Agenda Item G.4.a
Attachment 8
September 2007

Darkblotched Rockfish

STAR Panel Meeting Report

NOAA Western Regional Center
7600 Sand Point Way NE
Seattle, Washington 98115
July 16-20, 2007

STAR Panel
Tom Jagielo, Washington Department of Fish of Wildlife, SSC member, (Chair)
Patrick Cordue, Center for Independent Experts (CIE)
Stephen Smith, Center for Independent Experts (CIE)
Larry Jacobson, Northeast Fisheries Science Center (NEFSC)

PFMC
John Wallace (GMT) Representative
Pete Leipzig, (GAP) Representative

STAT
Owen Hamel, Northwest Fisheries Science Center (NWFSC)

Overview

A STAR Panel met at the NOAA Sand Point facility from July 16-20, 2007 to review a draft assessment of darkblotched rockfish. The assessment was conducted with SS2 and modelled one fishery and four fishery-independent surveys. Length, age, and conditional age-at-length observations were also included. The assessment was one of the first to employ conditional age-at-length data.

The Panel was initially concerned about the large differences in scale between the area-swept and GLMM indices for three of the four trawl surveys. However, the trends estimated by each method were similar, so the choice of method did not unduly influence the assessment results.

The Panel discussed the issue of properly computing effective sample sizes for multinomial distributions, particularly with respect to the conditional age-at-length data. Ian Stewart gave a presentation that helped the Panel understand the basis for the starting point used to tune effective sample sizes in the model. The subject of how to best determine effective sample sizes with respect to conditional age-at-length data was not resolved at the meeting.

The Panel and STAT discussed at length the value chosen for steepness during the meeting. Historical precedent, meta priors, and model sensitivities were examined. The Panel and STAT did not reach full agreement on this issue (see Areas of Disagreement, below).

STAR and STAT members agreed that the range of uncertainty was not fully captured in model runs. The major axes of uncertainty considered were steepness, and natural mortality, but, for example, uncertainty in catch was not evaluated. A full Bayesian MCMC analysis may provide a useful tool for evaluating the full range of uncertainty in the assessment (see Research Recommendations, below).

The Panel concluded that the final assessment represents the best scientific information available and that the assessment was suitable for use by managers. The Panel commends the STAT for excellent presentations, wellwritten and complete documentation, and their willingness to respond to the Panel's requests for additional analyses.

Requests Made to the STAT during the Meeting

Round 1 Requests

A: Compare absolute scale and trends of area-swept biomass indices with GLMM biomass indices. Also tabulate base model estimates of q for each time series and the "implied" estimates of q for the area-swept time series.

Reason: To determine if the scale and/or trends of the GLMM indices differed from those of the area-swept indices. Also, to perform a "reality" check on the estimated *q*s for each GLMM and/or the corresponding *q*s for the area-swept indices (as if they had been fitted instead).

Response: The GLMM indices were substantially larger than the area-swept indices for every survey except the AFSC slope time series. The trends were very similar for every time series except the NWFSC shelf survey. The estimated *q*s for the GLMM time series ranged from 0.14 to 0.44, and for the area swept time series they ranged from 0.18 to 2.0.

Discussion/conclusion: Concern was raised by some members of the Panel with regard to the large differences in scale between the GLMM and area-swept indices. However, because the trends were very similar for most time series (with the exception of a short "recruitment" series) it was decided not to pursue this issue at the current meeting. There was discussion with regard to whether the estimates of *q* could be used to perform a "reality check" in the absence of informed priors. The Panel had mixed views on this, but as there was no obvious cause for concern about the estimated *q*s in the base model, the decision was made not to pursue the issue at the current meeting.

B: Compare GLMM biomass time series on the same plot. Tabulate or otherwise display CVs.

Reason: To better judge if the trends of the different time series were consistent given their CVs.

Response/conclusion: The three "adult" biomass time series showed consistent trends (with the exception of the 2003 spike in the NWFSC slope series).

C: Tabulate or otherwise compare the number of trips and number of fish sampled in each length frequency series with the initial effective *n*. Also tabulate the final multipliers required to get to each effective *n* in the base run. Repeat for age frequencies.

Reason: To better understand the level of sampling underlying the length and age data, and how these related to the initial effective sample sizes and those used in the base model.

Discussion/conclusion: For the length frequency data some understanding of the temporal and spatial extent of sampling was gained and the sample sizes are now available for reference. For the conditional age-at-length data "sample sizes" were presented within length classes which was not particularly helpful. The original assessment calculated effective sample sizes for conditional age data independently for each length class in a year when the number of trips for a length class was taken to be the number of trips from which age data for that length group were obtained. The STAR panel noted that effective sample sizes for all length groups in the same year was linked. For example, the number of trips sampled was the same for all length groups while the number of ages collected from each length bin is a random variable.

D: Plot initial effective *n* and output *n* for each year for each time series (length and age frequencies).

Reason: This was a mis-specified request – see Request L below.

Discussion/conclusion: These were briefly looked at and, for some time series, the trends of the initial effective *n* differed from those of the output *n*. This was assumed to be due to the interaction between data sets (e.g., when a segment of a time series has no "competition", it will fit well and hence increase the effective *n* in these years; and conversely, when a segment of a time series is in conflict with other data its fit may be poor and the associated effective *n* will decrease).

E: Tabulate SD of standardized residuals for each time series.

Reason: To see if the tuning had been effective for the length and age data (and to see to what extent the residuals of the biomass indices were consistent with the input variance assumptions).

Response: This was not completed – some clarification on what was required was sought by the STAT.

F: Compare age data across states.

Reason: To determine if there was substantial variation in the age frequencies between states.

Response: This was not done due to time constraints.

G: Do likelihood profiles over R_0 – split into components down to stock-recruit component – tabulate current depletion and five year projections.

Reason: To ascertain which components of the total likelihood were influential in the determination of the base model estimate of biomass and where the "preference" of each component lay.

Response: A comprehensive table of likelihoods was produced over a range of R_0 values which resulted in a 2007 depletion range which was monotonic from 13% to 48%. The largest contributions to the total likelihood came from the age and length data (as would be expected). However, the greatest contrast in likelihood components were seen in the triennial survey abundance time series, the fishery and discard length data, the fishery and discard age data, the NWFSC slope age data, and a component labelled "catch". Two of the indices favored low R_0, two favored R_0 near the base estimate, and one favored high R_0 (the fishery and discard age data).

Discussion/conclusion: The components with the greatest contrast in likelihood are the ones which are potentially most "influential" in determining the base model estimate of R_0 (and hence depletion). It was concluded that there was some "tension" between the data sets, as is typical in many assessments, and that a change in their relative weightings could substantially alter the assessment results. It was not understood by meeting participants what the component labelled "catch" referred to. It was hypothesized that it must relate to some fitting procedure involving the input landings. See Request J below.

H: Sensitivity runs to base using sigmaR = 0.6 and 1.5 – tabulate output sigmaR.

Reason: To ascertain if there was any consistency in the output sigmaR, or if it was primarily dependent on the initial input value.

Response/conclusion: Input values of 0.6, 0.8, 1.0, and 1.5 gave corresponding output values of 0.74, 0.77, 0.83, and 1.72. It was concluded that there was some stability (except for very large values of sigmaR).

I: Sensitivity run to base with no fishery conditional age-at-length data.

Reason: To determine if the exclusion of the fishery age-at-length data made any substantive difference to the assessment results.

Response: The biomass trajectory for the sensitivity run reached a lower minimum level than the base model and had a depletion in 2007 of 13% compared to the base model of 23%.

Discussion/conclusion: It was concluded that this was a substantial difference and that the meeting participants would have to carefully consider how best to use the fishery age-at-length data.

Round 2 Requests

J: Determine what the likelihood component labelled "catch" is.

Reason: This likelihood component was potentially influential and nobody was sure where it was coming from.

Response: Investigation revealed that the landings data were being fitted because the option for continuous F (as opposed to Pope's approximation) had been selected by the STAT for the catch equation (i.e., Baranov catch equation). The STAT changed the control file to select Pope's approximation and reran the likelihood profiles on R_0. The results from the altered model were almost identical to the original profiles but the range of depletion increased somewhat as did the "preferences" of the other likelihood components (i.e., the R_0 at which they were minimized). The result was some easing of the "tension" between the data sets.

Discussion/conclusion: The option to use continuous F was a relatively recent innovation in SS2 and was not documented in the technical documentation. For this assessment it appears that the results are not affected by the choice of catch equation. However, the strong preference shown by the "catch" likelihood component was disconcerting as were the relatively large differences between the specified landings and the (model) predicted landings. It is not necessary to fit landings in order to implement the Baranov catch equation — total catch can be specified as a function of landings, discard rate, and discard mortality, and F calculated using an iterative procedure. This option should be offered as an alternative to that of fitting the landings.

K: Explore alternative treatments of the fishery conditional age-at-length data. In particular, enlarging of the smaller length bins, or truncation of the smaller length bins. Compare biomass trajectories.

Reason: The conditional age-at-length data appeared to contain quite a lot of data for fish between 20 cm and 30 cm, at which stage in their life cycle they average perhaps 3 cm of growth per year. Because sampling occurs throughout the year a basic assumption of the assumed multinomial distribution is violated to some extent (i.e., the proportions of age at given length are not constant). This issue is similar to better known problems in specifying length and size bins for fishery age-length keys. Two alternative treatments of the data were suggested, both of which mitigate the problem to some extent.

Response: Expanded length bins for smaller fish of 5 cm, tapering down to 2 cm, were used in one run. In the alternative run the length bins of 28 cm and less were deleted. For both runs the biomass trajectories were very similar to the base model.

Discussion/conclusion: For the base model the violation of the multinomial assumption of constant proportion at age is of no consequence.

L: As for Request D, but compare the input and output effective *n* for the *last* iteration (before achieving the base model).

Reason: To look at whether the tuning had been adequate.

Response: The average values of the effective sample sizes in the base model were similar to the output effective samples from the base model. Some differences in trend were seen, as in Request D.

Discussion/conclusion: Given similar average values, it appeared that the tuning had been adequate.

M: Complete requests E & F.

Reason: See Requests E & F above.

Response:

Request E: The SD of the standardized residuals for the biomass indices and length frequencies were tabulated. Two of the biomass time series had standard deviations substantially lower than 1. The standard deviations of the length frequencies ranged from 0.7 to 1.

Request F: Age frequencies and length frequencies for the fishery were presented graphically by year and state. No fish from California were ever seen in the plus group (30 years), but there were only two years when age data were available from all states for comparison. There were some years in which large fish were present in the California length samples.

Discussion/conclusion:

Request E: To satisfy the statistical assumptions of the model, it is necessary that standardized residuals have a SD not too different from 1. For short time series it is not unexpected to see large deviations from this expectation, and that is not a problem. Tuning appears to have been adequate for the length frequencies. The age data were not considered and it remains a topic for research as how best to jointly tune length and conditional age-at-length data.

Request F: It was concluded that although there may be an issue with spatial variation in age and unbalanced sampling, any corrective action would have no consequences for the output of the assessment. Therefore, the issue was not pursued.

Round 3 Requests

Candidate base model configuration:

- Use all existing data sets with conditional age-at-length for the fishery with expanded length bins for smaller fish.
- Estimate *q*s analytically as median unbiased.
- h = median of Dorn darkblotched prior.
- M = 0.07.
- sigmaR = 0.8.
- Tune using same procedure as in original base.

N: For all conditional age-at-length data, calculate initial effective *n* using Stewart's formula applied to the total number of trips and fish within year (rather than within length bin). Derive the effective *n* within length bins by scaling with the proportions of aged fish within length bins. Graphically compare the initial effective *n* calculated by the two methods. Run as an alternative candidate base model (fully tuned). Choose the base model.

Reason: There was concern that it was inappropriate to apply the equations for determining effective sample size for age-at-length data to the number of trips which delivered a number of fish within a given length class. Applying the equations to the total number of trips and aged fish appeared to be a reasonable alternative.

Response: The alternative method of calculating effective *n* for age-at-length data produced lower total effective *n* within each year of each time series. However, the pattern of sample sizes across length bins was very similar for both methods. The alternative candidate base model gave results almost identical to the first candidate base model. The STAT chose the alternative candidate as the base model.

Discussion/conclusion: Ian Stewart gave a brief presentation on the origins of the equations used to calculate initial effective *n* from number of samples and number of fish. The two equations (one for surveys and one for fisheries) were derived from a meta-analysis of the 2005 stock assessments. The data sources included in the study were mainly length frequencies, but there were also age frequencies and conditional age-at-length data. The Panel acknowledged the work as a good attempt to help standardize tuning of effective sample sizes in the 2007 round of stock assessments. However, there was concern that the equations developed were not appropriate for conditional age-at-length data due to the small number of such data sets included in the meta-analysis. The equations presented seemed to summarize current practice rather than estimate optimum values. There was also concern that the approach was not getting at the basic issue of assigning an appropriate level of observation error to the length and age data as a starting point for tuning (which involves the addition of extra variance as an acknowledgement of faulty model structure and compromised assumptions). Two alternatives were suggested. There were analytical options available for estimating effective sample size and a general bootstrapping approach could be used (which could be applied to length and associated conditional age-at-length data).

For this particular assessment, the Panel concluded that the results were not sensitive to effective sample sizes within the range explored.

O: Run four sensitivities to the base with low and high M (0.04, 0.10) and h (low, high – from prior).

Reason: To explore possible dimensions of uncertainty.

Response: The low and high *h* runs gave 2007 depletion ranging from 9% to 29% (base = 16%). The low and high *M* runs gave a 2007 depletion range of 4% to 50%. Two other sensitivities were also run: "no fish lengths" and "no fish lengths or ages" which gave depletion estimates of 12% and 9%, respectively. Estimated virgin spawning biomass showed little variation over all the runs (+- 10%).

Discussion/conclusion: The large sensitivity to *M* was noted compared to *h* but some of this was ascribed to the "larger range" used for *M*. An informed prior for *M* was constructed assuming a normal distribution with the mean equal to the base model *M* (0.07) and with the range between the low and high *M* taken to represent 95% of the density. New high and low values of *M* were then taken to be the mean of the lower and upper quartiles of the density (low *M* = 0.05, high *M* = 0.09). The Panel participants agreed to use *M* as a single dimension of uncertainty. This decision was revisited and *h* was added as an additional dimension of uncertainty (i.e., nine alternative states of nature were used in decision tables, being the combinations of low, high, and base values of *M* and *h*).

Later in the meeting the STAT requested reconsideration of the values of *h* to be used. The proposal was that *h* be estimated in the base model configuration using the Dorn prior and that the estimated *h* then be fixed and accepted for the base model (with low and high *h* determined as before assuming a similar variance around the point estimate). Members of the Panel argued against this approach asserting that the estimate so obtained could not be considered any more reliable than the median of the Dorn prior (despite all of the problems with the prior). The STAT agreed to think about the issues and adopt one approach or the other before proceeding to the final runs.

P: Do a retrospective analysis on the base model (4-5 years).

Reason: To check for retrospective patterns.

Response: There were no strong retrospective patterns in the biomass trajectories.

Discussion/conclusion: The Panel participants were divided on the usefulness of checking for retrospective patterns. However, because there were no strong patterns, the discussion was academic. There was agreement that a lack of retrospective pattern was not a useful diagnostic for the reliability of an estimator. However, there was disagreement on whether a retrospective pattern was an indicator of a problem with an assessment.

Q: Plot raw catch rates within strata for triennial surveys to compare distribution across years.

Reason: To see if there has been a major temporal shift in the spatial distribution of darkblotched rockfish within the survey area.

Response: The data were very noisy but there was no evidence of a general shift in distribution.

Discussion/conclusion: The area covered by the triennial survey contains a substantial, but unknown, proportion of non-trawlable ground. If there has been a substantial shift in the distribution of darkblotched rockfish and the species has a "ground preference", it is possible that a bias has been introduced into the relative abundance time series (e.g., if darkblotched densities tend to be higher on non-trawlable ground and there was a temporal shift from strata with little non-trawlable ground to strata with much more non-trawlable ground – for whatever reason). Given the presented data it was concluded that there was no obvious cause for concern.

Technical merits and deficiencies

- The use of conditional age-at-length data appears technically superior to the common practice of using dependent length and age frequencies (i.e., where the length data have been sub-sampled for age).

- The procedure used to specify initial multinomial effective sample size for tuning the model with age and length composition data has the advantage of standardization between assessments, but questions remain about its applicability and especially to conditional age-at-length data.

- GLMM diagnostics for the indices of abundance were not available for review.

- There is a problem in assuming constant proportions at age in conditional age-at-length, particularly for small fish where fishery samples are aggregated annually. The bins used to aggregate conditional age-at-length from the fishery were expanded for small sizes to accommodate rapid growth during the year while samples were collected. This procedure does not completely solve the problem.

- Conditional age data from the fishery were not scaled to account for differences in age-at-length and landings in different regions along the coast.

- Uncertainty about the catch history was not fully explored.

- Full uncertainty about model estimates was not explored as could have been done with an MCMC analysis. The asymptotic variances that were presented likely understate uncertainty in biomass, fishing mortality and other model estimates.

- Maps illustrating the spatial overlap of the various surveys, the fishery, and habitat were not available in the assessment but would have been useful in understanding and interpreting survey, fishery and other data.

Areas of Disagreement

- The STAT team and the Panel disagreed on procedures for establishing the steepness parameter *h*, which was fixed in the model and has substantial effects on model estimates and projections. All parties agreed that the model data contained little or no information on the value of *h*. The Panel advocated using the median of "Dorn's prior", calculated excluding darkblotched rockfish. The STAT team decided to estimate steepness based on model data and Dorn's prior in a preliminary model run and then fix steepness at the estimate for final runs. The STAT felt that the estimation procedure provided a better fit to trends in the survey data not necessarily reflected in the log likelihood.

Unresolved Problems and Major Uncertainties

- As in other West Coast groundfish assessments, there is considerable uncertainty associated with fixed and estimated parameters including natural mortality and steepness.

- Use of the triennial survey as an index of abundance for darkblotched rockfish was questioned because rocky habitats used by rockfish are not well sampled by trawl gear.

Concerns raised by GMT and GAP representatives during the meeting

The GAP and GMT representatives raised no major issues of concern during the meeting.

Research recommendations

<u>For the next assessment</u>

- GLMM survey index swept area biomass data for the NWFSC shelf and slope surveys were much higher than simple swept area biomass calculations. Although some differences might be expected, the magnitude and consistency of the differences was surprising. GLMM procedures and models used to standardize the survey data should be checked and differences should be explained.

- Assessment data and background information should be presented clearly and completely before dealing with assessment models and modelling results. Data tables should be distributed at the start of the review.

- Future assessments should include complete sets of model diagnostics for GLMM standardized abundance indices, and other types of model runs.

- Maps showing the spatial overlap of the darkblotched rockfish stock area, surveys, fishing grounds and prime habitat should be provided and considered in interpreting survey data.

General or long term

- Continued work to characterize effective sample size for length composition and, particularly, conditional age composition data is needed. For example, the procedure used to assign effective sample size initially for darkblotched rockfish was questioned in this assessment.
- A full Bayesian assessment.
- It would be useful to routinely check model estimates of survey catchability to determine if they imply implausible biomass estimates. This can be done by comparing the prior and posterior for q in a fully Bayesian assessment. Other approaches involve calculating bounds for plausible q values, comparison of model and minimum swept-area biomass estimates from trawl surveys.
- Assessment and review work would have been enhanced if the STAT had consisted of more than one person and if more time had been available to carry out the assessment.