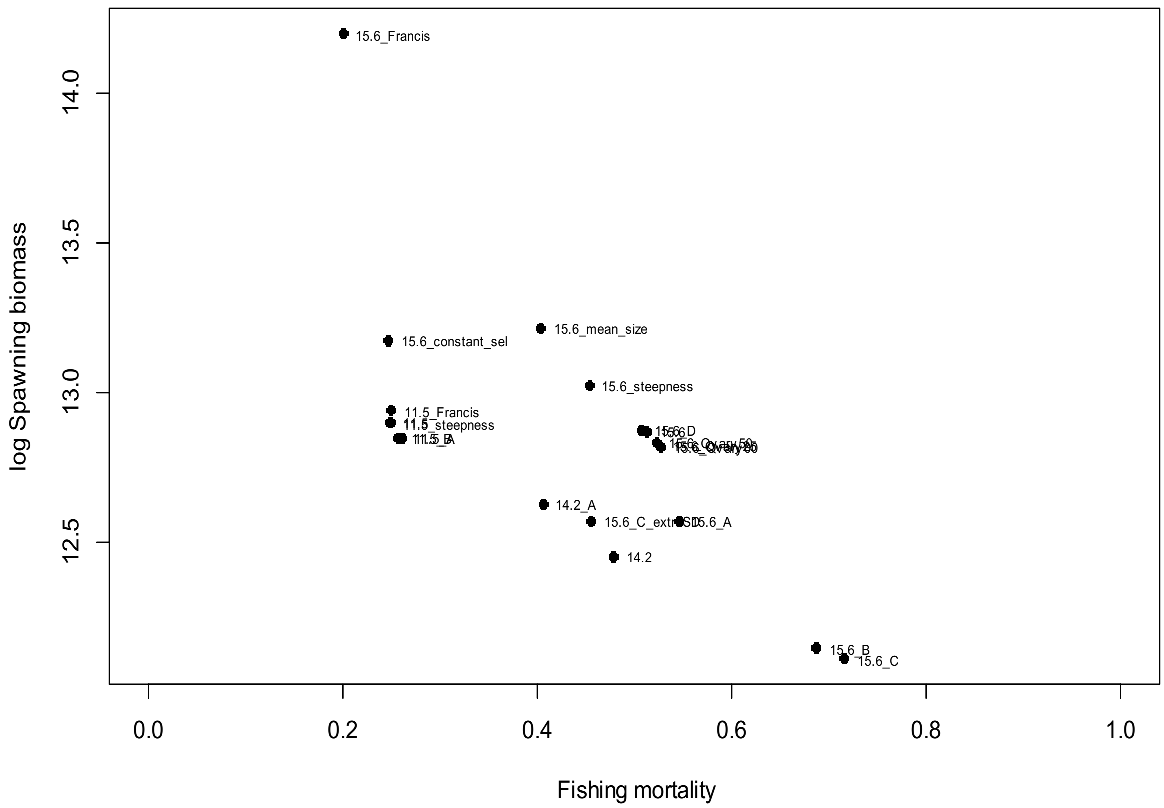


Figure 1. EBS model summary for end year biomass and fishing mortality

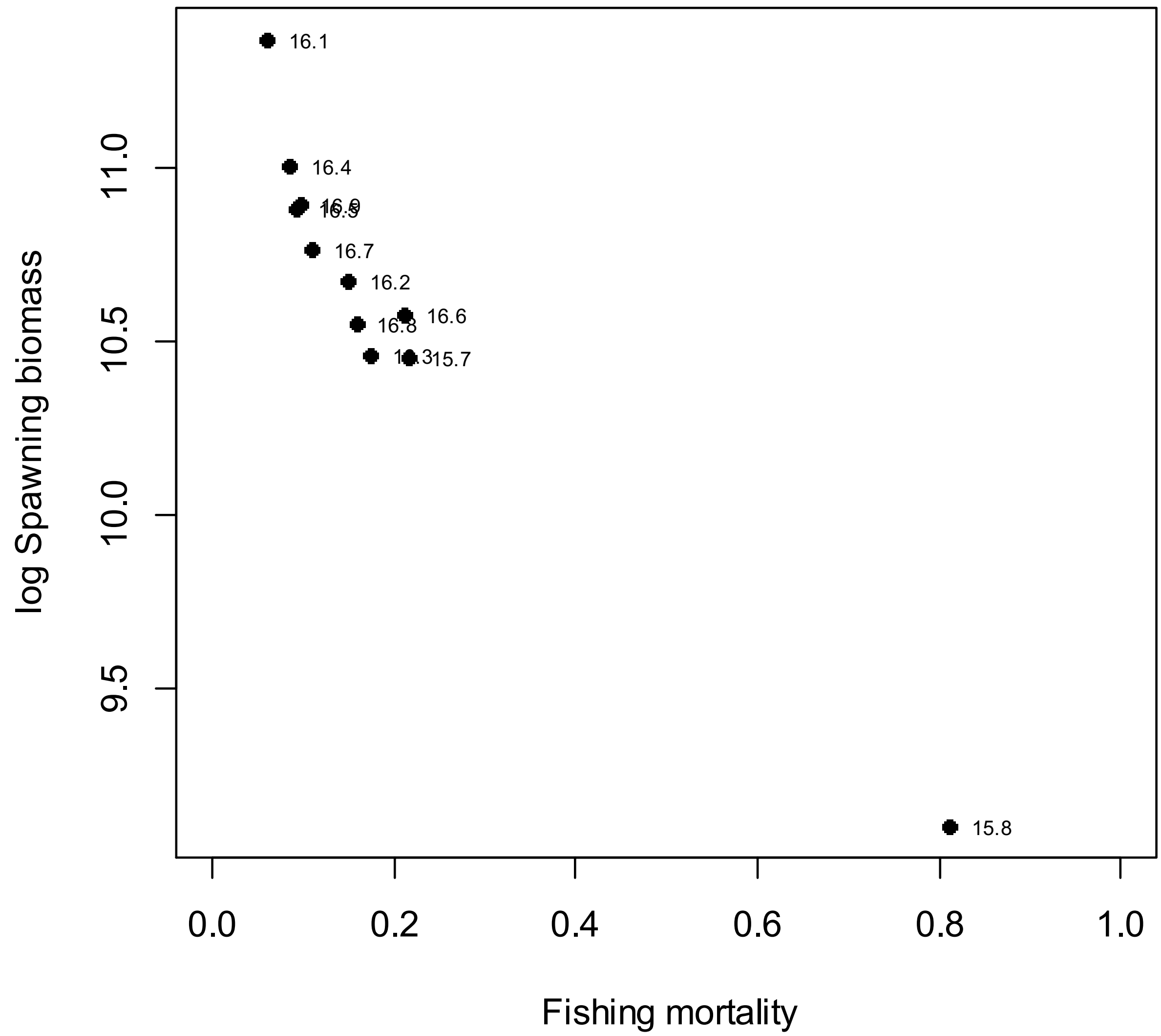


Figure 2. AI model summary for end year biomass and fishing mortality

Annex A. Exploratory analysis of EBS and AI data

An important assumption made by Synthesis is that fishing mortality is approximated by the observed catch and is treated as a known (error free) parameter. The population projection equations therefore use a variant of the Pope cohort approximation where the observed catch is subtracted from the current population. As a result, fishing mortality itself is not parameterised in the model, and is calculated after the model is fitted using the estimated log ratio of the numbers at age along a cohort. Many stock assessment models avoid this assumption by parameterising fishing mortality (F) and treat the catch as observed with error. This annex describes some exploratory analysis using alternative models that allow for errors in the catches and consider fishing mortality as a parameter and allow it to follow a random walk. In general, the models show similar trends to the Synthesis models but with much lower variability.

Eastern Bering Sea

This analysis considers only the total catch biomass and trawl survey catch at age data. The model is set out in Table A1. It treats the trawl survey as an index with unknown catchability (Q) and scales the population estimates to absolute quantities using the observed catch biomass. Fishing mortality at age follows a random walk with unknown process error. The model was coded and run in OpenBUGS. Two variants of the model were run; one with Q fixed over time and a second with Q allowed to vary as a random walk with unknown process error.
A summary of the results is given in Figure A1 where the output is compared to the preliminary Synthesis assessment models 11.5 and/or 14.2. In general, the trends in biomass, fishing mortality and recruitment from the Bayesian model are similar to the Synthesis model 14.2 giving some reassurance that the signal extracted from the models is a true reflection of the information in the data. It is worth noting that that the assumption in the Bayesian model of asymptotic selectivity for the survey is very different from the assumption in the Synthesis models of dome selectivity, yet similar stock trends emerge.

Aleutian Islands

A similar model was fit to the Aleutian Islands data, but at the time of the analysis no age structured data for the survey was available so the model treats all the individuals the same way regardless of size or age. The model is outlined in Table 2. Unlike the EBS model, recruitment is estimated as a random effect drawn from a lognormal distribution.

Fits to the catch and survey data are shown in Figure A2 and stock trends are shown in Figure A3. The model estimates the errors associated with the catch to be large and hence smooths the trend in catch considerably. However, the general trends in F and biomass are similar to Synthesis model 15.7, though heavily smoothed.

In these runs, the process error in F (σ_f) was set as 0.2. In principle it should be possible to estimate the process error within the model but this proved impossible. If σ_f was set to a larger value (e.g. =1), then estimates of F were extremely large and unrealistic. The problem is clearly associated with the very large inter-annual variation in the catch data which the model finds difficult to reconcile with the survey data that show a smooth decline over time. In the Synthesis model, all the variation in the catch is translated into estimates of F which show very large changes from year to year. In fact, the estimates of F are highly correlated with the catch observations. The implication if this is that very large changes in fishing mortality have little effect on annual changes in stock biomass. Given the very large estimates of F from Synthesis this deserves further investigation as it is unusual for F to vary so widely with relatively little effect on the biomass.

An alternative version of the mode was investigated where fishing mortality is not constrained by a random walk, i.e.:

$F_t \sim F_1 \exp\left(e_{f,t}\right), \;\; e_{f,t} \sim N(0, \sigma_f)$

This allows F to vary randomly around a constant mean. Results are shown in Figures A4 and A5. Here, the fit to the data is much closer (Figure A4) but fishing mortality, F, shows much more annual variation showing greater similarity to the Synthesis model.

Table A1: Population model used for EBS analysis

Population model

$N_{a+1,t+1} = N_{a,t}\exp\left(-F_{a,t} - M_{a,t}\right)$

$F_{a,t} = F_{a,t-1}\exp\left(e_{f,t}\right),\ \ e_{f,t}\sim N(0,\sigma_f)$

N=population number

F=Fishing mortality

M=Natural mortality (from Lorenzen Equation)

σ_f=process error of fishing mortality at age

Observation equations

$I_{a,t} = Qs_aN_{a,t}$

$\log(Q_t) = \log(Q_{t-1}) + e_{Q,t}\ ,\quad e_{Q,t}\sim N(0,\sigma_q)$

$I'_{a,t}\sim LogN\left(\log(I_{a,t}),\sigma_{a,I}\right)$

$$Y_t = \sum_{1}^{A} w_{a,t}\,F_{a,t}N_{a,t}(1 - \exp(-F_{a,t} - M_{a,t})/Z_{a,t}$$

$Y'_t\sim LogN(\log(Y_t)\,,\sigma_Y)$

I=Survey index (number at age a in year t)

Q= catchability of survey

s_a=survey selectivity

$w_{a,t}$= weight at age in the stock in year t

Y=catch biomass

$\sigma_{a,I}$=measurement error of trawl survey index

σ_q=process error of survey catchability

Data used

Trawl survey number at age

Total catch

Key assumptions

Natural mortality follows the Lorenzen equation where M is dependent on mean weight

Trawl selectivity is ***asymptotic*** with 50% retention at age 2.

Catchability Q is estimated in the model

Fishing mortality at age follows a random walk

Table A2. Population model used for Aleutian Islands

Population model

$N_{t+1} = N_t \exp(-F_t - M_t) + R_t$

$F_t \sim F_{t-1} \exp\left(e_{f,t}\right), \; e_{f,t} \sim N(0, \sigma_f)$

$R_t \sim LogN(r, \sigma_r)$

N=population number (survey scale)

R_t=Recruitment year t

r=mean recruitment (log scale)

F=Fishing mortality

M=Natural mortality (from Lorenzen Equation)

σ_f=process error of fishing mortality at age

σ_r=process error of recruitment

Observation equations

$Y_t = \overline{w_t} Q F_t N_t (1 - \exp(-F_t - M_t) / Z_t$

$N_t' \sim LogN(\log(N_t), \sigma_N)$

$Y_t' \sim logN(\log(Y_t)\,, \sigma_Y)$

Q= catchability of survey

$\overline{w_t}$=mean weight in the stock in year t

Y=catch biomass

σ_Y=measurement error on observed catch biomass

σ_N=measurement error on survey index number

Data used

Trawl survey index numbers and mean wt

Total catch in weight

Key assumptions

Natural mortality follows the Lorenzen equation: $3.64\overline{w_t}^{-0.304}$

Catchability Q is estimated in the model

Fishing mortality follows a random walk

Fishing mortality is scaled to 0.5 in initial year for identifiability

Figure A1. Results of fitting the model to EBS catch and survey data compared with models 11.5 and 14.2

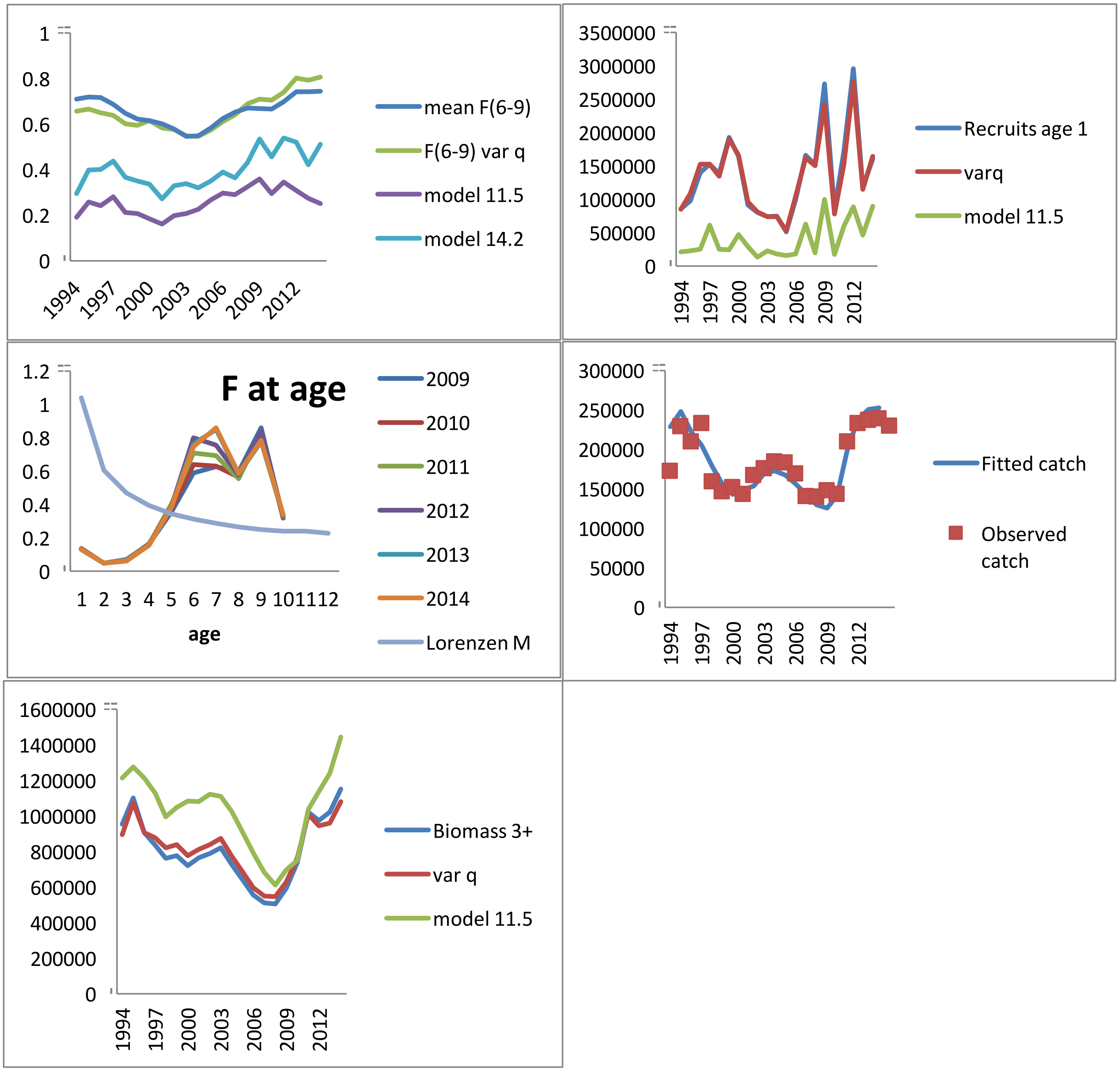
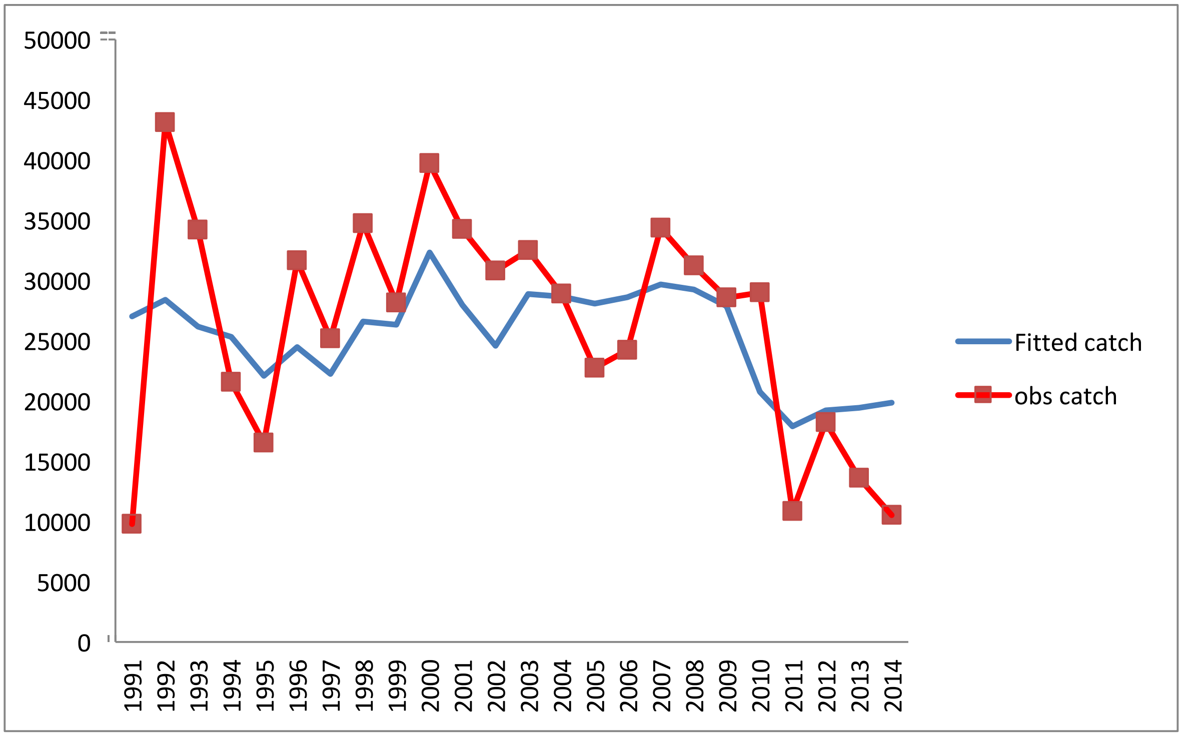


Figure A2. Model fits for the catch and survey index for the Aleutian Islands.



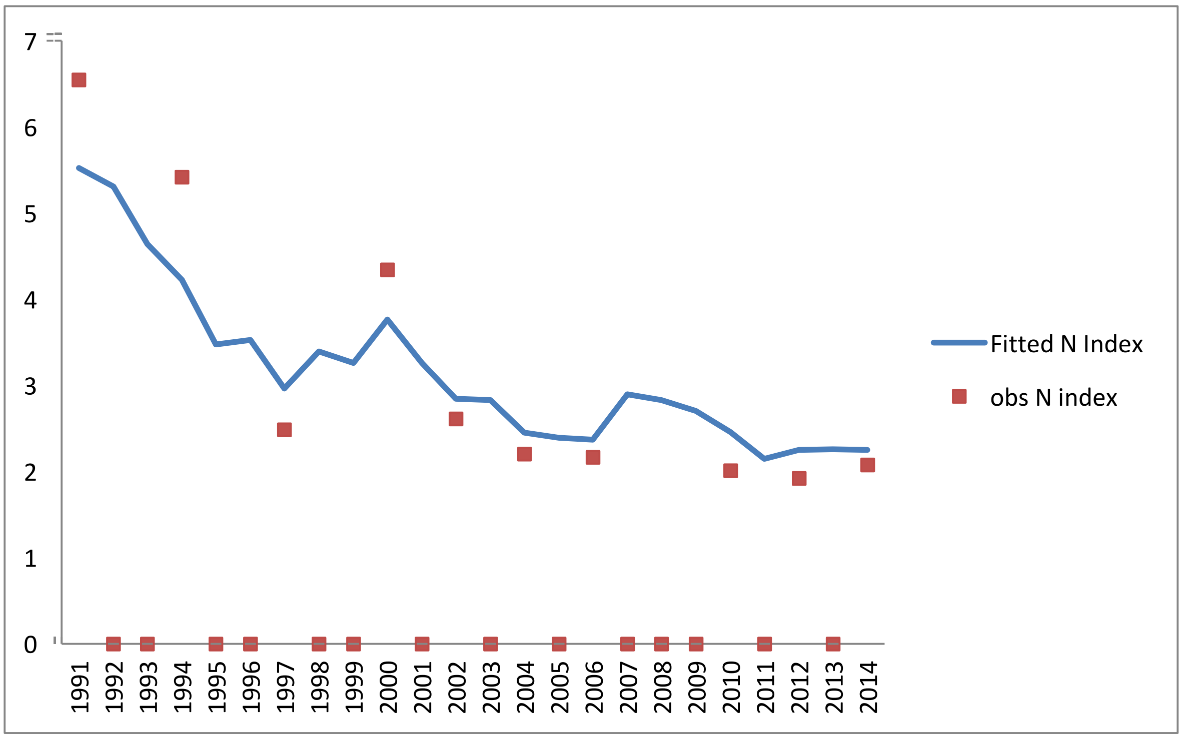


Figure A3. Model estimates compared to the Synthesis estimates from model 15.7. The time series are all normalised to their time series mean.

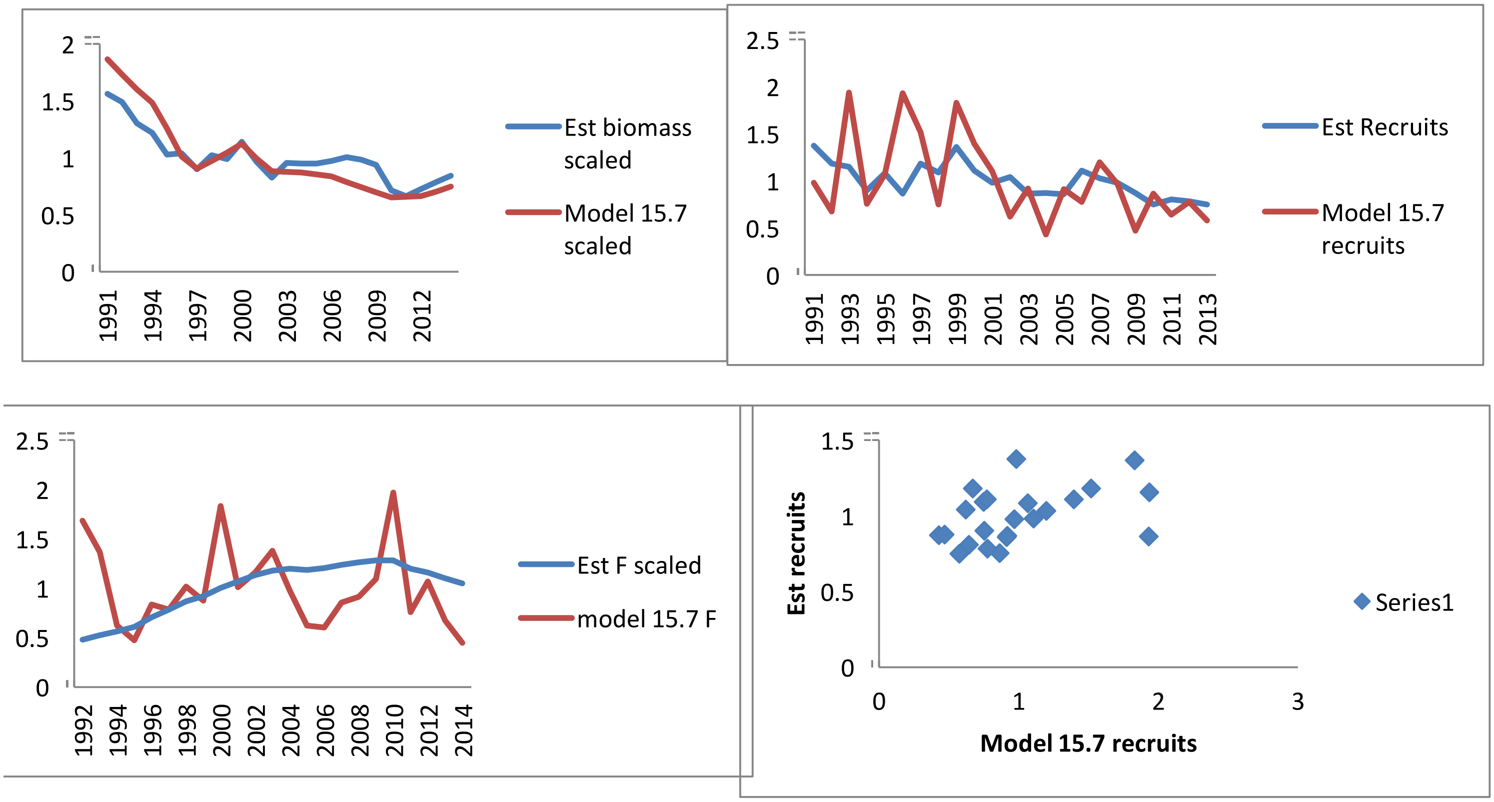
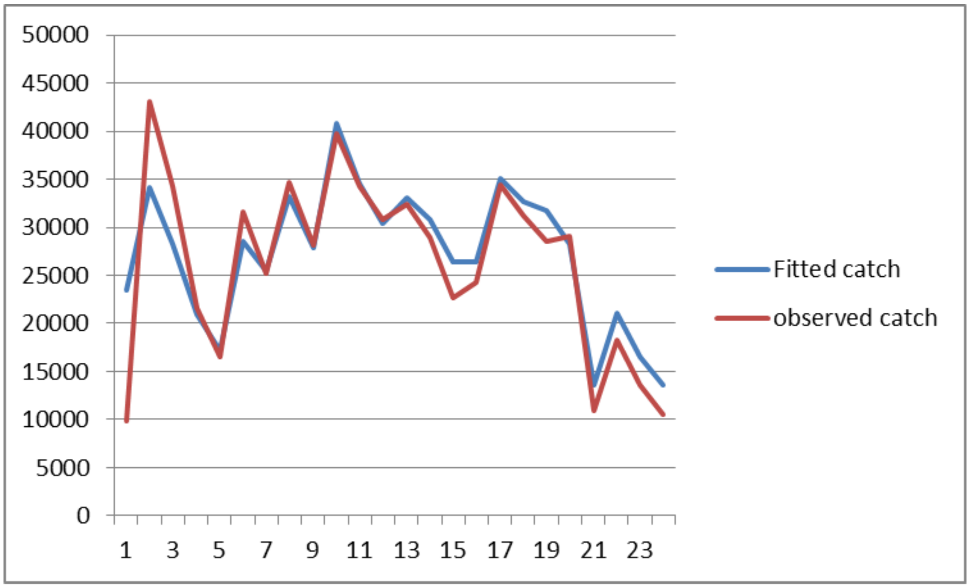


Figure A4. Aleutian Islands. Fits to the catch biomass and survey numbers when fishing mortaltiy is treated as a random effect with a standard deviation of 0.5.



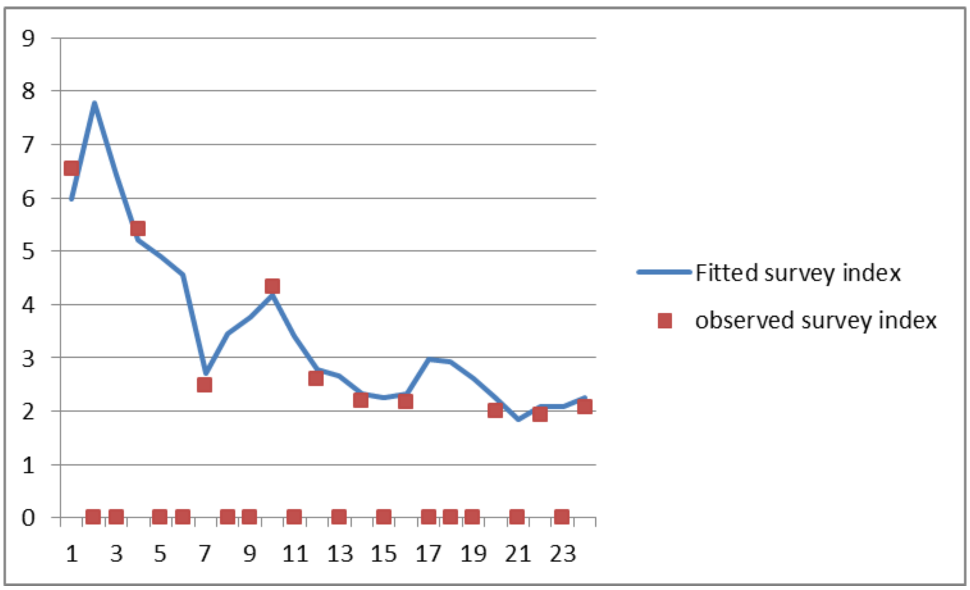
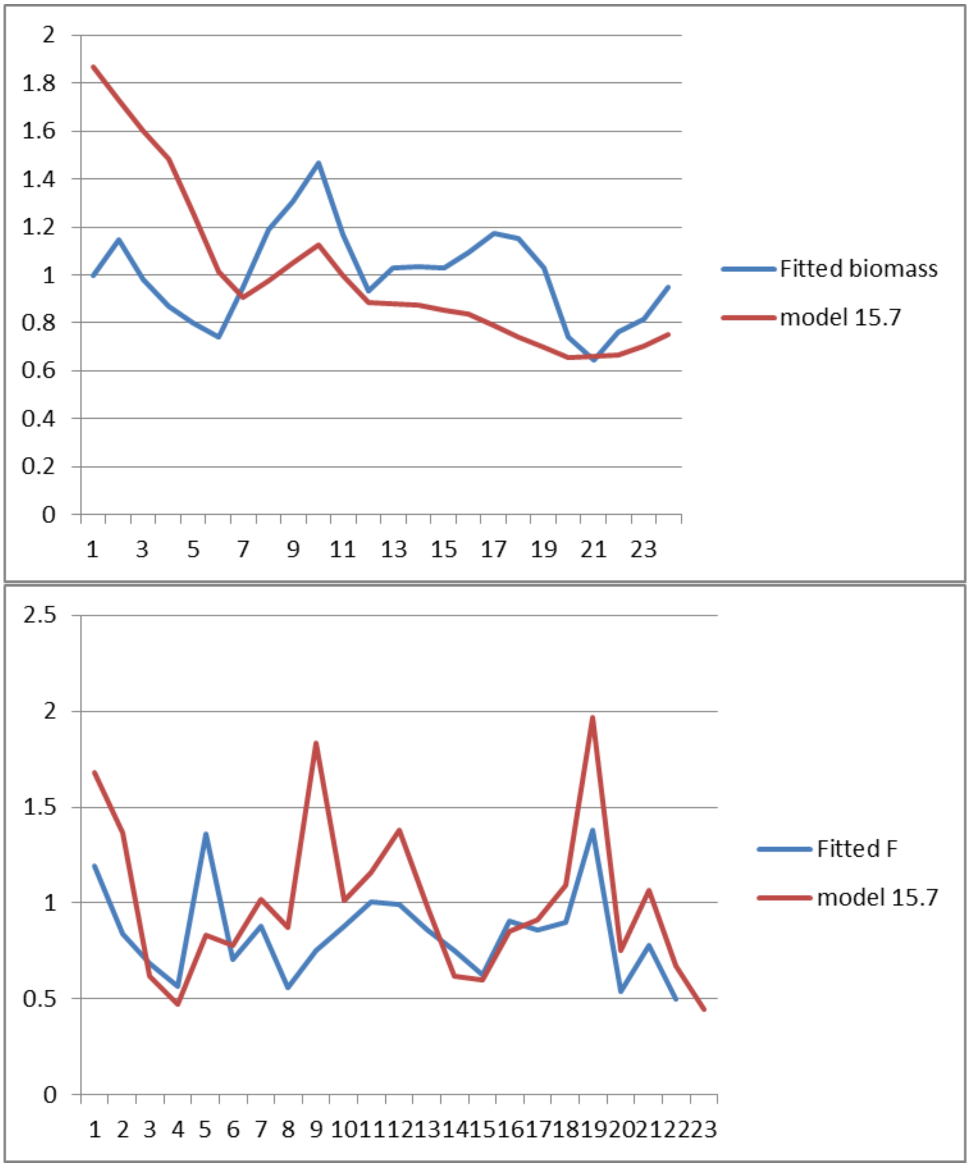


Figure A5. Aleutian Islands. Estimates of biomass and fishing mortality when F is estimated as a random effect with standard deviation of 0.5.



Annex 1: Bibilography

Copies of all documents for the review can be found at:

http://www.afsc.noaa.gov/REFM/Stocks/plan_team/2016pcodCIE/draft_assessments.htm

Pre-review background documents
Assessment of the Pacific cod stock in the Aleutian Islands
Assessment of the Pacific cod stock in the Eastern Bering Sea
Excerpt from the BSAI Groundfish Plan Team minutes of November 2015
Excerpt from the SSC minutes of December 2015
FLC comments - Pcod CIE review 2016
Thompson random effects variance estimation (version of 6-26-15, with figure legends)
Weinberg et al manuscript on field studies of selectivity (provisionally accepted)

Model files from 2015 assessments
ss3.zip
ss3tpl.zip
Preliminary assessment--Bering Sea.zip
Preliminary assessment--Aleutian Islands.zip
Final assessment--Bering Sea.zip
Final assessment--Aleutian Islands--Tier 5--Model 15pt6.zip
Final assessment--Aleutian Islands--Tier 5--Model 13pt4.zip
Final assessment--Aleutian Islands--Tier 3.zip

Presentations
Overview of data types used in the assessments (Thompson)
IPHC longline survey (Henry)
EBS trawl survey (Lauth)
Catch accounting system and in-season management (Furuness)
AI trawl survey (Palsson)
Observer program (Faunce)
NMFS longline survey (Hanselman)
Ageing (Helser)
Assessment history (Thompson)
Current assessments (Thompson)

Annex 2: Statement of Work

External Independent Peer Review by the Center for Independent Experts

Assessment of the Pacific cod stocks in the Eastern Bering Sea and Aleutian Islands

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: Despite exploration of a large number of alternative models and multiple levels of review each year, the annual assessments of the Pacific cod stocks in the EBS and AI continue to be controversial. Of particular concern currently is the estimation of catchability and selectivity for the bottom trawl survey in each area. However, review is requested of all aspects of the stock assessment models. The combined Pacific cod fisheries in the EBS and AI are of great economic importance, ranking second only to pollock in recent years. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment methods in general, and preferably Stock Synthesis in particular. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting ***scheduled in Seattle, WA during February 16-19, 2016.***

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Assessment of the Pacific cod stock in the eastern Bering Sea (220 p.), including a history of alternative models developed for assessing Pacific cod in the EBS (Appendix 2.3)
Assessment of the Pacific cod stock in the Aleutian Islands (143 p.), including a history of alternative models developed for assessing Pacific cod in the AI (Appendix 2A.3)
Comments on the final 2015 EBS and AI Pacific cod assessments by the Plan Team and SSC

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs cannot be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements).

The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

The review meeting will include three main parts: The first will consist of a series of presentations with follow-up questions and discussions by CIE reviewers, and will be chaired by an AFSC scientist or supervisor. The second will consist of real-time model runs and evaluations conducted in an informal workshop setting, and will be chaired jointly by the CIE reviewers. The third, time permitting, will consist of initial report writing by the CIE reviewers, with opportunity for additional questions of the assessment author.

<u>Contract Deliverables - Independent CIE Peer Review Reports</u>: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

<u>Other Tasks – Contribution to Summary Report</u>: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting ***scheduled at the Alaska Fisheries Science Center in Seattle, WA during February 16-19, 2016.***
3) Participate at the peer review meeting ***tentatively scheduled at the Alaska Fisheries Science Center in Seattle, WA during February 16-19, 2016*** as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than ***March 4, 2016***, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to mshivlani@ntvifederal.net, and CIE Regional Coordinator, via email to Dr. David Die ddie@rsmas.miami.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following ***tentative*** schedule.

January 11, 2016	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
February 1, 2016	NMFS Project Contact sends the CIE Reviewers the pre-review documents
February 16-19, 2016	Each reviewer participates and conducts an independent peer review during the panel review meeting
March 4, 2016	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
March 18, 2016	CIE submits CIE independent peer review reports to the COTR
March 25, 2016	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen Shimada@noaa.gov Phone: 301-427-8174

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:

NMFS Project Contact:

Grant Thompson, Alaska Fisheries Science Center
c/o Department of Fisheries and Wildlife
Oregon State University
104 Nash Hall, Corvallis, OR 97331
Phone: 541-737-9318
grant.thompson@noaa.gov

Annex 1: Format and Contents of CIE Independent Peer Review Report

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

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

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

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

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

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

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

3. The reviewer report shall include the following appendices:

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

Annex 2: Terms of Reference for the Peer Review

Assessment of the Pacific cod stocks in the Eastern Bering Sea and Aleutian Islands

1. Evaluate and provide recommendations on data used in the assessment models. In particular:
    a. Should data from the IPHC longline survey be used in either assessment?
    b. Should data from the NMFS longline survey be used in either assessment?
2. Evaluate and provide recommendations on model structure, assumptions, and estimation procedures. In particular:
    a. How should the various data sets be weighted?
    b. What form (i.e., Stock Synthesis "pattern") should be used for the selectivity functions?
    c. Should the models be structured with respect to season?
    d. Should the models be structured with respect to gear type?
    e. How much time variability should be allowed, and in which parameters?
    f. What constraints, if any, should be placed on survey selectivity at older ages?
    g. What constraints, if any, should be placed on survey catchability?
    h. How should large gradients be dealt with in otherwise apparently converged models?
    i. Anything else on which the reviewers care to comment.

Tentative Agenda

CIE Review of the EBS and AI Pacific cod stock assessment models

Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle, WA 98115
February 16-19, 2016
Building 4; Room 2039 (except Wednesday afternoon), Room 2143 (Wednesday afternoon)

Review panel chair: Anne Hollowed, Anne.Hollowed@noaa.gov
Senior assessment author: Grant Thompson, Grant.Thompson@noaa.gov
Security and check-in: Sandra Lowe, Sandra.Lowe@noaa.gov (206)526-4230

Sessions will run from 9 a.m. to 5 p.m. each day, with time for lunch and morning and afternoon breaks. Discussion will be open to everyone, with priority given to the panel and senior assessment author.

Tuesday, February 16
Preliminaries:
09:00 Introductions and adoption of agenda—Anne
Data sources (current and potential):
09:10 Overview of data types used in the assessments—Grant
09:20 Catch accounting system and in-season management—AKRO SF Division (via WebEx)
09:50 Observer program—AFSC FMA Division
10:20 Break
10:30 EBS trawl survey—AFSC RACE Division
11:00 AI trawl survey—AFSC RACE Division
11:30 IPHC longline survey—IPHC
12:00 Lunch
13:00 NMFS longline survey—AFSC Auke Bay Laboratory (via WebEx)
13:30 Age composition and mean-length-at-age data—AFSC REFM Division
Assessment models:
14:00 Assessment history—Grant
15:00 Break
15:10 Current assessments—Grant
16:10 Discussion—Everyone
16:40 Assignments for models to be presented on Wednesday—Panel
Wednesday, February 17 and Thursday, February 18
Review of models assigned the previous day—Grant
Discussion, real-time model runs—Everyone
Assignments for models to be presented the following day—Panel
Friday, February 19
Review of models assigned on Thursday—Grant
Discussion, real-time model runs—Everyone
Report writing (time permitting)—Panel

Annex 3. Participant list.

Review panel chair: Anne Hollowed, Anne.Hollowed@noaa.gov
Senior assessment author: Grant Thompson, Grant.Thompson@noaa.gov
Support assessment author: Steve Barbeaux, Steve.Barbeaux@noaa.gov

Name	Organization
Craig Faunce	AFSC
Craig Kastell	AFSC
Delsa Anderl	AFSC
Liz Chilton	AFSC
Ron Felthoven	AFSC
Sandi.Neidetcher	AFSC
Thomas Helser	AFSC
Tom Wilderbuer	AFSC
Farron Wallace	AFSC – Chair NPFMC SSC
Jeff Napp	AFSC – RACE Division Leader
Dana Hanselman	AFSC- Auke Bay Laboratory
Laura Junge	Alaska Pacific University
Anne Vanderhoeven	Bristol Bay Economic Development Corporation
Jean-Jacques Maguire	Center of Independent Experts (CIE)
Neil Klaer	CIE
Robin Cook	CIE
Chad See	Freezer Longline Coalition
Gerry Merrigan	Freezer Longline Coalition
Mike Peterson	Interested Fisherman
Patrick Lynch	NMFS Science and Technology
Mary Furuness	NMFS-Alaska Regional Office
Diana Stram	North Pacific Fishery Management Council (NPFMC)
Gwladys Lambert	NRC post-doc
Jennifer Cahalan	Pacific States Marine Fish Commission/AFSC
Steve Martell	SeaState