CIE Review for selected Gulf of Alaska and Bering Sea flatfish, 11–13 June 2012, Alaskan Fisheries Science Center, Seattle, Washington

Sven Kupschus

Prepared for

Center for Independent Experts (CIE)

Centre for Environment, Fisheries and Aquaculture Science
Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT
United Kingdom

Telephone: +44 1502 524454
e-mail: sven.kupschus@cefas.co.uk
www.cefas.co.uk

Contents

Executive Summary

Activities

The meeting was open to the public and was attended by a single observer; there was no industry representation. In general, the questions asked of the reviewers were related to assessment approaches and improvements to assessments and stock status determination rather than specific evaluation in terms of formal advice. For each stock, assessments were presented at the panel meeting in addition to assessments provided in written documents prior to the review, and this combination of material provided the basis of the discussions, with additional requests made by reviewers regarding presentation of input data and model diagnostics not contained in the documents. Assessment reports were well presented and, together with the presentations, formed a sound basis for the review. A more detailed presentation of some data summaries/analysis would have been desirable, as would a more detailed presentation of individual model diagnostics, including a detailed split of the likelihood components for comparison between different model settings, as presented for the rock sole complex.

In general all stocks examined at the review are currently lightly exploited, and there seems little pressure from either managers or industry to increase the level of catches. With the exception of yellowfin sole, the stocks have not been heavily exploited over the period of the assessments. This leads to a lack of contrast in either spawning stock biomass (SSB) or F, making it difficult to scale the management quantities or reference points accurately, particularly with respect to stock productivity. The difficulty of determining selectivities accurately when there is relatively little variation in interannual variation in year-class strength and relatively high estimates of ageing error suggests that uncertainty in fisheries selectivities and fishing mortality reference points is likely to remain high while the time-series of available information remain relatively short. The review panel generally felt that for Dover, rock and rex sole, the assessments can at best be used for the provision of formal advice at the level of Tier 3 or greater, but is almost certainly not suitable for tiers 1 or 2.

BSAI yellowfin sole suffers less from the aforementioned difficulties, having been exploited more heavily prior to the declaration of the exclusive economic zone (EEZ) and with the stock having recovered from low levels of abundance in response to greatly reduced fishing effort associated with the development of a domestic fishery. In addition, significant interannual variation in cohort strength deduced from the survey suggests that the assessment is able to provide advice to managers on a tier 1 level. Consequently, much of the discussion on yellowfin sole focused on the appropriateness of reference points for the stock and the effects of fisheries rationalisation, as well as the issue of the potential northward expansion of the stock into the northern Bering Sea.

1. BACKGROUND

In accordance with the Statement of Work (SOW: Appendix 2), I was contracted to participate as a CIE independent review panellist for the 2012 CIE Review for selected Gulf of Alaska and Bering Sea flatfish. This document represents my own findings and interpretation of the information provided, and is based on the panel meeting and discussions. However, some of the thoughts and conclusions were formulated in the process of writing this report, so may not have been discussed in specific detail at the review. Unlike some assessment review panels, the focus of the group convened for this task was more on the suitability of the assessments, possible further improvements to assessments, and the most suitable basis of advice, rather than specific recommendation of exploitation levels/OFLs for the coming year.

2. REVIEW ACTIVITIES

The 2012 selected flatfish review was held at the Alaska Fisheries Science Center (AFSC) in Seattle, Washington, from 11 to 13 June 2012. The bibliography consulted is listed in Appendix 1, and the Terms of Reference for the CIE panel in the SOW (Appendix 2: Annex 2).

A list of participants including panel members, Scientific and Statistical Committee (SSC) representation and observers are listed in Appendix 3. The meeting was open to the public, and was attended by one observer, but there was no industry representation. In general the questions asked of the reviewers were related to the assessment approaches and improvements to assessments and stock status determination rather than a specific evaluation of formal advice. For each stock, the assessments were presented at the panel in addition to the assessments provided in written documents prior to the review and this combination provided the basis of the discussions, with some additional requests made by reviewers regarding presentation of input data and model diagnostics not contained in those documents.

3. FINDINGS

Findings are provided by species and in response to each term of reference (TOR).

GOA rex sole TOR

CIE Reviewers shall evaluate, and make recommendations for improvements on, the current approach to determining stock status and future harvest reference points (ABC and OFL).

The review group was asked to comment on the proposal to continue providing information on the basis of the Tier 5 approach (recommendation by the assessment team) rather than the Tier 3 approach consistent with the available output from the age structured assessment model; the reason given for this was the uncertainty in the necessary F reference points required for the latter advice. Although this term of reference appears to be a relatively short and simple one to consider given an assumed assessment, it soon became apparent that the decision is almost certainly conditional on the quality of the assessment and the perceived accuracy and precision in the relevant management quantities, so required an in-depth look at the assessment. What follows are bullet points of assessment concerns and a subsequent summary of conclusions regarding assessment uncertainty, before answering the original question of the basis of advice. These points were discussed in some detail at the meeting but are shortened here for brevity because they were not part of the specific TOR.

1) Generally speaking, the quantity of data and the contrast in the information over time provide considerable challenges to the assessment in terms of convergence, which is why it is necessary to constrain the model considerably through assumptions. Results of the assessment are therefore highly dependent on the appropriateness of the assumptions.
2) The emphasis in the model on matching catch observations provides a good fit to the catch information in general. This is not surprising given the arbitrarily high weighting of 20 compared with most other data sources at 1. The residuals from the model, however, imply temporal bias in the catch figures, with the period 1993-2003 indicating consistent underestimation. Although catches are known with some certainty, weight-at-age is poorly estimated. Consequently, if weight at age is not allowed to vary over time, enforcing catches this strongly with an incorrect weight-at-age will incorrectly estimate year-class strength.
3) The fitting procedure for the steepness parameter of the selectivity function requires refinement. As parameterized, there is no effect on the model penalty function (flat log-likelihood) of steepness greater than 1, yet as implemented, AD Model Builder (ADMB) searches over this parameter space, potentially causing problems with convergence.
4) The length frequency information from the fishery is not well matched in the early part of the time-series, suggesting perhaps that a cohort has been missed. One possibility is that the relative proportion of the smaller vs. the larger fish is underestimated, because in the early years there is considerably more length information available than used, because some fish could not be sexed. This situation was worst in 1990, but continued to be a problem until 1996.
5) The survey information provided at the meeting (I believe this has not been corrected for the variable survey coverage, nor has 2011 been included) shows reasonable consistency in cohort strength estimation. The estimation of cohort strength is consistent between sexes (Figure 1). The plots of mean standardized abundance for both sexes suggest that 1990 appears to show some strong year-effects, with older ages being far less abundant than expected and younger ones dominating the catch. (This is one of the surveys that covered waters <500 m, but if this were the problem it should continue to be the case in the subsequent year. It is not, nor is it a problem with age sampling, because 270 otoliths were aged.) Despite this situation, cohort effects are quite clear, suggesting that after a peak in recruitment in 1988, recruitment dropped to a low in 1994 from whence it has recovered to near average levels in recent years. The assessment does not provide the same view of recruitment,

but smoothes the historical fluctuations while suggesting that recent recruitment has been generally twice the historical peak and about 4 times the historical low (Figure 6.15 in the assessment report). The 2007 assessment appeared to pick up these cohort trends more satisfactorily, although it is difficult to judge what the effect of the most recent survey data would have been.

6) The Markov Chain Monte Carlo (MCMC) posterior marginal distributions suggest significant asymmetry and departure of the maximum likelihood estimate from the median value of the posterior distribution for some selectivity parameters, suggesting that the joint posterior distribution may not be symmetrical. It may be that the joint probability is more informative than the marginals and that the width of the uncertainty in the selectivity slope parameter is in fact associated with otherwise highly unlikely outcomes, which are not appropriately penalized in the model (see 3 above).

Basis of advice

Given the above-mentioned uncertainties, my advice would be to provide information on the basis of Tier 5, but unlike the assessment team's consensus view, I would not base my estimate of spawning stock biomass (SSB) on the assessment, but rather on the survey or a form of survey-only analysis. This conclusion is made in part because I feel the assessment is smoothing through cohorts, indicating to me that it is acting more as a biomass dynamics model than a catch-at-age model. To estimate SSB on the basis of the survey only requires an estimate of survey selectivity, which it should be possible to derive from catch curves (the model suggests there is a great deal of information on 50% selectivity in the data but less information on steepness), and catchability is fixed at 1 (modified by the depth strata sampled as in the assessment).

I find the notion of using the biomass estimate from the assessment, but not the F/selectivity information, problematic. Separating the uncertainty in the biomass trends from the uncertainty in the F trends within an age-based assessment is in fact much more tenuous than is suggested by the MCMC marginal posterior distributions in Figure 6.18 (in the assessment report). Moreover, modes of the posterior distributions do not appear to match well with the maximum likelihood estimates. With survey catchability fixed, the problem is scaled and the only determinant that may lead to asymmetry in uncertainty between SSB-ratio (e.g. B_{40}) and the long-term F required to attain that SSB is uncertainty in selectivity.

Whatever that selectivity estimate currently is, it would modify the current estimate of SSB, and if indeed it is poorly defined, then estimates of the 2012 estimate of SSB from the future assessment will be equally variable. It is inconsistent to take one but not the other from the assessment, particularly because the model appears to be having a number of implementation concerns. For me, the only two options are to base the advice on Tier 5, but to do so on the basis of the survey information, or to provide information on the basis of the assessment in the form of Tier 3, in which case the model has to be implemented in a way that will lead to less uncertainty in the selectivity or at least show that the conditional uncertainty on the basis of the joint likelihood is smaller than the marginal likelihood.

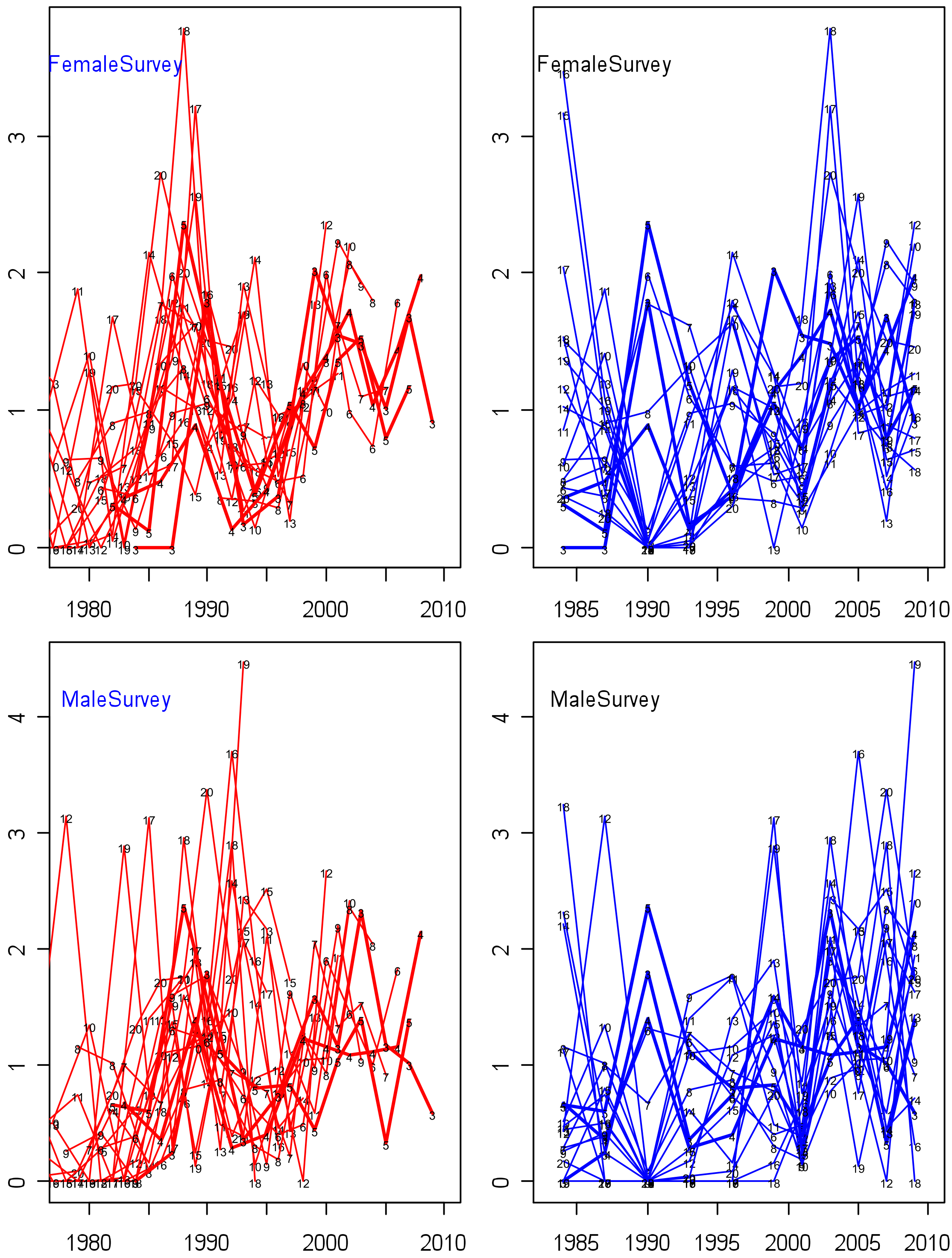

Figure 1: Mean standardized survey abundance indices for rex sole plotted by cohort (left) and year (right), with the first three ages (recruiting ages) shown as thicker lines.

Point of clarification

Median fishing mortality according to the text is estimated to be 0.015 per year (p. 10 in the assessment report, though table 6.13 suggests it is 0.067 [exp(−2.6945)]). The latter is also consistent with the marginal posterior (top right of Figure 6.14 in the assessment report). Presumably the posteriors of F_{40} (top right of Figure 6.18 in the assessment report) should be on the same scale (*x*-axes), but they appear to be on a different scale, because otherwise there are few if any scenarios in the MCMC that would make any sense from the perspective of population dynamics (F is far too high). In addition, the method of estimation of F_{40} within the model should be laid out clearly in the report. From the perspective of known population dynamics, the MCMC should not be able to explore negative Fs, which is why they are estimated on a log-scale (at least the median F value). In other words it should not be necessary (uncertainty) or mathematically possible (bias) to throw new fish in the sea to attain 40% of the virgin biomass.

GOA Dover sole TOR

CIE Reviewers shall evaluate the current model assumptions and make recommendations for improvements thereof, including: (see below)

In principle of course, the idea of an age-structured model is appropriate given the data available. The model is similar to the one used in evaluating rex sole (above), so the assessment suffers from some of the same problems. More importantly, though, the similarity in the data sources, such as a relatively short time-series with little contrast in F and only some indication of interannual cohort variation, seem to apply to this stock too. Relatively little information is contained in the data in terms of appropriate fisheries selectivity, so the assessments seems to rely heavily on the survey information when available. Maybe not too surprisingly, the difficulties in the assessment are similar and the approach to advice is the same. What differs in terms of the perspective of the review is the TORs. Many of the concerns/comments provided under the TORs for Dover sole also apply to the rex sole model, but may not be specifically mentioned there because the terms of reference did not request the information.

Use of age data, including:

use of age composition data

Every time I read the report on Dover or rex sole, I discover more about the model and its implementation. There are some unconventional implementations in the model assumed. Although clearly these have been present for some time, only examination of the code in ADMB can bring these out, something for which there is insufficient time to do in detail for each stock. However, since the implementation in MCMC, some of the aspects are becoming more apparent. Sadly, however, it is difficult to determine whether the source of the problem is implementation or whether it is the data/biology/fishery operation or process error that are giving rise to the difficulties in assessing the stocks.

Age reading for Dover sole is more difficult than for the other species, with not only a sizeable percentage disagreement, but also a much larger percentage error for individual fish. This may account to some degree for the relatively flat age composition information past the age of 10. Given that there is little in the way of a decline in the percentage occurrence of older ages, my concern is that even with a low M of 0.085 in the population, how is the model going to cope with essentially no fish dying. I was therefore surprised when examining the mean standardized indices from the survey (Figure 2) that there is reasonable internal consistency of the survey information with respect to the cohort signal, despite relatively low age-sampling levels, apparently poor reader agreement (though for younger fish highlighted in the plots this is less of a concern), and a strong year effect in 2001 (no eastern portion of the survey, corrected in the assessment, but not in the data used here).

The plots suggest that recent recruitment has been improving generally and that there has been at least one very large cohort recruiting in 2003 / 04. Survey data from the 1980s suggest that recruitment then was much lower, but more recent surveys paradoxically find plenty of fish from those cohorts, almost as though the catchability for the younger fish had changed. The survey age compositions (Table 5.6 in the assessment report), particularly for males, suggest that during the early part of the time-series the abundance of older fish was lower, being specifically evident for 1987, 1993, 1996 and to a lesser extent, 2001. These indices have not been corrected for the difference in the spatial extent of the survey and are consistent with the idea that the older fish are found in deeper water. The correction for the survey biomass as conducted in the assessment would then be appropriate. If one were to multiply the selectivities of the shallow surveys by 0.82 (to account for the difference in q) it would appear from Figure 5.12 (in the assessment report) that the survey selectivities are reasonably matched at older ages , but that they diverge around age 30 consistent with the difference in the age compositions in the survey. However, this would suggest that juveniles are evenly spread across the whole area. This does not appear to be the case, so a better way to model the stock would be to assume a constant q for all surveys (i.e. the whole of the juvenile population is monitored each year), and then to use a dome-shaped selectivity to estimate the effect of survey depth on older ages.

A further consideration with respect to the scaling of survey catchability to 1 is whether given evidence of low catchability (0.3) on the basis of gear selectivity studies (Figure 3), it is appropriate. From the latter information it is not clear how well defined this particular information is, but it certainly does not appear to be consistent with a catchability of 1. Freeing up this parameter in the model is likely to lead to strong correlation with other parameters, because of the lack of contrast in SSB and F over the time-series, so is likely to cause convergence problems. Nevertheless, the sensitivity to this assumption would benefit from being tested.

A more formal use of gear-efficiency data to estimate survey selectivity on the basis of length instead of estimation is difficult to support for two reasons. First, the spatial effects are likely to be significant for anything but the youngest ages, so additional spatial effects would have to be estimated separately, and these would then have to be correlated to the choice of gear selectivity curves so would not change the assessment output. Second, the size-at-age information does not appear to be sufficiently accurate / precise to do this.

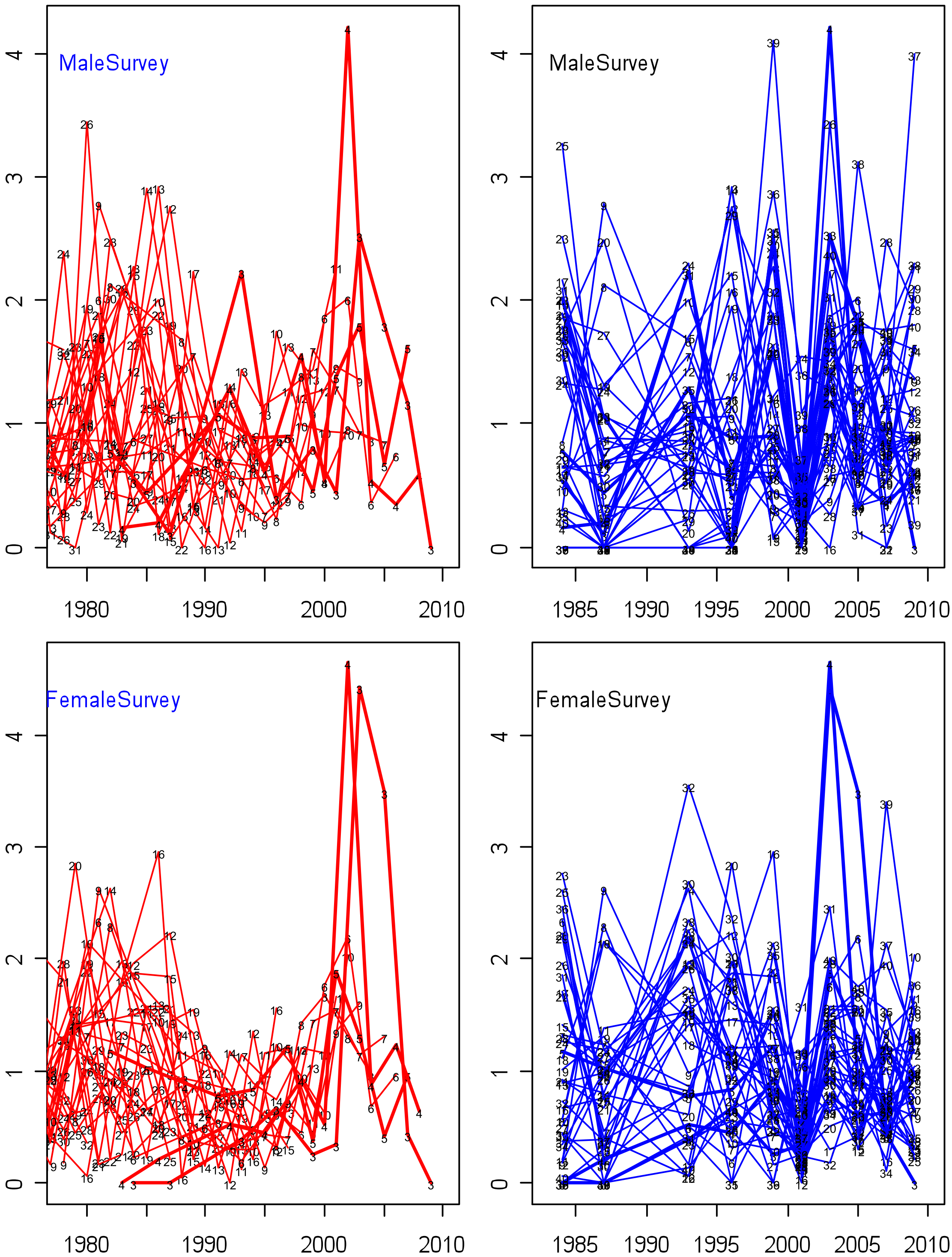

Figure 2: Mean standardized survey abundance indices for Dover sole plotted by cohort (left) and year (right) with the first three ages (recruiting ages) shown as thicker lines.

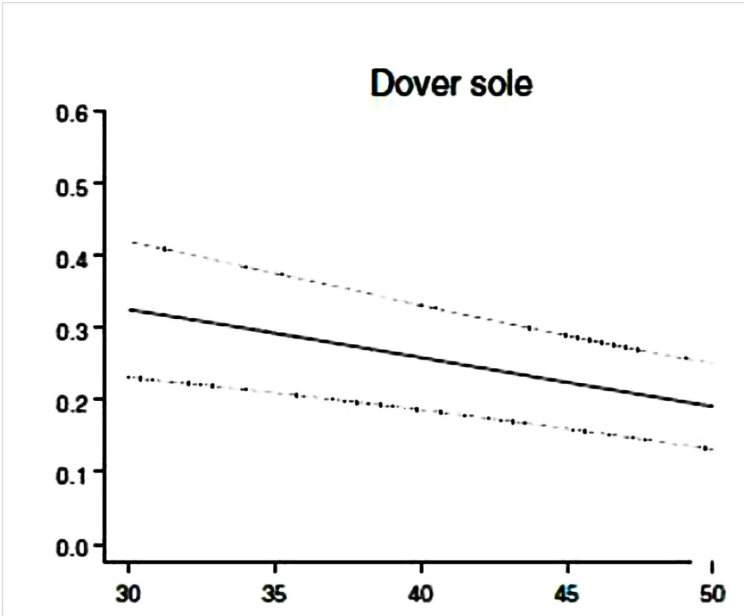

Figure 3: Estimate of gear selectivity of Dover sole suggesting a large percentage of the fish escape. Figure taken from Dover sole presentation provided at the Review meeting, with original citation unknown.

appropriateness of age range and binning

The size of the plus group seems large, indeed the largest group in the survey age compositions for many years, so should really be reduced, especially if dome-shaped selectivities were to be applied, to avoid cryptic biomass issues. However, given the relatively poor ageing repeatability, it seems that there are few fish at that age that can be assigned to a particular cohort with any certainty, so increasing the age range is almost certainly not a viable option. Decreasing the age range to the range where age reading is more certain may help in terms of tracking cohorts (less ageing error), but from an assessment perspective is unlikely to be helpful because the species matures late and is exploited at even greater ages.

estimation of size-at-age relationship and variability (external vs. internal to model)

I really do not see the source of the high-quality data needed to carry out an analysis of size-at-age. Even if information on weight was available with greater precision, the poor reader agreement on ages mitigates against this as an option.

inclusion of ageing error

The uncertainty in the assessment does include ageing error. What is probably meant here is formal inclusion. However, I fail to see how this would be possible. It is certainly possible to provide the model with a form of variance estimate in terms of ageing. However, the problem is the asymmetry of an ageing error. If an ageing error is found in assessments, it is usually only observable around very strong cohorts, and where adjacent cohorts are poor.

Basically, in early years it shows up as a single cohort, but later the cohort splits/spreads to adjacent cohorts. Unless there is a preconceived notion/information on the variability of cohort strength or its estimate (perhaps from surveys at recruitment age) models just cannot separate out his effect. Size information/growth curves can help in some circumstances, but they need to be derived from independent age information such as tagging information or when cohorts are separated well by size LFDA.

Generally

Use of size data, including:

use of survey size composition data

Given the relatively poor age determination and what appears to be extensive overlap of the size between ages, I am not sure that this adds much to the model. It would be worse if the size-at-age was changing over the time period, potentially reducing the effectiveness of the age data. The issue is really a matter of how much trust the authors have in each data source.

It should be remembered that although that the underlying stock dynamics are working on the basis of cohorts, when the conversion lengths to ages contains multiple confounding errors not specifically or appropriately addressed in the error structure, the use of length information will be detrimental to model accuracy and precision. Having looked at the problem only briefly, the model does not seem to fit the survey length compositions particularly well for the period 1984−1996, at least potentially missing a large cohort in the length data, which appears to be the only signal coming from that information (for both sexes).

use of fishery size composition data

The same comments as above would of course apply to the fishery length data, were it not for the fact that no age information is available for the fishery. These data seem very variable and it is difficult to observe any cohort signal, but of course it is necessary to determine selectivity in the model by some means. Whether this will lead to an accurate assessment of the selectivity function for the fishery is open to debate, and indeed there is some indication that this is poorly defined. I would argue therefore that it is necessary to include this information, but preferably along with better age information from the fishery.

The number and functional forms of estimated selectivity curves, including:

fitting different selectivity functions to data from different survey years based on survey depth coverage

(see section on age composition data)

types of selectivity curves considered

(see section on age composition data)

use of age-based vs. size-based selectivity curves

Size based selectivities are generally preferable when the major component of the variation in the catch is due to gear selectivity and not to spatial interaction with the gear. In this case, lengths are not very informative in terms of age given both the variation in size-at-age and the very small rate of growth of the ages in the fishery. The length−age conversion matrix (Table 5.18 in the assessment report) suggests that there is very little opportunity to distinguish cohorts in terms of their length, and in conjunction with the ageing difficulties leads me to believe that a length-based selectivity would on practical grounds be inferior even if it were desirable on theoretical grounds.

allowing for annual variability in fishery selectivity

Again in theory this would be a “nice to have” if it was affordable, especially because the fishery is not heavily targeted at the species and affected significantly by the regulations on other species. As I understand it, the effect here is less than that for rex sole, which is a true bycatch species. However, the model already has a great deal of difficulty in fitting selectivities (see section on convergence), so adding more of the same is likely to over parametrize it. That does not mean to say that it would not be interesting to try this option. It may indeed be informative in terms of the problems the model is currently facing with respect to fitting selectivities, but for management purposes it would be difficult to justify.

use of size-based selectivity curves for survey data based on trawl net catchability experiments

(see section on age composition data)

fixing (and updating) the natural mortality rate, based on Hoenig (1983).

The Hoenig method assumes that natural mortality decreases with age. The effect of changing the natural mortality rate in this assessment depends on the overall level of M to which the M-at-age is scaled. Generally, the idea is to scale this to the first age at full selection. For the Dover sole fishery, this would mean to ages >10. There is little change in size following that age, and there is very little maturity before age 10, so for such an assessment (assuming the current M is taken to be the M at age 10), there would be little impact on the SSB and F estimates, but the survey selectivity curve at the younger ages would change to account for the steeper decline in abundance of a cohort attributable to the higher assumed values of M. Given the available data, the difficulties in the assessment and the likely minimal impact on management, I would consider this to be a very low priority from an assessment perspective.

If a different form of scaling was chosen, one that alters the overall M, the productivity estimates of the stock and hence the management outcomes would change. However, this is likely to be because of the revised value of M, not the scaling of M with age. Unfortunately, to use the assessment (fit) to determine whether the revised value of M is suitable, scaled or not, it would be necessary to have considerable contrast in F and a highly reliable cohort structure, neither of which is available, and the choices would then have to be made more on theoretical grounds.

Model convergence diagnostics

Presumably this concern is based more on the previous assessment than this one, but of course the principle question remains the same. ADMB has some diagnostics that can help provide some information on convergence, such as the eigenvector file and simply observing the parameter estimation process on screen. Sadly, however, this information is not currently available in text file. It would be very informative to see which parameters are difficult to estimate, though (note, however, that a slow computer would be needed because otherwise the results flash by too quickly to be seen).

Other methods for obtaining a better feel for convergence would be to change the order in which the parameters are estimated. This is not always possible, though it can be helpful along with changes in the initial parameter settings and to see if the final estimates are sensitive to the starting values. Of course, it is possible to calculate manually the log-likelihood over a range of settings and to see whether it is possible to attain a lower value. However ADMB essentially does this and based on the response, determines new options to try. The MCMC function in ADMB does this essentially in reverse and this has been implemented in the model provided, so issues of lack of convergence are likely to be of less concern. Certainly the fact that the maximum likelihood estimates of the management measures now fall in the middle of the credibility interval suggest better convergence (Figure 5.18 in the assessment report). In contrast, the selectivity steepness parameter estimates for both the shallow survey and the fishery imply that the MLE is located towards the edge of the MCMC-searched space. The less than smooth marginal distributions in the $\alpha 50$ parameter for the shallow survey selectivity suggest that possible further minima may be located at higher ages.

The posterior on the slope parameter in the fishery selectivity is a further concern, only a very small fraction of the possible parameter space being investigated (see *x*-axes), strongly correlated parameters still trapping the MCMC algorithm in a local minimum because the response surface is flat and MCMC parameter selection based on the posterior distribution, which may not be searching a wide enough parameter space. This is certainly a parameter for which I would try to do a wider search. However, I think there are better ways to set up the model for survey selectivities (see above), which may make such a test obsolete.

In short all the right things are being looked at I terms of Dover sole, but this is still no guarantee of convergence; in the end, vigilance and experience will be the best way to avoid the pitfall of 2009.

GOA northern and southern rock sole TOR

Of the assessments reviewed at the meeting, the rock sole was the most complex to assess. It suffers from the same difficulties as the rex and Dover sole assessments of low data contrast, short time-series and the need to deal with the variation in survey coverage, but additionally has to deal with the problem of two species, which have only been separated later in the time-series, since 1996 (for most of the spatially constricted time-series) and 1997 (in the commercial information; a small percentage of the catch until 2009).

1. *Evaluation of findings and recommendations on quality of input data and methods used to process them for inclusion in the assessment (specifically fishery and survey data).*

The statistical catch-at-age model approach and the data and its usage seem to be fundamentally appropriate. Questionable, however, is the use of the external growth function from the literature, which appears to be inconsistent with the survey and catch data used in the model. This was said at the workshop not to be used in the fitting of the model for survey information, although a likelihood component is presented for all potential models in Table 4A.5 (in the assessment report). Moreover, the assessment report states that length rather than age compositions were used for the survey data until 1993 (presumably because species were not separated in survey catches). A clearer understanding is needed of how the survey data are used (tables of all the input information, especially the survey-catch-at-age in the assessment report is necessary). For the fishery data certainly, an alternative method is required to estimate size-at-age because the model fits to the data are poor for most years.

Other than this there are many ways theoretically that the assessment can potentially be improved, but the current model does not seem to replicate the trends in the stock dynamics appropriately, i.e. the residuals show temporal bias/trends rather than being random. Statistical catch-at-age models attempt to ensure that the deviations follow assumed error distributions, but make no assessment of the order of the residuals. The uncertainty estimates for both the MLE and the MCMC parameter estimates are sensitive to this assumption, i.e. they cannot distinguish between process and random error, but base the uncertainty and ultimately the parameter estimates on the total deviance, assuming it is random.

In this case the parameters are either hitting bounds, or there is little or no information to estimate them, with the result that the model more or less comes up with the same answers irrespective of the sometimes substantial changes to its configuration. This lack of response makes it difficult to decide which changes should be made to the model configuration, especially given that the model ends up being quite complex with the need to split the data into species and survey time-series. Starting with a much simpler model, then adding different options to see what the major contributing factors to the deviance are, would seem to be a more efficient approach in building a model. As it stands, I find it difficult to support the model's current implementation as the basis of advice beyond stating that the stock is lightly exploited and certainly exploited sustainably. The main indicator of this is the wide age structure (given the longevity of the species), and hence the breadth of fisheries selection (relative to L_{max}). Survey biomass estimates also indicate that current catches are not strongly linked to abundance. If stock status was poor, exploitation at sustainable levels should have resulted in a dramatic increase in abundance that is not at all apparent from fisheries-independent information. The assessment is consistent with those conditions, but the same conclusions could have been reached from simple examination of the data. For management purposes, the exploitation rate is the most important estimate following stock status estimation, and here I am not sufficiently confident in the estimation of the selection parameters. However, for Tier 5 this is less of a worry, but the survey selection parameters are seemingly equally uncertain and because of a lack of species-specific survey information in the early part of the time-series, it is not possible to develop absolute estimates of abundance from the survey information, unlike the case for rex sole.

The concerns with assessment estimates of concern are:

- The historical time series of recruitment estimates for southern rock sole are very low, which means that the biomass estimates in the model are virtually 0, and this in conjunction with an average catch leads to unreasonable estimates of F. The situation is less dramatic, but similar, for the northern stock, but because the southern stock makes up most of the biomass, this effect is less visible. There seems to be no variation in the southern stock either, suggesting that this parameter may not be estimated well, but it is going to be very influential in the estimates of selectivity.
- There is an offset of 1 year between the two species in terms of the recent strong recruitments over the period 1996–2007, a period over which the survey provides species-specific age information. This may be a case of increment formation taking place at different times of the year as a result of different spawning periods, or it may represent a difference in the time of the first annulus formation, but it does suggest that there are common dynamics in the species. If this situation can be confirmed independently in the survey information, i.e. that it is not an artifact of the size difference and the selectivity curve, the number of parameters needed would be reduced greatly. Certainly, the consistency of the sex ratio in the survey supports the idea.

2. *Recommendations for further assessment improvements for management in both the long and short term.*

As described above, although there are a number of things that could be improved in the assessment, it is currently unclear which factors are most important and how the model would respond to the improvements. The main reason for this is the complexity of the model necessary to describe the species split appropriately and the inability to distinguish ages in the early part of the time-series owing to a lack of age information and poor separation of ages by length. What follows below is therefore a recommendation for an approach, rather than specific things to change in the assessment.

I feel the model is too complex to evaluate the effects of individual changes given the information content in the data. It seems that any change made is countered by re-estimation of other parameters, and some parameters do not appear to be estimated at all (they deviate little from the initial estimates), suggesting that the model is over-parameterized.

What is important is to determine the process that has the greatest effect and that produces sensible and consistent results. An approach would be to start with a simple model, in this case perhaps combining the two species into a single species complex. Presumably at this stage it should be possible to use the undiscriminated age information from the survey to provide a better idea of the historical age structure, to use a single survey selectivity curve and to fix catchability at a reasonable level (e.g. 1). The model output at this stage would hopefully then indicate higher biomasses in the early period.

It should then be possible to run alternatives to this basic model, one model splitting the species, another freeing up catchability estimates of the survey, another adding additional selectivity periods, etc. The choice of different options should be based on detailed examination of the residuals. The location of systematic residuals vs. random ones will provide clues as to unrealistic process description within the model. It would then be necessary to choose the most appropriate model as the new base one and to try some further models each differing slightly from the base model. Increasing complexity slowly and understanding the effects of each change both in terms of the residuals and the population dynamics estimates will be important when it comes to determining the level of complexity sufficient to explain enough of the variation while avoiding over-parameterization. Small gains in precision (i.e. AIC or equivalent) are not necessarily justified. In cases where a lower AIC is attainable by the addition of additional parameters, this may be based on smaller, but systematic, residuals, either because the process or the error structure is inadequately described. Common sense needs to be employed when evaluating the appropriate model complexity, not strict statistical criteria, and some of the recommendations for the Rex and Dover sole assessments apply here too.

Comments on improvements to management are even more difficult to make, because they depend heavily on the outcome of model development. Using simple but effective indicators of fish stock dynamics as indicated in response point 1 of the rock sole TOR above are currently sufficient for managing the stock in the short term, but may represent difficulties in terms of the legal requirements for advice.

BSAI yellowfin sole TOR

Evaluation of the analytical approach (application of a statistical ADMB integrated catch-age model) and model assumptions used to assess stock *status and stock productivity.*

Compared with the other stocks reviewed here, the available data for this stock contain significantly more information on stock and fisheries dynamics, so a more-detailed catch-at-age model and an attempt to provide advice under Tier 1 is justified.

<u>The data</u>

The abundance-at-age information provided by the survey has good internal consistency in the estimates of year class, with sufficient interannual variation in recruitment to generate confidence in the estimation of fishing mortality trends over time. Moreover, the trends are consistent between the sexes. Although in the assessment, age composition and abundance information are treated separately in the penalty function of the model minimization, there is no reason to assume that the information on abundance-at-age derived from the survey is not being picked up by the assessment, as indicated by the good fit of the model to the age information and the SSB trend from the survey.

Catch data indicates significant variation in catch over the period of the assessment, with large catches early on during the exploitation by the international fleet, with more modest catches around a hundred thousand tonnes with some peaks to twice that during the development of the US fleet. The corresponding age information unfortunately is not provided in raw format. Catch-at-age is provided only as model predictions despite the reference to raw catch-at-age information on page 585 (in the assessment report). Model residuals do not provide any serious concerns regarding the models ability to interpret the catch-at-age composition though so one has to assume the data are reasonably consistent with the observed data.

Overall, discards from the fishery have been small in recent years, with just the Pacific cod and pollock fisheries (in 2010) having large discard ratios, and even their overall take is small compared with the targeted fishery. Consequently, the estimation of discards in recent years at least is relatively unproblematic with respect to the accuracy and precision of the assessment. In the 1980s and early 1990s, the estimates of the overall discard fraction are larger (up to 40%), and sampling levels seem to indicate that the estimates carry with them much greater levels of uncertainty, which it is not clear are represented in the overall assessment uncertainty. Given the recent low levels of exploitation and the heavy reliance of the assessment on the abundance information from the survey, it seems unlikely that this would have significant implications for the assessment of current F and SSB, but potentially could affect the estimation of some management reference points and consequently affect stock status estimation. It is not clear how the accuracy and precision in the discard rate estimate could be assessed objectively external to the model, so it may be necessary to estimate these inside the model (unlikely given the high degree of parametrization in the model for fishery selectivity and catchability) or alternatively to perform a sensitivity analysis on the assumptions of accurate discard estimation in the past. It is difficult to guess the likely influence of these effects on future management, but this is a minor issue given the light exploitation of the stock. Strictly speaking, however, this is a modelling issue and should be discussed further, but because of the likely low impact it is merely mentioned here.

Length-at-age information is not used directly in the model, but the average length- at-age information from the entire survey time-series is used to estimate weight-at-age in a number of different ways (see the model section below). Maturity information is based on a scientific analysis of survey information and is assumed to be constant throughout the time period. No information has been presented as to whether this assumption is or is not still appropriate, but it is assumed that maturity in the population is relatively stable compared with maturation in the individual, so it is unlikely that changes in the information would be discernable from the relatively few maturity samples that are available. Information on natural mortality was evaluated independent of the model based on additional data sources, but it appears to be consistent with the current assessment. From the perspective of fish life history, given the age, size and distribution, the value appears to be consistent with what one might expect for other species and stocks of this nature. If assessment estimates were to suggest that stock status was much closer to reference points, however, it would be sensible to provide a sensitivity analysis with respect to the assumptions of M. However, such an analysis is unlikely to yield significant differences in the provision of advice when there is little contrast in F in the recent period. The main effect would be a scaling up or down of SSB and the reference points.

<u>The model</u>

The model used in its most basic configuration is an age-structured statistical separable catch-at-age model implemented in ADMB. However, a sizeable number of “bells and whistles” allow for more-complicated, parameter-intensive relaxations of the basic model assumptions. These appear to have been implemented in the model in response to the assessment requirements specifically for this stock, although they do share many similarities with the model used to assess the pollock stock.

Given the available data and the apparent information content in terms of cohort strength information in the data, such a model is indeed appropriate for evaluating this stock in principle, and the model fit indicates in general that it should provide a sound basis for management. Because of the overall appropriateness of the model, the reasonable fit to the data and the lack of effect on the management quantities of some of the potentially questionable assumptions used in the assessment, this model is deemed to be suitable for the provision of advice. I do, however, think that some of the modelling and forecasting aspects are worth considering further and may be appropriate for refinement to advice. To do this effectively, detailed comments on these points are provided below, rather than going through each assumption in the model, and agreement with the authors on the appropriateness of other assumptions not mentioned is therefore implicit.

Assessment and catch forecast assumptions

Temperature effect on q

The effect of temperature on the survey efficiency was discussed in detail at the meeting. The assertion, and there is good evidence for it (Nichol, 1998 and others), is that the annual timing of the spawning migration is temperature-dependent, and because yellowfin sole spawn in shallow coastal areas outside the coverage of the survey, the proportion of the stock not sampled in a given year depends on the number of fish remaining on the spawning grounds at the time of the survey .This in turn depends on water temperature (in the model based on the average water temperature from the survey) in a given year. What is less clear is the quantification of this effect and the functional form of the relationship between q and temperature. Here it is modelled as an exponential multiplier on q, whereas the true form must inevitably be more complex. Moreover, q is applied to all ages including those not taking part in the spawning migration, which is a substantial part of the survey biomass. Given this strongly non-linear effect on the survey abundance estimates, a concern would be that this parameter ends up being a catch-all for residuals, reducing the information content to the assessment. This concern is exacerbated by the statement that “allowing q to be correlated with annual bottom temperature provides a better fit to the bottom trawl survey estimates” though having little effect on the estimation of management measures. No residual diagnostics are provided on this effect, but Figure 4.16 (in the assessment report) indicates significant temporal bias in the earliest and most recent periods with survey estimates, with the model under-predicting survey abundance, and most of this has nothing at all to do with temperature. Compared with the fixed q-model (fixed at 1.15) shown in Figure 4.18 (in the assessment report), the improvement is almost entirely attributable to five data points (1999, 2000, 2003, 2004, 2005) and at the cost of a reduced fit during the early 1990s.

The overall effect of scaling q is minor on the management quantities and is unlikely to change advice this year, so is inconsequential to the provision of advice this time. It does, however, provide potential problems for future assessments, if as suggested here, it allows the assessment to reduce the effect of the survey inappropriately if a large conflict between the survey index and the fisheries information were to arise for reasons other than the described migration effect. In any case, the stability and suitability of this parameter needs to be examined carefully in future assessments if it is going to be retained. A retrospective analysis of the development of the parameter would also represent a useful analysis to aid decision-making.

In general, I am in favour of including ecosystem and environmental effects in assessment models where supported, but including this temperature effect on catchability may reduce the chance of detecting other effects of temperature on the stock dynamics. For example, the growth chronology presented in Matta *et al*. (2010) appears to be a much better fit with temperature than the modelled effect of q, but shares a similarly strong increasing trend over the period of warming followed by cooling (1999−2006). If this effect on otolith growth is matched by a similar effect on weight-at-age, then this may be a much better explanation of the residuals in survey SSB than the catchability effect (see further considerations in the section on weight-at-age).

The consideration that the effect of temperature on q is related to escapement activity is poorly supported and to my understanding would compensate for the proposed migration effect. In any case, average mean bottom temperatures for the survey would appear to be inappropriate, and the survey would have to be corrected by tow to correct this properly. The effects of temperature on the “scope of activity” are generally minimal over the differences of temperatures observed here between the lowest and highest values, but of course this does not withstand the possibility that there may be some threshold level that could be interfering with it. In this case the implemented model is a poor proxy, so this theory is not considered further here.

Weight-at-age

The proposed assessment model uses weight-at-age fits from the survey to determine weight-at-age where it is available, or uses other empirical data sources or averages over certain periods. Underlying all of this is the assumption that weight-at-age varies. Unfortunately, the only way the model appears to be able to deal with imprecision in this estimate in the calculation of fisheries catch is by creating residuals in the overall catch. This is heavily constrained by virtue of the weighting in the penalty function in the model towards ensuring that the catch is close to the actually reported catch. This means that the numbers-at-age derived from the model are heavily affected when the weight-at-age estimates from the survey are not representative of the population or the catch.

Given the extended period over which the fishery operates relative to the survey, there is almost certainly an intraannual trend in weight-at-age, as a result of both growth and spawning. The effect cannot be investigated because comparative data sources do not exist. Although a single weight-at-age estimate would not serve to improve the situation, in combination with an estimate of the likely variability of growth around the weight-at-age taken from the standard error of the annual weights-at-age it may improve the need to tie the numbers-at-age in the catches so closely to the weight-at-age.

Alternative modelling approaches were suggested in the assessment fitting to age and year effects, whereby the latter were also tied to temperature effects. Although this approach may be appropriate for growth, it seems inappropriate for modelling weight, which is the cumulative of growth. The potential bias in this effect introduced when there are many ages in the population and annual growth represents only a small portion of the total weight near the asymptote. Accounting for cohort, age and year will likely result in strong colinearity between the effects, rendering parameter estimation difficult. It would, however, be possible to fit weight-at-age in the model conditional on the available data by deriving year (or temperature) and age effects on growth, then summing growth to determine weight-at-age. Including this in the model-fitting procedure could also avoid the effect of the weight-at-age on catch numbers described above. Personally, I find it difficult to accept growth model 2 or 3 (described in the assessment report) in the assessment, particularly given the strong constraint on the precision of the catch estimates.

Stock−recruit relationship

Given the low level of exploitation of this stock, there seems little question that this stock (as for the others reviewed here) is currently not subject to overfishing. Furthermore, it is very unlikely that the stock is in an overfished state, but to be sure of this, a reference point is required, because this is a target fishing mortality for providing advice on future TACs. In this Tier 1 assessment, it is proposed that this be achieved formally through the investigation of the likely stock−recruitment relationship following the determination of recruitment deviants and estimation of SSB (see Ricker spawner−recruit relationship).

Three different time periods are examined, the complete time-series, an early one and a late one, referred to as A, C, and B, respectively. Plotting the full time-series produces a convincing stock−recruit relationship, despite the fact that for flatfish generally such fits are rare, because recruitment is frequently driven mainly by environmental conditions and the effects of spawners are much less discernable. Aside for some possible autocorrelation in the recruitment, this then represents a plausible estimate of B_{msy}, and with a given selection pattern allows for the determination of F_{msy}. The historical time-series only (model C) provides much the same picture, generally suggesting that the relationship has not changed over time. However, the recent time-series estimates B_{msy} to be significantly lower (model B), and the assessment team felt that given the documented regime shift in 1978 (hence the split of the time-series), it is uncertain whether the historically good recruitments were not caused by environmental effects rather than low stock size, and consequently opted for the more conservative approach.

I have some fundamental issues with this approach:

1) F_{msy} is an equilibrium concept. Long-lived stocks with sizeable recruitment variation and possible temporal autocorrelation in recruitment will never reach such an equilibrium state, so I am not sure that the measure is of use here without testing with stochastic simulations that the population is sustainable even over periods of prolonged poor recruitment caused by sporadic or periodic recruitment.
2) F_{msy} is based around the idea of exploiting populations when the population growth rate is maximum for the prevailing environmental conditions. That is why there may be justification for shortening the time-series. However, there is an evolutionary consideration that population growth rates greater than those necessary to reach equilibrium population size in excess of the carrying capacity only make sense if there is a need for buffering against changing environmental conditions. By taking the portion of the population in excess of that at which population growth peaks, we significantly increase the risk to the population, because under some environmental conditions, what appears to be currently maximum exploitation is suddenly no longer sufficient to

maintain the population. Setting different values of F_{msy} for different environmental periods is therefore technically unacceptable. Although in reality it could be concluded that one should always take the most conservative approach, in reality periodic environmental conditions are unlikely to have varied sufficiently in any fisheries assessment to have any notion of the additional risk added by this approach. In addition, much less may be taken from the fisheries than possible if one moves away from the equilibrium concept and determines the current rate of population growth from the assessment directly, i.e. allowing for variable exploitation rates rather than a single rate. There is likely to be either a single level of F_{msy}, or no F_{msy}. The concept of periodic F_{msy} values is neither estimatable (because of the short time-series) nor internally consistent with the theory underlying the idea in the first place.

3) If despite the previous points, Tier 1 reference points are desirable, conservationists would require the most conservative one to be taken or that a determination be made of the risk in the uncertainty around a changing reference point. The more conservative reference point is surpassed during the more recent time period, but if we look at the variation in biomass included in the short time-series, it is clear that it provides little or no data on stock productivity at lower biomasses and can therefore not be informative on appropriate levels of F_{msy}, and the risk in foregoing considerable yield would appear to be large. The fact that it is a lack of information at lower stock levels that is causing recent B_{msy} to be much lower, rather than a specific change in the productivity in the stock, is underscored by the fact that the full and early time-series are similar, whereas if there had been evidence of a change in productivity, then the early and the full time-series would have differed. The risk of collapsing the stock by fishing at the higher level of F_{msy} in this case is very low, because even if there was no compensatory mechanism in recruitment, it would take many years to fish down the stock to the more conservative level of B_{msy} (i.e. 400 million tonnes) even at the less conservative level of F_{msy}, at which point it should be clear which level of F_{msy} ($F= 0.11$ or $F = 0.16$) is more appropriate given the current conditions. The same would be true of fishing at the more conservative F_{msy} level, except that the period taken to approach B_{msy} would be considerably longer and, assuming some compensatory effect, would take a long time to approach equilibrium at B_{msy} (400 million t), at the same time foregoing a vast potential yield.

I therefore suggest that using the full time-series is sufficiently conservative not to endanger the stock or the fishery while having the least risk of unnecessarily reducing potential yield. In terms of immediate management, however, there seems little pressure from the fishery or indeed scope for increasing yields given the restrictions on capacity and effort based on other more-limiting resources which are taken in association with the fishery. It appears that some progress has been made in rationalizing the BSAI fisheries, which may in future allow for more effective exploitation of the stock.

Catchability >1

A catchability of 1 is often considered problematic, which is why there is usually a rush to explain why it might arise. The authors of the assessment explain the small deviation from 1 (0.15) for an area-based survey on the basis of herding, citing a number of papers that refer to it. No doubt there is some herding by the sweeps and bridles, but it is doubtful whether this is the only reason why the average q is >1. There are many possible reasons for this, a likely candidate being an inaccurate measure of M. In fact, if one freely estimates q for almost any assessment that used a swept-area-based approach for a survey, one would almost always end up with a value of q different from 1. Assessments are working are almost certainly working on a relative level and the accuracy on an absolute scale is much less than the precision on a relative scale. Providing the conclusion that herding is the main reason for this value of catchability is unlikely, which is why at first reading, it distracted from the feeling of quality about the assessment. Personally, I do not think that it needs mentioning, let alone justifying, but traditions in the US may be different.

Evaluation of the implications of using the northern Bering Sea research results as an index of abundance if yellowfin sole increasingly occupy this area with changing climate

The question in general is of course an interesting one, but how to deal with spatial changes in the distribution of a species outside the current survey area is perplexing. In general, surveys should be of a design capable of adapting to changing conditions while maintaining a handle on catchability. The rex sole survey dealt with the changes in depth distribution of the survey in a pragmatic and useful manner, but the situation here is different because it is unknown whether there has always been a population in the northern Bering Sea or if the population expanded there with the recent period of global warming. Before discussing how one might handle such additional information, it is important to consider whether the northern Bering Sea stock represents an entirely separate population or at least a subpopulation with significantly restricted mixing from the stock currently being surveyed.

It is of course difficult to determine the above with any certainty, given that only a single year of data is available. In addition, there is no fishery operating in the area that might provide a longer time-series of information. Nevertheless a simple plot of abundance by station spatially rather than interpolating surfaces should provide a better indication of whether the population is at least contiguous with the southern one. Evaluation of the age distribution (relative cohort strength) should also provide further information as to whether the population dynamics are consistent with those of a single population. The historical paper on the migration pattern of yellowfin sole (Figure 4 in this report, reproduced from the presentation) suggests that the whole area is used as a spawning ground. Age, size and maturity information from the single survey may provide further clues as to whether the northern Bering Sea still serves this function for the southern population, or if there are even sufficient juveniles to support the biomass estimated for the area as an independent population.

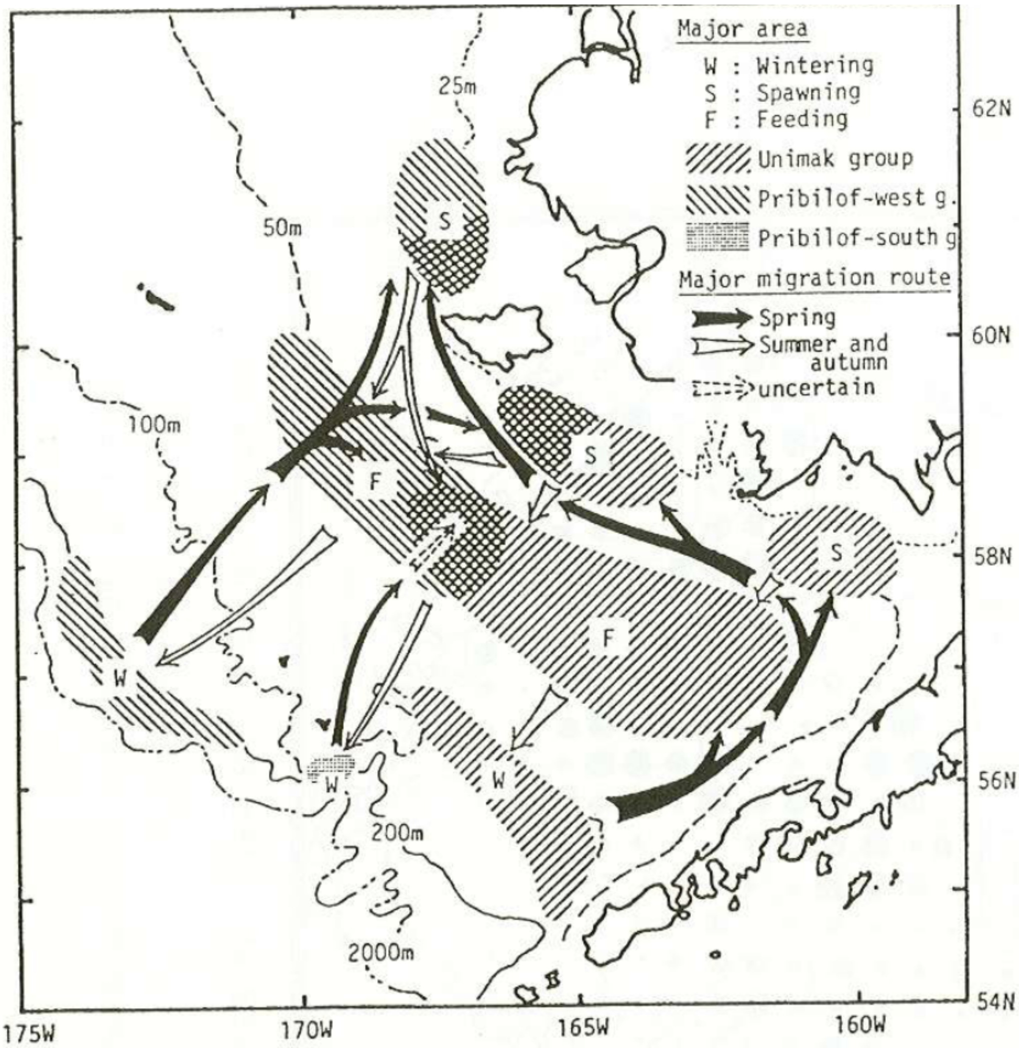

Figure 4: Schematic of annual migration pattern of yellowfin sole. The source of this Figure is not known, but it is taken from the assessment presentation.

Assuming that the evidence suggests that the population is the same, dealing with it will be difficult on the basis of a single year of data. The only way currently available is by use of a deterministic multiplier on q (possibly based on area, as in rex sole). Because there is no fishery in the area, this will merely scale up the population and scale down the value of F. From a management perspective, a 15% gain in catch would be possible assuming the ratio for the northern and southern survey areas, but given the uncertainty around metapopulation structure, this may incur risks for which there currently seem to be little justification owing to the limitation of the fishery by other factors. The additional unexploited biomass in the northern Bering Sea may make the population more resilient to exploitation than is currently assessed.

If the concern regarding northward population expansion or movement with climate change is a real concern, further information will be necessary, at which point the surveys should be designed so that they can address the specific question of population mixing and the relationship with temperature.

Determination of whether the assessment represents the best available science for the stock assessment of BSAI yellowfin sole, including considerations of fishery rationalization on timing and selectivity of fishery

The advice provided by the assessment in my opinion is consistent with that provided by the best available science. This is slightly different from the question, because I struggle with some of the implementations in the model that I feel could be parameterized better (yellowfin sole TOR 1 above). At present, however, the areas of the assessment that I consider would be worthy of further investigation for model improvement are unlikely to change the perception of current estimates of SSB or F.

More likely to affect management is the consideration of an appropriate F_{msy} or proxy thereof is the choice of stock−recruitment relationship (see detailed description above) and the methodology for providing estimates of future stock biomass. Estimation of F_{msy} is based on an assumed stock−recruit relationship, yet stock status in the future with respect to biomass is independent of SSB. Recruitment estimates are drawn from an inverse Gaussian distribution parameterized by the assessment estimates from the assessment. To me this is not an appropriate function to describe recruitment trends in the assessment, appearing to be more bimodal than inverse Gaussian. The situation should therefore be investigated, but more significantly it is strongly inconsistent with two findings in the assessment (although I do not necessarily agree with those conclusions). First, it is suggested that there is an appropriate stock−recruit relationship, and if so it should be used with some deviates. Second, there was a strong environmental shift in the ecosystem prior to 1978 and recruitment from that period is no longer seen as appropriate for estimating present stock dynamics, so those estimates should not be included in estimating the variation in recruitment. If it is deemed that such recruitments may be possible again in future if the environment changes again, then some form of autocorrelation will have to be used or the population will appear to be more stable than it really is during the periods of lower recruitment.

At present, it seems that none of these issues is likely to have an effect on future exploitation of the stock, and management decisions on other species will ultimately determine the levels of exploitation. This is not to say that this stock itself could not be managed for higher productivity, at which point the estimate of stock productivity would be more important in management. Therefore, the basis of advice should be internally consistent and appropriate.

The other question with respect to future exploitation was how appropriate the selectivity from the current assessment might be given the fisheries rationalization programme in the Bering Sea. Given the low rates of exploitation, the effects of any selectivity change would be hard to detect in the population. It would be easier to do so in the time-series of fisheries information, but the assessment model appears to find it difficult to pick up significant changes since 2008.

The fact that selection appears to be more stable in recent years is likely an artefact of the model. Numbers-at-age estimates at the younger ages are more uncertain than at older ages (because of the number of available estimates). Changes in selectivity at older ages are confounded with the estimation of F for the year, because of the asymptotic nature of selection. Selectivity, though able to alter year on year, appears ultimately to be constrained in the assessment by a lack of evidence to the contrary, i.e. the model attributes any variation in the catch-at-age elsewhere.

There is, however, what appears to be a significant shift in the level of fishing mortality commensurate with the fisheries rationalization, suggesting at least for the older ages that the available capacity is able to exert almost twice the fishing mortality it currently does since the abandonment of the old derby-style of fisheries.

Points of clarification

In the process of writing this report and looking into some finer details, some additional questions became apparent. These are points that may require explanation in text or correction, because it is not always clear what was done or what is being presented.

Ricker spawner−recruit parameters

At the meeting and in the presentation it was suggested that the GM recruitment and deviates were the only parameter estimates providing information on recruitment. This is consistent with the description of the key equations used (Table 4.11 in the report). However, the statement differs from the spawner−recruit estimation section (p. 591 of the report) which suggests that annual recruitment estimates were constrained to fit a Ricker model, although at only a later phase of the estimation. The “Parameters Estimated Conditionally” section (p. 590) is inconclusive on the matter, because although it states that two parameters are estimated for the Ricker model, this could be the result of post-model fit. The method employed should be clarified in the text. If the recruitment is constrained to the Ricker model, its effect appears to be relatively minor; at least, recruitment does not track SSB well. Nevertheless, the lack of independence between the two estimates does pose some problems in the estimation of likely stock productivity. This is discussed further in the section on advice below.

Modelling weight-at-age

The assessment model applies a number of different approaches to estimating appropriate weight-at-age to convert numbers-at-age to biomass estimates of spawners and catches. There are some confusing statements in the assessment report with respect to what was actually done. The first and most detailed explanation seems to refer to what was done previously (assuming a single length-at-age, but with a deviate on the weight-at-age applied to the length−weight relationship). Following this, things become more nebulous, but as far as I can tell, none of the new modelling options presented use any form of length information. However, the growth model section (p. 592) implies that Model 1 uses length information, in which case it is not clear how it is used. This should be clarified or the reference removed.

The authors consider constant weight-at-age to be inappropriate because it is not supported by the data. Having looked at the weight-at-age information used in the assessment, I cannot find any long-term trends in the available information. There appears to be much variability in the use of the actual survey data, and the information on the prior period is aggregated over many years so appears to be more constant. A plot justifying the statement that this is not supported by the data would be helpful, or at least an indication of opinion of the authors as to what is meant by "'not supported by the data". Presumably there is good reason for this statement, but I am not able to verify this appropriately. The appropriate choice of model is discussed further below.

Stock productivity

The text on p. 591 of the assessment report suggests three models for recruitment estimation based on different period of recruitment. Models A and C are the full and historical time-series, whereas model B is the recent time-series. The legend in Figure 4.12 top (in the assessment report) is inconsistent with this for the historical time-series because this only goes up to 1977. Figure 4.12 bottom, the recent time-series includes a 1977 point apparently prior to the ecosystem shift. The assessment suggests using Model B, the period with the lowest estimate of F_{msy}, but this appears to be mislabelled in Figure 4.13 (top), where model A is more consistent with what is described as Model B in the text.

Fishing mortality estimates

Tables 4-14 and 4.15 of the assessment report show fishing mortality estimates. Assuming that selectivity at the oldest age shown (age 20 in Table 4-14) represents fishing at full selection, then this should be consistent with the F at full selection vector provided in the subsequent table, but this does not appear to be the case; the difference is unclear to me.

Appendix 1. Bibliography of materials provided for the review of selected Gulf of Alaska and Bering Sea flatfish

Documents cited in this report:

Matta, M. E., Black, B. A., and Wilderbuer, T. K. 2010. Climate-driven synchrony in otolith growth-increment chronologies for three Bering Sea flatfish species. Marine Ecology Progress Series, 413: 137-145.

Nichol, D. R. 1998. Annual and between-sex variability of yellowfin sole, *Pleuronectes asper*, spring-summer distributions in the eastern Bering Sea. Fishery Bulletin US, 96: 547-561.

Stockhausen, W.T., M.E. Wilkins and M.H. Martin 2011. Assessment of the Deepwater Flatfish Stock in the Gulf of Alaska, pp82.

Stockhausen, W.T., M.E. Wilkins and M.H. Martin 2011. Assessment of the Rex Sole Stock in the Gulf of Alaska, pp64.

Turnock, B.J., T. A'mar T. and T.K. Wilderbuer 2011. Assessment of the Shallow-water Flatfish Complex in the Gulf of Alaska for 2012, pp134.

All files (dates size and version) of materials supplied in the drop box:

12/06/2012 04:15 a.m. 13,534 Agenda for 2012 flatfish CIE (1).docx
14/06/2012 04:03 a.m. 873,910 BSAIintro.pdf
12/06/2012 07:40 a.m. 3,505,945 BSAIyfin.pdf
11/05/2012 09:02 a.m. 41,984 CIE agenda for flatfish.doc
11/05/2012 08:42 a.m. 103,478 CIE agenda for flatfish.pdf
10/05/2012 10:20 a.m. 82,944 CIE Flatfish assessment SOW.doc
10/05/2012 10:26 a.m. 147,222 CIE Flatfish assessment SOW.pdf
12/06/2012 05:45 a.m. 1,631,359 CIE Review NPGOP_slides.pdf
12/06/2012 03:57 a.m. 5,006,251 CIE Review NPGOP_slides.pptx
14/06/2012 08:43 a.m. 170,164 CIE review.GOArs.survey_data.pdf
14/06/2012 08:11 a.m. 158,208 CIE review.GOArs.survey_data.ppt
13/06/2012 08:47 a.m. 336,563 CIEreview.GOArs.2012.pdf
13/06/2012 08:08 a.m. 1,209,856 CIEreview.GOArs.2012.ppt
12/06/2012 06:43 a.m. 1,049,578 ciereviewoftheebsbottomtrawlsurvey.zip
13/06/2012 06:30 a.m. 1,607,043 CIEReview_DoverSole.pdf
13/06/2012 04:06 a.m. 5,343,982 CIEReview_DoverSole.pptx
13/06/2012 06:29 a.m. 969,839 CIEReview_GOARexSole.pdf
13/06/2012 04:03 a.m. 3,305,853 CIEReview_GOARexSole.pptx
14/06/2012 09:17 a.m. 226,511 Copy of yfin tempq vs fixed q.xlsx
13/06/2012 02:53 p.m. 16,967 DoverSole_AgeCompsData.xlsx
13/06/2012 04:06 a.m. 24,912 DoverSole_DiffAgeCompComparisonsFemale.pdf
13/06/2012 04:06 a.m. 24,995 DoverSole_DiffAgeCompComparisonsMale.pdf
14/06/2012 09:26 a.m. 297,665 EBS flatfish catchability.pdf
14/06/2012 10:50 a.m. 0 filelist
14/06/2012 08:54 a.m. 6,182 FlatFishCIE_2012.htm
12/06/2012 04:57 a.m. 2,251,218 Flatfish_A&G_CIE review.pdf
12/06/2012 04:42 a.m. 8,348,723 Flatfish_A&G_CIE review.pptx
13/06/2012 04:08 a.m. 3,272,053 GOA2012SAFE_DeepwaterFlatfish.pdf
13/06/2012 04:04 a.m. 2,142,740 GOA2012SAFE_RexSole.pdf
12/06/2012 04:25 a.m. 1,486,832 GOAAOFlatfishCIEpalsson.pdf
12/06/2012 03:56 a.m. 46,803,904 GOAAOFlatfishCIEpalsson.pptx
12/06/2012 09:34 a.m. 590,720 Haynie Abbott CIE A80.pdf
12/06/2012 08:54 a.m. 1,595,864 Haynie Abbott CIE A80.pptx
12/06/2012 04:28 a.m. 2,020,044 Nichol_CIE_6_11_2012.pdf
11/06/2012 03:49 a.m. 8,261,983 Nichol_CIE_6_11_2012.pptx
14/06/2012 02:07 a.m. 30,467 Notes for Dover sole and assmnts.docx
02/12/2011 02:01 p.m. 384,933 Nov_2011_BSAI_Minutes.pdf
02/12/2011 02:00 p.m. 455,094 Nov_2011_GOA_Minutes.pdf

02/12/2011 01:59 p.m. 356,409 Nov_2011_JPT_Minutes.pdf
12/06/2012 10:27 a.m. 2,956,288 overview.ppt
14/06/2012 09:07 a.m. 11,513 Participant list for CIE.docx
13/06/2012 08:47 a.m. 5,539 pbtb.tmp
14/06/2012 09:31 a.m. 12,720 rex sole survey-at-age.xlsx
13/06/2012 04:14 a.m. 14,564 RexSoleAgeCompsData.xlsx
13/06/2012 05:23 a.m. 48,028 RexSoleSurvey.docx
13/06/2012 04:04 a.m. 16,777 RexSole_DiffAgeCompComparisonsFemale.pdf
13/06/2012 04:04 a.m. 16,716 RexSole_DiffAgeCompComparisonsMale.pdf
13/06/2012 04:08 a.m. 616,335 SomertonEtAl2007_FishBull.pdf
13/06/2012 12:20 p.m. 745,100 SSC Dec2011Draft.pdf
14/06/2012 08:45 a.m. 531,896 Stark Somerton growth of rock soles.pdf
12/06/2012 11:43 a.m. 23,040 SUMMARY_2011_BSAI.xls
12/06/2012 03:43 a.m. 26,112 SUMMARY_2011_GOA.xls
14/06/2012 09:17 a.m. 226,511 yfin tempq vs fixed q.xlsx
13/06/2012 05:59 a.m. 87,717 yfs age comps.xlsx
14/06/2012 08:51 a.m. 2,243,401 yfs presentation to cie panel.pdf
13/06/2012 11:59 a.m. 5,231,019 yfs presentation to cie panel.pptx
11/05/2012 08:53 a.m. 3,581 _pgbrz_temp.htm

Appendix 2: Attachment A: Statement of Work for Dr Sven Kupschus (Cefas)

External Independent Peer Review by the Center for Independent Experts

Peer Review of the BSAI and GOA flatfish stock assessments

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: The Alaska Fisheries Science Center (AFSC) requests a Center of Independent Experts (CIE) review of 4 Bering Sea/Aleutian Islands (BSAI) and Gulf of Alaska (GOA) flatfish stock assessments. They include: GOA northern and southern rock sole, GOA Dover sole, GOA rex sole and BSAI yellowfin sole. The BSAI and GOA flatfish resources are large, subject to significant fisheries and are key components of the BSAI and GOA ecosystems. The flatfish stock assessments routinely undergo thorough review by the AFSC, the North Pacific Fisheries Management Council's Groundfish Plan Teams and Scientific and Statistical Committee, and members of the public. However, the BSAI and GOA flatfish stock assessments have not had the benefit of a CIE review since 2007. Since 2007, several modifications to existing assessment and projection models have been implemented, and a new assessment for Gulf of Alaska northern and southern rock sole has been developed. These innovations have not been reviewed by the CIE. The Alaska Fisheries Science Center desires an independent peer review of these stocks to assess the quality of the assessments and to ensure that the North Pacific Fishery Management Council bases its decisions on the best available information. Therefore, a CIE review in 2012 would be timely. 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, including population dynamics, survey methodology, estimation of parameters in complex nonlinear models, and the AD Model assessment program in particular. Reviewers should also have experience conducting stock assessments for fisheries management. 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, Washington tentatively during June 11-13, 2012.

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/sponsor.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.

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 can not 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.

Contract Deliverables - Independent CIE Peer Review Reports: 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.

Other Tasks – Contribution to Summary Report: 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 in the panel review meeting in Seattle, Washington during June 11-13, 2012.
3) In Seattle, Washington during June 11-13, 2012 as specified herein, conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than July 9, 2012, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to 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 schedule.

May 1, 2012	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
June 1, 2012	NMFS Project Contact sends the CIE Reviewers the pre-review documents
June 11-13, 2012	Each reviewer participates and conducts an independent peer review during the panel review meeting
July 9, 2012	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
July 23, 2012	CIE submits CIE independent peer review reports to the COTR
July 30, 2012	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) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each 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:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Tom Wilderbuer, NMFS Project Contact
NMFS Alaska Fisheries Science Center
7600 Sand Point Way NE., Seattle, WA 98115-6349
tom.wilderbuer@noaa.gov Phone: 206-526-4224

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: Tentative Terms of Reference for the Peer Review

2012 CIE Review for selected Gulf of Alaska and Bering Sea flatfish

GOA Dover sole TOR

CIE Reviewers shall evaluate the current model assumptions and make recommendations for improvements thereof, including:

1. Use of age data, including:
   a. use of age composition data
   b. appropriateness of age range and binning
   c. estimation of size-at-age relationship and variability (external vs. internal to model)
   d. inclusion of ageing error
2. Use of size data, including:
   a. use of survey size composition data
   b. use of fishery size composition data
3. The number and functional forms of estimated selectivity curves, including:
   a. fitting different selectivity functions to data from different survey years based on survey depth coverage
   b. types of selectivity curves considered
   c. use of age-based vs. size-based selectivity curves
   d. allowing for annual variability in fishery selectivity
   e. use of size-based selectivity curves for survey data based on trawl net catchability experiments
4. Fixing (and updating) the natural mortality rate based on Hoenig, 1983.
5. Model convergence diagnostics

GOA rex sole TOR

CIE Reviewers shall evaluate, and make recommendations for improvements on, the current approach to determining stock status and future harvest reference points (ABC and OFL).

GOA northern and southern rock sole TOR

3. Evaluation, findings and recommendations of the analytical approach (application of a statistical ADMB integrated catch-age model) used to assess stock status and estimation/presentation of uncertainty.

4. Evaluation findings and recommendations on quality of input data and methods used to process them for inclusion in the assessment (specifically fishery and survey data).
5. Recommendations for further assessment improvements for management in both the long and short term.

BSAI yellowfin sole TOR

Evaluation of the analytical approach (application of a statistical ADMB integrated catch-age model) and model assumptions used to assess stock status and stock productivity.
Evaluation of the implications of using the Northern Bering Sea research results as an index of abundance if yellowfin sole increasingly occupy this area with changing climate.
Determination of whether the assessment represents the best available science for the stock assessment of BSAI yellowfin sole, including considerations of fishery rationalization on timing and selectivity of fishery.

Annex 3: Tentative Agenda

Note: Final Agenda to be provided two weeks prior to the meeting with draft assessments and background materials.

CIE Flatfish assessment review

NMFS Alaska Fisheries Science Center
7600 Sand Point Way NE, Building 4
Seattle, Washington

4. AGENDA *JANUARY 20 DRAFT VERSION* June 11-13, 2012

Monday June 11th

9:00 Welcome and Introductions, adopt agenda **Sandra**
9:15 Overview (species, biology, surveys, fishery, catch levels, ABCs, TACs, bycatch) **Tom**
10:00 Bering Sea trawl survey **RACE Division**
10:30 Gulf of Alaska trawl survey **RACE Division**
11:00 Coffee break
11:20 Observer Program **FMA Division**
11:50 Age Determination **Delsa and Beth**
12:30 Lunch
1:30 Effect of rationalization on flatfish fisheries **REFM Economic subtask**
2:30 GOA rex sole **Buck**

Tuesday June 12th

9:00 Gulf of Alaska Dover sole **Buck**
11:00 Coffee break
11:20 Gulf of Alaska Dover sole (continued) **Buck**
12:30 Lunch
1:30 Gulf of Alaska northern and southern rock sole **Teresa**

Wednesday June 13th

9:00 Bering Sea yellowfin sole **Tom and Jim**
11:00 Coffee break
11:20 Bering Sea yellowfin sole (continued) **Tom and Jim**
12:30 Lunch
1:30 CIE panel discussion (assessment authors will be available)

Appendix 3: Participants in the 2012 AFSC Flatfish Review Panel, 11–13 June 2012, AFSC, Seattle, Washington, USA

Sven Kupschus	Cefas Lowestoft UK
Yan Jiao	Virginia Tech
Kevin Stokes	
Anne Hollowed	AFSC Status of stocks
Buck Stockhausen	AFSC Status of stocks
Teresa A’mar	AFSC Status of stocks
Tom Wilderbuer	AFSC Status of stocks
Sandra Lowe	AFSC Status of stocks
Jim Ianelli	AFSC Status of stocks
Loh-Lee Low	AFSC International coordination
Alan Haynie	AFSC Economics program
Dan Nichol	AFSC Bering Sea survey program
Wayne Palsson	AFSC Gulf of Alaska survey program
Tom Helser	AFSC Age and growth program
Lisa Thompson	AFSC Observer program