Dover Sole STAR Panel Report

National Marine Fisheries Service
Hatfield Marine Science Center
Newport, Oregon
July 25-29, 2011

Review Panel Members:

Vidar Wespestad (Chair), SSC representative
Kevin Stokes, Center for Independent Experts
Sven Kupschus, Center for Independent Experts
David Sampson, Oregon State University and SSC

STAT Team Members Present:

Allan Hicks, Northwest Fisheries Science Center

Advisors:

John DeVore, Pacific Fishery Management Council (PFMC) Staff
Joanna Grebel, PFMC Groundfish Management Team representative
Gerry Richter, PFMC Groundfish Advisory Subpanel representative

Overview

During 25-29 July 2011 a Stock Assessment Review (STAR) Panel met in Newport, Oregon to review a draft stock assessment document for Dover sole that had been prepared by prepared by Hicks and Wetzel (2011). The Panel operated under the Pacific Fishery Management Council's (PFMC) Terms of Reference for the Groundfish Stock Assessment and Review Process for 2011-2012 (PFMC 2010). This same panel also reviewed a draft assessment for sablefish.

The Dover sole stock assessment was conducted using a recent version of Stock Synthesis 3 (SS3 ver3.21f) applied to data from several sources. Although a draft assessment document was distributed on schedule before the Panel meeting, subsequent to the distribution but prior to the meeting the STAT discovered an error in its reconstructed catches for WA. Prior to the meeting a new base case was run and this was used during the meeting as a starting place. The corrected run produced only minor changes to distributed base case assessment results; nevertheless, the results presented in the draft document were erroneous. Results presented by the STAT during the STAR Panel meeting were based on runs using the corrected catch data. Because the corrected catch streams make little difference, conclusions drawn from analyses using the uncorrected data are still valid and were referred to during the review. The author stated that tables and figures in the body of the assessment document will be corrected for the version of the document that will be reviewed during the September PFMC meeting.

The Dover sole assessment assumes a unit stock in the waters off Oregon, Washington, and California. All model configurations included available length, age, and biomass data from five (splitting the Triennial survey) bottom trawl surveys of the slope or shelf, and available trawl fisheries length and age data from CA, OR and WA. The estimated catch history, split by the three states, extended back to 1911. All parameters were freely estimated except for steepness, which was fixed at 0.8.

The new assessment model for Dover sole is much simplified compared to previous models. The STAR requests to the STAT focused on exploring whether the simplified model structure provided an adequate representation of the stock dynamics, sufficient to provide a basis for management advice. In particular there was some difficulty reconciling tensions between signals from length and age data which was compounded by the way male and female selectivities are linked in SS3. After reviewing results from several exploratory analyses prepared by the STAT the Panel was satisfied that the proposed base model adequately fit the available data and that uncertainty was suitably characterized.

The STAR recommends the new Dover sole stock assessment as the best available science and that it provides a suitable basis for management decisions.

Requests by the STAR Panel and Responses by the STAT

REQUEST 1 i) Present likelihood components from model runs with normal vs log-normal assumption on growth; and ii) Plot histograms of size across selection of ages

RATIONALE Usually length-at-age is assumed to follow the normal distribution, despite the fact that growth being a multiplicative random process should lead to a log-normal distribution. Nevertheless the effect of this departure should be examined in terms of model fit and to confirm the appropriateness in the original data.

RESPONSE The STAT delivered the required information. Plots of the data indicated some skew in the length distributions particularly for the females in the raw data. Log-likelihood components indicated an improved fit for both the age and length information, with a small but significant overall improvement over assumed normal error distributions. Plots of the fit grouped across a number of ages showed little difference overall, but the modes tended to match better with the raw data for the lognormal distribution.

REQUEST 2 i) Run model with simplified selex on the two slope surveys (use double normal and with female offset); ii) Run model with slope surveys dropped.; and iii) Run model with 5 node cubic spline on combo surveys.

RATIONALE The aim was to investigate how the slope survey data influence model interpretation of cohort strength/progression, and the overall effect on model fits and outputs. In particularly the way the model might interpret the slope surveys to see if it could not be forced to accept a bimodal distribution in the population rather than selectivity.

RESPONSE The STAT ran the required alterations from the base model and summarized the information. The simplified selectivities (Request 2 i) of the two slope surveys severely dampened the double humped selectivity curve in the female structure as expected, but at a considerable cost to the overall likelihood, though improving the age contribution. Adding additional complexity (Request 2 iii) lead to some odd behavior in the male selectivity, as there was no data available at the largest lengths for males for the spline function to work on but through the linkage with the female selectivity catchability rose exponentially with estimates of q rising to over 4. Bi-modality was retained in the female selectivity although the relative size of the modes changed. Removing the slope surveys altogether (Request 2 ii) could not be compared using log-likelihoods, but management parameters output by the model indicated very little difference in the results.

REQUEST 3 Run model turning off asymptotic selex on males in commercial fleets.

RATIONALE Concern at the persistent bias in length fits (too few large males in the catches) in the commercial fleets. There was concern that the asymptotic assumption was constraining the model in appropriately.

RESPONSE The STAT provided the requested run. Although increasing the flexibility of the commercial selectivity for males did significantly improve the overall fit to the likelihood, but at the cost of what was deemed to be excessive complexity in the commercial selectivities with some unexpected interactions between male and female selectivity. In terms of model results there was little effect on the changes to the important management parameters.

REQUEST 4: investigate calculation of Pearson residuals in age-at-length bubble plots when using log-normal errors on length.

RATIONALE The plots appear wrong, with high preponderance of dark (i.e. underestimated) points and few open bubbles.

RESPONSE The STAT examined the output as requested. The Pearson's residuals are based on the differences between observed and fitted after retransformation so should not be influences by the log-normal assumption. Increasing the size of the plots confirmed the similarity between the lognormal and normal length modeling.

REQUEST 5: Provide summary table of retrospective estimates of key parameters, including natural mortality, selectivity, growth and discard rates (and check if converged).

RATIONAL The concerns were that there was indication of a retrospective bias in the data and the concern was that the effect of the retrospective was to reduce the influence of the NWFSC combo survey which implied different estimates of natural mortality compared to other data sources. Decreasing the weight of this series should therefore have had the effect of decreasing natural mortality, but the management parameter estimate output suggested biomass estimates were not ordered sequentially with decreasing time period. The panel wished to examine how the estimation of key parameters including natural mortality and selectivity changed during the retrospective process.

RESPONSE The STAT provided the requested summaries from the retrospective runs performed. The log-likelihood components for other surveys tended to increase as the length of the NWFSC combo survey decreased underlining the conflict in the information between the different surveys. Virgin biomass estimates changed relatively little over the series with some variability in the estimate of current biomass until the 2005 run, when there was a step in the estimate of natural mortality leading to a significant increase in natural mortality for both sexes which results in a sizeable increase in the virgin biomass, but predicts current biomass to be very similar to the base case. The panel examined this inconsistency further under request 6.

REQUEST 6: tabulate Log Likelihood components and key model parameters when holding female M at 0.115 and male M across a range of values.

RATIONALE The panel felt the need to understand the tensions between the different model parameters and data sources to better understand the model dynamics in order to determine the appropriateness of the model dynamics to evaluate the uncertainty.

RESPONSE The STAT provided a graph showing the log-likelihood components illustrating the conflict between the different data series (Figure 1). Survey indices contained little information regarding a minimum natural mortality but increased monotonically with natural mortality. Length data suggested that very low levels of natural mortality were unlikely, but again contained little information on the appropriate choice of male natural mortality above 0.12. As expected age provided the most contrast in the data to inform on natural mortality, but also indicated the greatest degree of conflict between the information sources. Commercial data suggested natural mortality to be 0.11 (Oregon, Washington) to 0.13 (California), with the NWFSC slope survey indicating natural mortality of 0.13 or greater (little change in LL above).

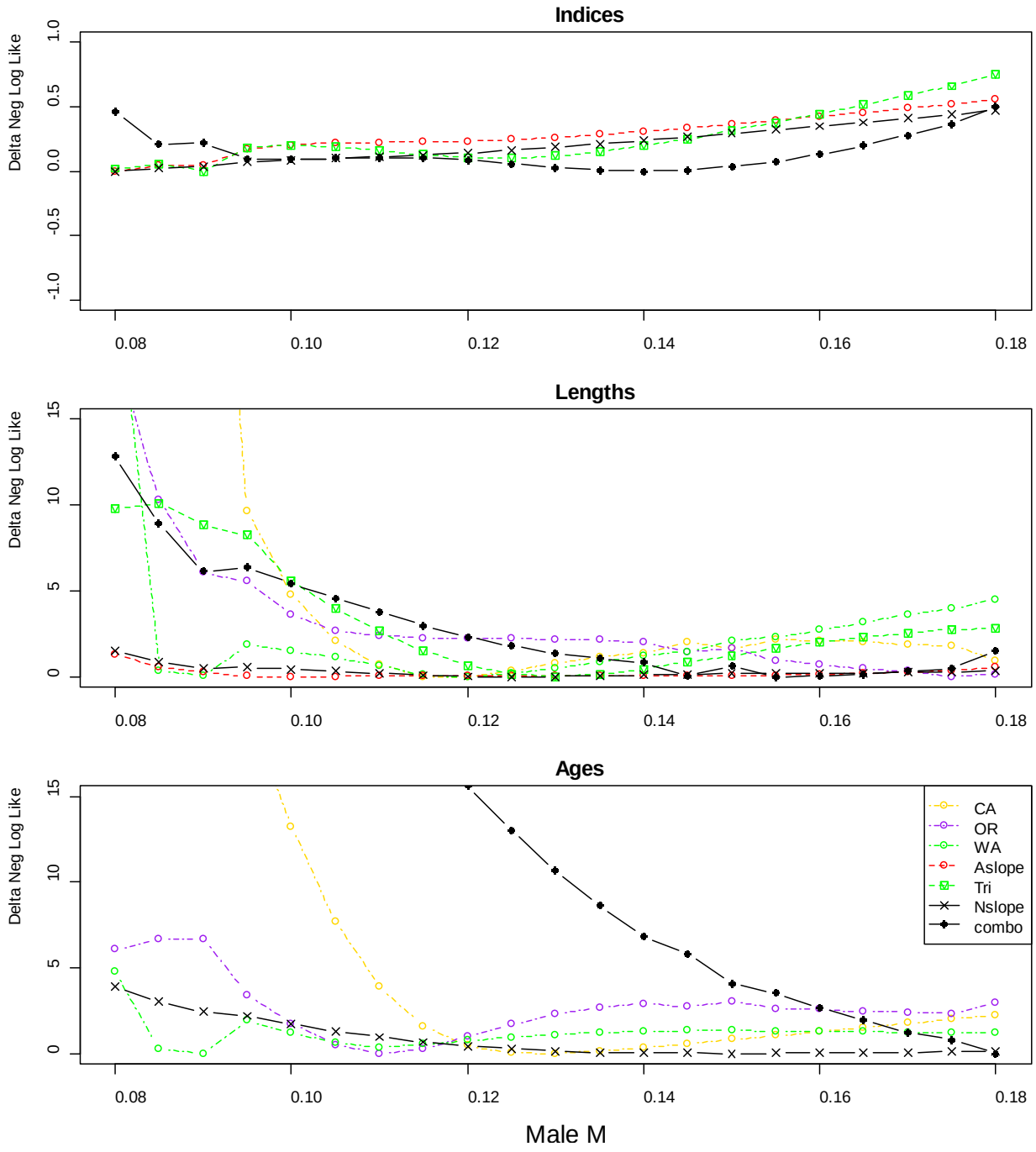

Figure 1: Illustrating the log-likelihood contributions of the various survey sources to the estimation of natural mortality (M) simulated here by fixing male natural mortality in the base model over the range of 0.08 to 0.18.

In contrast, the NWFSC combo survey indicated much higher levels of natural mortality not reaching a minimum over the range examined. Commercial male data is assumed to have an asymptotic selection and therefore should contain the most information on natural mortality, where as the NWFSC combo survey with dome shaped selection should contain less information

due to the correlation between the parameters, so that these LL components are conditional on the choice of the selectivities applied in the model. It is interesting to see that the length information from the NWFSC combo survey indicates far less conflict with lower natural mortality so that there is unresolved conflict even within the information from a single survey. Changing the selectivity of the survey did not dramatically improve the log-likelihood overall, though its effect on the uncertainty of natural mortality was not examined. As there are persistent differences in the sex ratio between the length and age information from the combo survey this may imply some non representative sampling.

The STAT provided summary plots of the changes in selectivity pattern with changes in male natural mortality based on the runs conducted for the sensitivity analysis. One thing that became very clear from this investigation is the apparent link between male natural mortality and the estimate of female SSB, two parameters that intuitively should not be correlated. As male natural mortality increases under the model structure, female biomass increases dramatically. The link between the male and female selectivities was found to be the root cause. In SS3 female selectivity is modeled as an offset to male selectivity and with differences in the sex ratio forcing higher selectivity on males. Increasing male natural mortality decreases numbers of older males, which increases the selectivity of older males (modeled as asymptotic) and consequently smaller males (Figure 2). This influences the offset applied to smaller females and through the restrictive double normal the selectivity of larger females decreases their susceptibility, hence increasing numbers of large females and consequently estimates of spawning biomass.

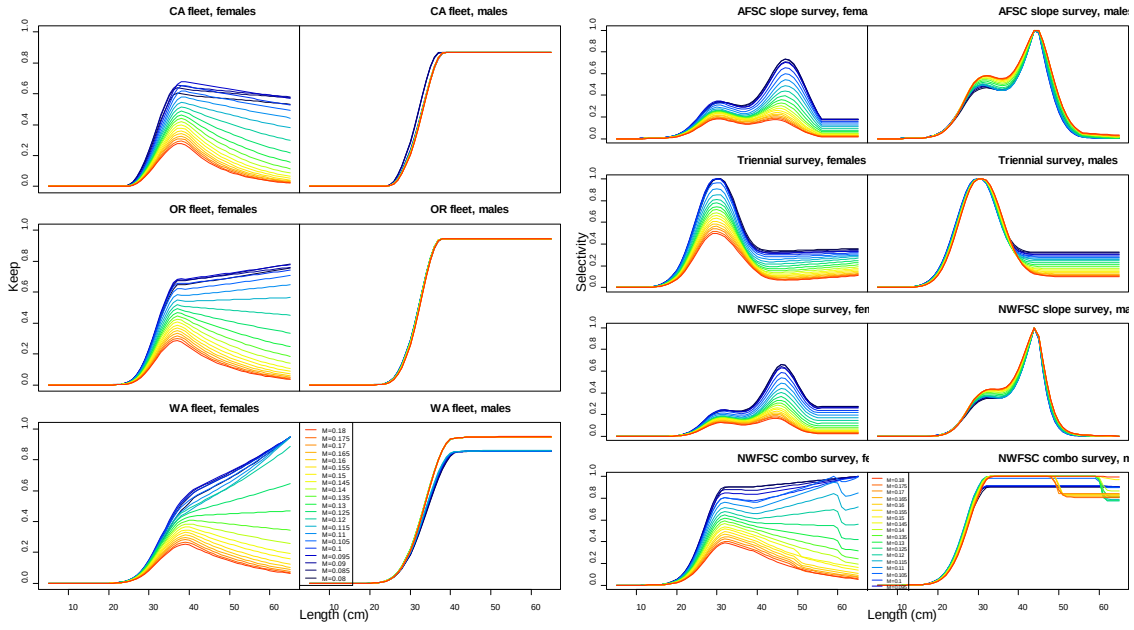

Figure 1: Changes in selectivity patterns for fleets (left panel) and surveys (right panel) in response to changing male natural mortality across an interval of uncertainty in SSB as a consequence of the offset selectivity function implemented in SS3. Over the range of male M considered (0.08 – 0.18) the female spawning biomass varied from 73,000 to 2,000,000 mt.

REQUEST 7 Examine the possibility of using age-specific natural mortality (by sex).

RATIONALE Part of the problem with the assessment is the asymmetry in the sex ratio at older ages. The Panel wished to investigate whether allowing for age-specific natural mortality (different by sex due to growth differences) could help to reduce the conflict.

RESPONSE The STAT implemented the Lorenzen M option in SS3 and provided the requested model output. The requested change resulted in a small increase in the overall log-likelihood, but more importantly the model was still unable to resolve the conflict at the larger ages, because the difference in growth between sexes was insufficient to resolve the discrepancy of the sex ratio at older ages.

REQUEST 8 Provide plots of SSB and depletion with MLE-based confidence limits for the base model and for an equivalent run of the base model with male M fixed at the value estimated in the base model, as well the correlation between virgin and current SSB based on an MCMC simulation for the base model.

RATIONALE The uncertainty in current SSB is large due to the formal inclusion of the uncertainty in natural mortality. Although it is commendable to include this parameter formally in the model estimation there are concerns regarding the interpretation of the information. Virgin and current SSB change in unison so that the uncertainty around depletion, the management target in this case, is very much less dependent on M than the estimate on current SSB potentially causing some misinterpretation of the uncertainty when evaluating the P* approach with respect to current SSB alone.

RESPONSE The STAT provided the requested plots (Figure 3,4). The results confirm the reasoning behind the rational and show that the depletion on which the management target is based is much more certain than the estimates of the current biomass. It would be worthwhile considering if in these cases it would not be more appropriate to investigate uncertainty based on depletion rather than current SSB.

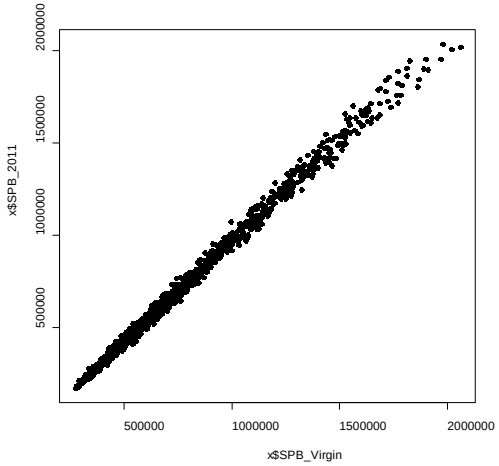

Figure 3: MCMC simulation results illustrating the interdependence on virgin and current SSB estimates in the base model.

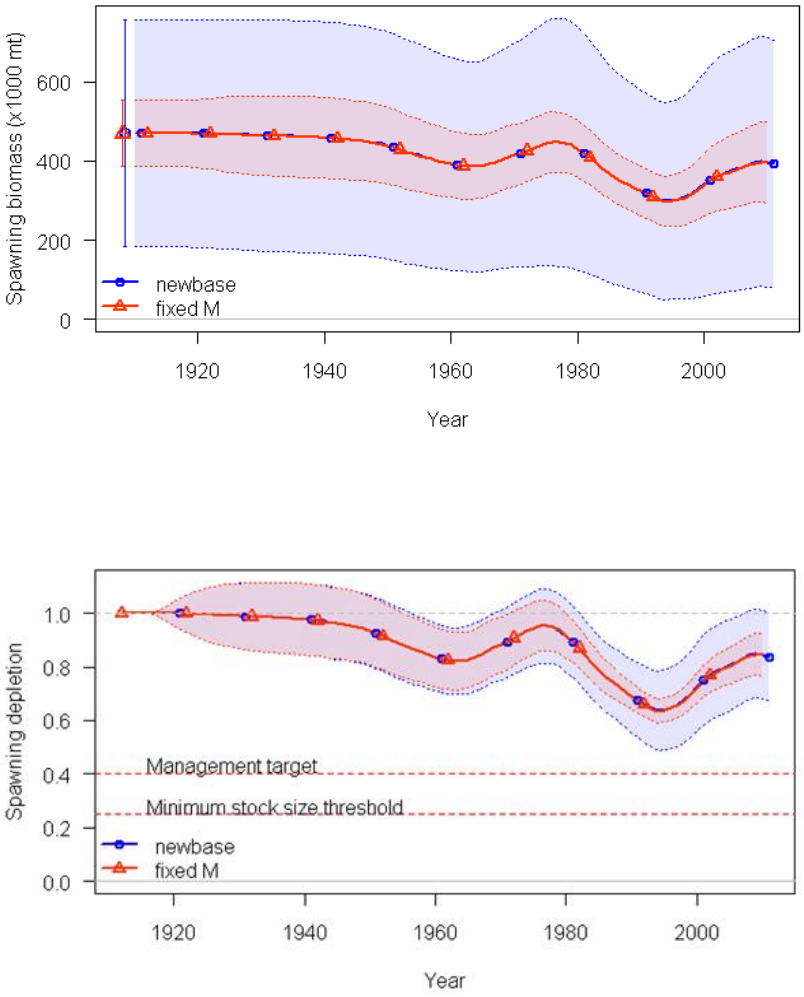

Figure 4: Plot of MLE confidence limits around SSB and depletion for the base case and the same model with M fixed at the level of the base case M estimate to illustrate the difference in the dependence on M estimation for these management parameters.

Description of base model and alternative models used to bracket uncertainty

The premise behind the modeling approach was an attempt to simplify previous assessment models for this stock in the hope of increasing parsimony in the assessment. The base model was developed in SS3 (ver3.21f) using age and length based catch data in three commercial fleets (California, Oregon and Washington) using a plus group age of 60. Fisheries independent information consisted of four survey time-series (ASFC Tri-annual, ASFC slope, NWFSC slope

and NWSFC shelf slope combo) providing length information. The NWFSC surveys also provided age information.

Investigations conducted during the panel indicated that the estimation of natural mortality (gender specific) represented the major uncertainty in the model with respect to biomass estimates within the model. As described above there was an undesirable linkage between male natural mortality and female spawning biomass estimates caused by the SS3 implementation using an offset linking male and female selectivities. However, the panel determined that although the large uncertainty around SSB was questionable, this estimate of variability, as covered by the range of natural mortality defined as the states of nature, was likely overly precautionary despite the compromised stock dynamics and as such are useful measures of the uncertainty of the current SSB.

The TORs required the development of different states of nature on the basis of the inter-quartile range in the uncertainty. This information is presented in the final assessment report as the uncertainty in current SSB. However investigations conducted during the review indicated that depletion was less sensitive to uncertainty in natural mortality due to the low exploitation of the stock and wondered whether a P* approach based on depletion would not be more appropriate than one based solely on the estimate of current spawning biomass, which varies in concert with virgin biomass estimates in this assessment.

Comments on the assessment

- Very complete pre-STAR draft assessment document.

Technical merits:

- The model was much simplified structurally compared to previous assessments. This followed a careful examination and consideration of data availability and possible complexity in underlying population dynamics and fishing patterns. The simplification of the model necessitates that fits strike a balance when fitting data and results in some residual bias. This is inevitable but is appropriate and allows for better characterisation of uncertainty than fitting a more complex, over-parameterized model with sparse data (and low information).
- The resulting high uncertainty in parameter estimates encompasses much of the uncertainty seen through extensive sensitivity testing allowing for a simpler development of the decision table accounting for the identified major axes of uncertainty (using male and female natural mortality combinations chosen to be consistent with base case model 12.5 and 87.5 percentile estimates of spawning biomass).

Technical deficiencies:

There are no technical deficiencies relevant to providing short term management advice for Dover sole, although Fmsy, long-term management or management at significantly higher levels of exploitation will be compromised if based on this assessment.

Explanation of areas of disagreement regarding STAR Panel recommendations

Among STAR Panel members (including GAP and GMT representatives)

There were no major disagreements.

Between the STAR Panel and STAT Team

There were no major disagreements.

Unresolved problems and major uncertainties

Two areas of the assessment were investigated in detail during the meeting. The first related to the information contained in the NWFSC slope survey, the other concerned the implementation in SS3 of linkage between male and female selectivity.

Selectivity information provided by the AFSC and NWFSC slope surveys was modelled as a cubic spline smoother resulting in a bi-modal selectivity curve by length for both females and males. On close examination the second selectivity mode for males was at lengths greater than those recorded for males in the survey and was based on the dependence of female selectivity on male selectivity, as implemented in SS3. Considerable discussion developed on the merits of the spline function used to model selectivity and on the imposed linkage between male and female selectivity imposed in SS3. The issues could not be completely resolved during the meeting and there is a need for further work in this area.

One issue to be resolved with the assessment relates to the dynamics of the model revealed by investigations into the sensitivity to natural mortality (se Request 6). Estimates of current female SSB are dependent on the estimates of male natural mortality, for which there are no obvious biologically plausible reasons. Close examination of the model identified the cause as the linkage between male and female selectivity imposed by SS3. Commercial male selectivity is modelled as asymptotic, with female selectivity modelled as double Normal but linked to male selectivity by a common offset parameter. The idea behind this is that selectivity is a function of length-based, gear interactions. However, in this case, selectivity represents a multi-dimensional process representing spatial segregation of males and females in addition to the gear related process; the dependence between the gender specific selectivities is therefore no longer appropriate. In the current implementation of SS3 it is not possible to decouple the selectivities. It is recommended that future implementations of SS should allow independent modelling of the selectivities.

Given the process error inherent in the misspecification of selectivity, estimates of uncertainty in biomass generated from the base model may be over-stated.

Management, data, or fishery issues raised by the GMT or GAP representatives during the STAR Panel.

The issues that were raised are adequately described elsewhere in this report.

Recommendations for future research and data collection.

General (affecting more than one assessment)

1. Complete and review the Washington catch reconstruction and review the California and Oregon catch reconstructions. The accuracy and wide availability of consistent basic information is essential to the development of Pacific coast assessments. In addition to the raw data, the reliability and availability of more spatially dis-aggregated forms of the data should be investigated to determine if they could be used to develop more spatially explicit models without causing sacrifices in accuracy.
2. The difficulties encountered in the Dover sole assessment and some other flatfish assessments with respect to the linkage between selectivities require addressing. Although in many instances sized based selectivity may be appropriate, when sexes separate spatially there is a requirement for models to at least be able to investigate complete independence between genders. It is important that this be implemented in an updated version of SS3.
3. The panel investigated the use of age-specific natural mortality in both assessments presented during STAR 4. In each case, one of the reasons for exploring different mortality schedules was the difficulty in fitting the imbalanced abundance at age information (as seen through residuals to fits), either in the sex ratio at older ages (Dover sole) or the ratio of young to old fish (Sablefish). The use of Lorenzen M based on a decline in natural mortality by the inverse of the growth rate implies a link with predation; however, wider use and development of some guidance on the appropriateness of the implementation in other stock assessments should be investigated.
4. Currently the only available error distribution for age information is the multinomial probability function. It appears that this may have some impact with respect to underestimating strong year-classes and it would be desirable to explore the use of alternative error assumptions in order to analyse survey information, in particular where variance estimates in catches-at-age may be less than independent on abundance.
5. There should be new studies of maturity by length and age based on more comprehensive coastwide and depth-based sampling and using histological techniques for determining maturity stage. Given that there is uncertainty regarding the temporal stability of maturity schedules, there should be periodic monitoring to explore for changes in maturity.
6. Update the STAR Terms of Reference to ensure that assessment documents include standard plots (or tables) of likelihood profiles that include likelihood components by data source and fleet. Such plots are an important diagnostic tool for displaying tensions among data sources.profiles.

Specific to Dover sole

1. Researching ageing error, particularly aging bias, is important for Dover sole given the current base models difficulty with reconciling some tensions between different data sources regarding the sex ratio at the oldest ages. In addition, the ability of the model to track cohorts accurately

would be significantly disrupted if there were severe size-based bimodality in cohorts caused by vastly different times of settlement (Dover sole are thought to have a larval period of 6-18 months). Consequently, larval period should also be examined.

2. For the NWFSC combo survey, raw age and length information appeared to imply persistently different sex ratios when viewed in isolation. The concern is that there is some unrepresentative sampling occurring in the age distribution as ages are sub-sampled from length. The sampling procedure should be investigated more closely and potentially improved.
3. The conclusions of the NMFS workshop on developing priors on catchability were not available to the Panel. These should be made available and the information reconsidered specifically with respect to Dover sole, in an attempt to reconcile the relatively low catchability estimates for the surveys, particularly the NWFSC combo survey which is thought to cover the majority of the stock distribution.
4. Having simplified the model compared to previous assessments, especially with respect to uniform growth, it is important to continue investigating if this is likely to introduce undesirable levels of bias into the assessment process as more information becomes available. Spatial information on the distribution by age/size of females, particularly in the southern part of the rang, particularly across the stratification boundaries of the survey as well as between stocks, should be the primary focus of this work.

References

Pacific Fishery Management Council (2010). Terms of Reference for the Groundfish Stock Assessment and Review Process for 2011-2012.

Hicks, A.C., and Wetzel, C. (2011). Status of the Dover Sole off the U.S. West Coast in 2011.