Pacific Coast Groundfish Stock Assessment Review (STAR) Panel Report

For Data-Moderate Assessments

April 22-26, 2013
NMFS, Southwest Fisheries Science Center
110 Shaffer Road
Santa Cruz, CA 95060

Methodology Review Panel Members:
Martin Dorn (Chair), Scientific and Statistical Committee (SSC), NMFS, AFSC
Vivian Haist, Center for Independent Experts
Selina Heppell, SSC, Oregon State University
Yan Jiao, Center for Independent Experts
André Punt, SSC, University of Washington

Panel advisors:
John DeVore, PFMC Staff
John Budrick, Groundfish Management Team
Gerry Richter, Groundfish Advisory Panel

STAT:

Edward Dick, NMFS, SWFSC
Alec MacCall, NMFS, SWFSC
Melissa Monk, NMFS, SWFSC
Braden Soper, NMFS, SWFSC
Jason Cope, NMFS, NWFSC
Chantel Wetzel, NWFSC, University of Washington

1. OVERVIEW

A review of data-moderate assessments was conducted by a STAR Review Panel (Panel) at the Southwest Fisheries Science Center, Santa Cruz Laboratory, during 22-26 April 2013. The review panel consisted of three SSC members and two CIE reviewers, and was advised by representatives from the GAP, the GMT and Council staff. The Panel followed the Terms of Reference for the Groundfish and CPS Stock Assessment and Review Process (November 2012). Dr. John Field welcomed the participants on behalf of the Southwest Fisheries Science Center. Dr. Martin Dorn, the Panel chair, reviewed the terms of reference, and clarified the role of STAR panel members, advisors to STAR panel, and the STAT. The Panel was provided extensive background material, including a number of primary documents, through an FTP site, two weeks prior to the review meeting. The STAT gave a number of presentations to the Panel during the meeting, and responded to Panel requests for additional information.

This was the first STAR panel of data-moderate assessments organized by the Pacific Fishery Management Council (PFMC). An overview by Dr. James Hastie described the historical development of the data-moderate assessment methods as a management tool. The need for robust assessment methods that are intermediate between catch-only methods and full age-structured models (e.g., Stock Synthesis) has been recognized for some time. The first workshop that considered data-moderate methods (along with data-poor methods) was held in April 2011. None of the proposed data-moderate methods were endorsed, because they were not sufficiently developed at that time to be used to provide management advice. In June 2012, a methodology review panel was held to review further progress on developing data-moderate assessments suitable for stocks with the exploitation history and biological characteristics of those managed by the PFMC. The 2012 Panel concluded that two data-moderate assessment methods, XDB-SRA (Extended Depletion-Based Stock Reduction Analysis) and exSSS (Extended Simple Stock Synthesis), were sufficiently well developed to form the basis for data-moderate assessments in the next assessment cycle. A comparison of data-moderate assessments results with outputs from full assessments suggested that data-moderate methods can provide improved results over data-poor approaches, such as DB-SRA and SSS.

The STAT provided a draft document with data-moderate assessments for nine stocks. The stocks included a group of nearshore rockfish, including brown rockfish, china rockfish, copper rockfish, and vermillion rockfish, and a shelf-slope group, including English sole, rex sole, sharpchin rockfish, yellowtail rockfish, and stripetail rockfish. Two of these stocks, English sole and yellowtail rockfish, had been previously assessed with full assessments, but the assessments were considered no longer current under NMFS guidelines. Given limited capacity to conduct and review full assessments, it was considered unlikely that these stocks would be prioritized for full assessment in the immediate future. The selection of stocks was guided by several principles, including likelihood of a successful assessment based on an initial evaluation of data availability, and the intent to look at different stocks with different data sets available for assessment in this first data-moderate review panel to explore the utility of the approach.

The first major task of the Panel was to review the data inputs for the assessments. For data-moderate assessments, data inputs consist only of historical catches and abundance indices. Methods to derive historical catches were similar to those used previously for data-poor assessments. While historical catch reconstructions have been used extensively, the Panel notes that historical catch estimates are highly uncertain, and there has been regrettably little effort to evaluate the magnitude of that uncertainty. Catch reconstructions for both commercial and recreational fisheries have also not been completed for all states. Consequently it was difficult for this Panel to evaluate the effects of uncertainty in catch reconstruction on assessment results. This concern applies to all assessments that use these data, but is particularly acute for data-poor and data-moderate approaches in which historical catch is a major determinant of assessment results.

Two types of abundance trend indices were used for data-moderate assessments: fishery independent indices from trawl surveys, and fishery dependent indices based on catch per unit effort (CPUE) in the recreational fishery. In March 2013, the Scientific and Statistical Committee reviewed methods for developing abundance indices for data-moderate assessments, endorsed some methods and made recommendations for improvement. Generalized Linear Mixed Models (GLMM) were used with the trawl survey data to obtain an abundance index. Initial GLMM runs included an extreme catch event (ECE) component to the model. The STAT noted there were no suitable model selection criteria for use to evaluate the benefits of including the ECE component in a GLMM model, beyond visual interpretation of model residuals. The Panel was concerned about this lack of model selection criteria, and the introduction of a new feature without a greater understanding of how to identify when it should be implemented, and consequently recommended that the GLMM models exclude the ECE component.

The Panel also reevaluated the now-standard approach of splitting the triennial time series into early and late periods. The issue is particularly important for data-moderate assessments because indices are the only source of information on stock trends in the assessment. Adding a break greatly reduces the inferences that can be made about long-term trends in abundance. After evaluating model fits with and without the triennial break, the Panel recommended base models that included the split in the triennial survey for English sole, rex sole, and sharpchin rockfish, a base model with a single triennial time series for yellowtail rockfish.

The STAT applied the same approach used previously to construct the RecFIN trip-based CPUE indices by filtering trips using the Stephens-MacCall (2004) method, followed by the use of a delta-GLM analysis. The Panel endorsed this approach. The Panel reviewed and approved an index derived from an expanded (both spatially and temporally) CPFV data set obtained by combining historical CDFW observer CPFV data with more recent data from ongoing CDFW and ODFW observer programs, and data from a CalPoly sampling program. The Panel recommends continuation of efforts to obtain drift-specific data from the earlier California program to allow easier merging of data sets, and exploring improved methods of filtering data using habitat maps.

The Panel developed and applied a standard set of criteria for deciding whether the assessment should be considered acceptable. Generally these criteria pertain to model convergence, goodness of fit, sensitivity to assumptions, plausibility of parameter estimates, and whether the prior is updated as expected given the model inputs (see table below). For the nearshore group, the Panel evaluated both coast-wide assessments and split-area assessments. The Panel considered split-area models to be the most appropriate for nearshore species when acceptable split-area models could be developed because of the greater likelihood that these stocks would exhibit local-scale population dynamics.

The STAT provided results for most stocks for both XDB-SRA and exSSS with several algorithms (MCMC, SIR, AIS) for sampling from the joint posterior distribution. In general, results for the two methods were similar, but not identical. This is not surprising given that the methods differ in terms of population dynamics as well as how priors for stock productivity are specified, and given the uncertainty associated with the outputs from both assessment methods. The Panel made some progress in comparing the two approaches, but ultimately was unable to develop criteria to identify a preferred method for a particular stock. For pragmatic reasons, the Panel adopted a suggestion by the STAT to focus only on XDB-SRA model runs for the nearshore group, and on exSSS model runs for the shelf-slope group.

For the nearshore group, the Panel recommended a coastwide base model for brown rockfish, and split-area base models for china and copper rockfish. For the shelf group, the Panel recommended base models for English sole, rex sole, yellowtail rockfish, and sharpchin rockfish. For stripetail rockfish, the Panel concluded that the stock was above the $B_{40\%}$ target level, but did not endorse a base model for setting the OFL.

Due to a lack of time the Panel was unable to review the draft assessments of vermillion rockfish and yellowtail rockfish south of Cape Mendocino, and cannot make recommendations regarding their use for Council decision-making.

The Chair thanked the SWFSC for hosting the meeting, acknowledged the assistance of SWFSC in providing a meeting room and helping with meeting logistics, and thanked the participants for spirited discussions and a constructive atmosphere during the review, the results of which should help inform the Council and its advisory bodies determine the best available science for the assessment of groundfish.

2. Data Inputs

2.1 Catch data

The catch time-series for the nine species were constructed in essentially the same manner as was the case when DB-SRA was applied during the 2011 assessment cycle to calculate OFLs for Category 3 species. PacFIN was used to determine the commercial landings north of Cape Mendocino for 1981-2012 while CalCOM was used to determine commercial landings to the south of this for 1969-2012. The catches north and south of Point Conception were based on splitting catches by port between Santa Barbara and Morro Bay. The recent (1980-89; 1993-2012) recreational catches were taken from RecFIN (type A+B1). It was noted that recreational catches for Oregon and Washington in RecFIN are based on State sampling programs. However, it was not clear to the Panel whether some of the earlier catches for Oregon (pre-mid-1990s) were still based on using the results from telephone surveys to determine effort. The historical foreign catches were based on the reconstruction by Jean Rogers, while bycatch in the at-sea whiting fishery was extracted from NORPAC (1992-2012). The pre-1969 commercial and pre-1980 recreational catches off California were based on the reconstruction by Ralston *et al.* (2010), while the pre-1981 commercial catches off Oregon were based on the unpublished reconstruction by V. Gertseva. A variety of data sources (e.g., Tagart [1985], the PMFC data series, and the Pacific Fisherman Yearbooks) were used to reconstruct the Washington landings. The discards in the commercial fishery were based on WGCOP and (where available) data from the Pikitch discard study. E.J. Dick noted that removals (including discards) for English sole were taken from the last assessment update.

The Panel noted that the approach used to construct the historical removals was the best available. However, the historical catches, in common with those for most other Council-managed groundfish, are subject to considerable uncertainty. The uncertainty regarding species compositions was noted as a particularly pervasive problem. The Panel noted that the impact of this uncertainty could be addressed in the future with sensitivity tests and/or by adding errors to catches when applying exDB-SRA and exSSS (see Givens and Thompson, 1985). The Panel highlighted the importance of a data workshop to vet the catch series before assessment reviews.

A. Request: Explain how the pre-1981 recreational catches for Oregon were specified.
Rationale: The document does not include this information.
Response: This request was not completed before the end of the Panel meeting. It will be specified in the final assessment document.

B. Request: Create a set of tables (one table for each species) of the historical catches, with columns for each data source. Plot the data by source.
Rationale: The catches are only provided in plots by area, which makes evaluating the uncertainty associated with the catch data difficult. The Panel wished to understand which data sources were most influential on total removals.
Response: This request was not completed before the end of the Panel meeting. The tables will be provided in the final assessment document.

C. Request: Plot the time-series of discard rates by species.
Rationale: The Panel wished to determine how historical discard rates were created.
Response: This request was not completed before the end of the Panel meeting. The plot will be provided in the final assessment document.

2.2 Fishery-independent abundance indices

1.2.1 AFSC Triennial Shelf Survey / NWFSC Slope-Shelf Survey

The focus for the Panel evaluation of the Triennial and NWFSC survey indices was the application of the GLMM model for constructing the indices. The software on which the GLMM analyses were based was evaluated by the SSC in March 2013. The SSC endorsed the new software, and recommended that it be used for stock assessment purposes. It also recommended that documents presented to this data-moderate panel: 1) compare alternative error models (e.g. gamma vs lognormal) when developing indices of abundance using Q-Q plots, posterior predictive checks, and average deviance, and 2) test whether effort impacts the probability of a catch being non-zero. The analysts did not test whether effort impacts the probability of a catch being non-zero.

Model selection for the GLMM involved using mean deviance to select between gamma and lognormal error models, then using visual examination of Q-Q plots to select between models with and without Extreme Catch Events (ECEs) components.

D. Request: For the GLMM-based indices, show Q-Q plots comparing models which include components for ECEs and those that do not. Plot histograms of deviance for each of the four GLMM analyses for each species.
Rationale: The Panel wished to examine the basis for selecting models with ECEs, which was based on the Q-Q plots.
Response: The plots were provided. As expected, the differences in the Q-Q plots for models with and without ECE components was larger for rockfish (e.g. sharpchin) than for flatfish. However, the basis for selecting models with ECE components over those without such components using the Q-Q plots was not obvious. It was noted that the Q-Q plots for some of the "rejected" models would have been "acceptable" had they been provided for, for example, a CPUE standardization.

E. Request: Plot the distribution of the positive catches.
Rationale: The Panel wished to see whether ECEs are evident in the data, especially for species (such as English sole) which do not aggregate.
Response: The plots were provided, but they were not in log-space and so were difficult to interpret. There was some indication that schooling species like yellowtail rockfish showed greater evidence of extreme catch events than non-schooling species like English sole

F. Request: Plot the four GLMM indices (ECE vs non-ECE; lognormal and gamma), along with the design-based index, with associated confidence intervals by species.
Rationale: The Panel wished to more fully understand how different the outcomes from the GLMM are given various assumptions regarding inclusion (or not) of model components for ECEs, as well as the choice of error model.
Response: The plots were provided. In general, the log-normal (no ECE component) result differed the most from the remaining indices.

G. Request: Create time-series of GLMM-based indices for the Triennial survey which (a) include the data for 1977, and (b) analyzes all of the data to form a single time-series.
Rationale: The SWFSC analysts include the Triennial data in assessments as single time-series in their preliminary results, but used the design-based estimates rather than the GLMM estimates which made comparisons very difficult.
Response: The time-series were created. It was noted that the GLMM indices for the entire time-series mimicked those when the data were analyzed separately for 1986-1992 and 1995 onwards for some species (English sole and rex sole), but differed quite markedly for others (e.g. yellowtail north).

In discussion, the Panel agreed that the method proposed for model selection is not ideal because there is no objective basis to select between models with and without ECE components using Q-Q plots, and that the use of average deviance ignores differences in numbers of effective parameters among models. Although the STAT followed the recommendations of the SSC, this problem of model selection has yet to be fully resolved. It therefore recommended that the model selection criteria be a topic for future research. DIC (Deviance Information Criterion) was identified as a possible approach for model selection. The analysts noted that Eric Ward and Jim

Thorson (developers of the GLMM software) do not support the use of DIC for model selection. However, no documentation of the reasons for this was available to the Panel.

Although the Panel is generally supportive of the ECE approach as a potentially useful improvement in GLMM models for survey data, the Panel is concerned about the lack of model selection criteria, and introduction of a new feature to the GLMM models without a greater understanding of how to identify when it should be implemented. For these reasons, the Panel recommended limiting consideration to GLMM models which exclude ECE components.

2.2.2 *Southern California Hook and Line Surveys*

The Panel noted the concerns by the SSC that effort (in terms of *Fishing Time*) in the model is determined by how quickly the hooks are being occupied rather than being treated as an independent variable, meaning that the catch influences *Fishing Time* instead of the other way around. The SSC therefore concluded that including *Fishing Time* in the model was likely inappropriate. The SSC recommended exploring versions of the statistical model that do not use *Fishing Time* as a covariate to examine the sensitivity of the abundance index to the assumption that hook saturation has no important effect. They also identified other analytical approaches that can be used to evaluate the importance of a gear saturation effect, such as dropping the data from a particular hook location as a cross-validation exercise and treating the Y-variable as the number of the 5-hooks per line that caught fish. The abundance index from the Hook and Line Survey included in the assessment for vermillion rockfish ignored the *Fishing Time* covariate as recommended by the SSC.

2.3 Fishery-dependent abundance indices

2.3.1 RecFIN trip-based CPUE

The analysts applied the same approach (Stephens-MacCall [2004] filtering of trips, followed by the use of a Delta-GLM analysis approach) to construct the RecFIN trip-based CPUE indices. The Panel endorsed this approach.

2.3.2 Observer-based Recreational CPUEs from CPFVs

Data are available from four primary data sources (historical observer CPFV data from CDFW, data from recent CDFW and ODFW observer programs, and data from a CalPoly sampling program). The data from the first of these programs is by “fishing site” whereas the data from the latter three sources are available by drift. The analysts aimed to analyze these data to construct indices for southern California, central California and northern California/Oregon. The basic approach involved adding “buffers” about each point with a positive catch from the recent data to define regions, and combining regions which overlap to create “sites”. The data which will be potentially used in subsequent analysis are then the positive catches, along with the zeros which are located in the areas which define the sites. This data set is then restricted in several ways: a) a site needed to be visited at least five times, b) start or ending lat/longs should be available (CDFW and Calpoly), c) there should be no erroneous location or time data (CDFW and Calpoly), d) drifts should not be in San Francisco or San Diego Bays, e) at least 30% of or more of the observed catch should be groundfish (CDFW only), f) the sites should be within 60 fathoms (southern California) or 40 fathoms (Orgeon), and g) the sites should not be not in Conservation Areas. The data were further restricted to exclude records outside the 2.5% and 97.5% quantiles of fishing times and 2.5% and 97.5% of anglers.

The Panel noted that the approach was highly innovative, and attempted to make best use of the available data. However, it is a very complicated and computationally intensive approach, which makes evaluating it difficult. The Panel identified several requests to further explore the sensitivity of outcomes from the approach to some of its specifications.

H. Request: Compare the standardized indices from the proposed method with indices constructed by applying the Stephens-MacCall approach to the data aggregated by trip.
Rationale: The Stephens-MacCall approach has been the standard way to analyze the recreational CPUE data.
Response: There was insufficient time to conduct this analysis during the meeting.

I. Request: Repeat the analysis of the CPFV data for brown rockfish in the central California region based on the final model, changing the buffer to 0.03.
Rationale: The Panel wished to understand how sensitive the outcomes from the proposed method are to changing its assumptions. The size of the buffer determines how many zeros will be included in the analysis, making it an appropriate factor to vary. Brown rockfish in the central California region was selected because this is a species the index for which is based on the historical as well as the recent observer data.
Response: The results were insensitive to the changing the buffer.

J. Request: Provide the graphs for the RecFIN indices for brown rockfish.
Rationale: These graphs were omitted from the document provided to the Panel.
Response: The graphs were provided when the results for brown rockfish were shown to the Panel.

3. Modeling Issues (Technical merits and deficiencies)

3.1 Criteria for evaluating assessment adequacy

Review of model performance focused on model convergence, goodness of fit, sensitivity to assumptions, and plausibility of parameter estimates and whether the prior is updated as expected given the model inputs. The Panel discussed and agreed upon the following set of performance criteria for evaluating the adequacy of data-moderate assessments.

Performance criteria for evaluating data moderate assessments
1. Do the diagnostics for the posterior sampling algorithm indicate that the model has converged?
2. Are the indices used in the assessment sufficiently precise to provide a signal for the assessment?
3. Is an adequate fit achieved to indices of abundance used in the assessment?
4. Does the model capture the evident trends in the abundance indices, or, if not, can the residual pattern be explained by model's inability to account for increases in recruitment? This would only occur when the index is trending more strongly upwards than the model predictions.
5. Do sensitivity analysis indicate the results are robust to uncertain model assumptions?
6. In comparison to catch-only assessments (SSS and DB-SRA), does the addition of index data update the prior distributions in a sensible way, rather than giving strongly divergent results?
7. Is the updating of the distribution of key parameters from prior to posterior reasonable given the likely information content of the indices? For example, a posterior distribution of B_{MSY}/B_0 that is very different than the prior distribution could be a concern because the data are not likely to be very informative about this quantity.
8. Are the estimates of catchability for survey indices within reasonable limits ($0.1<q<3$) for assessments that use survey indices?
9. In cases where a previous assessment has been accepted for the stock, are results reasonably consistent with the previous assessment?

Evaluation of performance criteria was necessarily limited to the collective opinion of the STAT and Panel members, but usually it was straightforward to rank fits for alternative model runs from good to bad. It was more difficult, however, to decide on a threshold to distinguish between adequate and inadequate model fits. The Panel found the checklist approach useful, but acknowledges that these criteria are in need of refinement, and notes that quantitative metrics would assist future Panels in reviewing data-moderate assessments. The use of more objective thresholds for evaluating model performance would also help to assure consistency across assessments. However, even full assessments occasionally display consistent patterns of lack of fit and other assorted problems, and it is important not to apply more stringent criteria to data-moderate assessments than are applied to full assessments.

3.2 Comparison of modeling approaches

One goal of this meeting was to compare outputs from XDB-SRA and exSSS for the same species and provide guidance on their applicability to different stocks. The Panel was able to do this in a general way, but ultimately it was not possible to conduct a formal model comparison for all species due to time constraints and differences in input data in the original model runs (e.g. whether the analysis used the 1977 data point for the triennial survey, split or do not split the triennial index).

In general, XDB-SRA and exSSS gave similar results when input data were the same and, for some stocks, when priors on key parameters were constrained to force the models to behave in similar ways. Both models were considered acceptable for use. The analysts were generally able to sample parameter vectors from the posterior distribution. The Sample-Importance-Resample (SIR) algorithm performed adequately for some stocks, but Adaptive Importance Sampling was able to deal with situations when SIR was inefficient as the posterior differed markedly from the prior. For exSSS, the MLE estimates were not considered useful for providing management advice since the model has a Bayesian formulation, but were useful to rapid exploration of the sensitivity of runs to assumptions.

Due to time constraints, it was not possible to re-run all models with standardized input data for a formal model comparison. In particular, exSSS can take several hours for each model run. To ensure that completed model runs and analyses were available for each species, the STAT proposed and the Panel agreed to apply XDB-SRA for the nearshore species (copper, brown, and china rockfishes, combined and split by region) and apply exSSS to the offshore species with fishery independent data (English and rex sole, yellowfin and sharpchin rockfish).

For the rockfish assessments, exSSS repeatedly led to posteriors for F_{MSY}/M which had most of their mass at far higher values than past assessment models have attributed to rockfishes. For age-structured assessments with a Beverton-Holt stock recruit relationship, a high F_{MSY}/M ratio occurs when the estimated steepness is close to 1.0, but also depends on fishery selectivity and the maturation curve. Maturity and selectivity are assumed to be the same in exSSS which might explain this behavior[1], along with the fact steepness was estimated to be high for stocks with rapidly increasing abundance. Nevertheless, the Panel remains puzzled about the high values for F_{MSY}/M which are given support in the posterior and recommends this be explored further.

XDB-SRA, which has greater flexibility in how productivity changes with stock size, was less likely to assign considerable mass to high F_{MSY}/M ratios than exSSS (note that XDB-SRA imposed a prior on F_{MSY}/M where F_{MSY}/M is a derived quantity of exSSS). XDB-SRA assumes that the priors for F_{MSY}/M and B_{MSY}/B_0 are independent, but the priors are based, at least to some extent, on meta-analysis of stock assessments where these quantities are likely to be strongly correlated. It would not be too difficult to design a simulation experiment to evaluate correlation in F_{MSY}/M and B_{MSY}/B_0 in an age-structured population with more flexible stock-recruit relationships.

When the base models for exSSS and XDB-SRA were compared for the same species, it was clear that some of the differences in model results were due to differences in how the priors for stock productivity were specified. Modifying priors to make the productivity-linked parameters of each model more comparable led to similar model behavior and outputs. Further consideration is needed to specify priors for stock productivity that are appropriate for the life history and biology of the species being assessed. Additional analyses are also needed to compare the performance of exSSS and XDB-SRA; it may be that species life history or catch history can be used to choose a model *a priori* for future assessments, and that the development of a single, "best" approach to modeling will not be possible.

[1] Review of the code for stock synthesis suggests that these were actually not the same.

Chantel Wetzel presented preliminary results of a Management Strategy Evaluation (MSE) to compare the performance of harvest control rules based on data-poor and data-moderate assessment methods for rockfish-like and flatfish-like life histories[2]. Rapid generation of replicate simulations for exSSS is not currently feasible computationally, precluding a full evaluation of this method. The MSE approach is valuable for determining the performance of different models and management strategies under conditions of varying population dynamics and data quality, and will provide future guidance for recommending which assessment methods(s) should be applied to different stocks.

K. Request: For the brown rockfish coastwide assessment, re-weight each point in the posterior by the ratio of the prior probability for F_{MSY}/M for exSSS to that for XDB-SRA and constrain the F_{MSY}/M – B_{MSY}/B_0 point to the relationship between F_{MSY}/M and B_{MSY}/B_0 underlying XDB-SRA.
Rationale: Eliminate the model structure as an issue in the comparison
Response: The marginal posterior for F_{MSY}/M from this version of XDB-SRA was closer to that from exSSS. However, full results from this comparison were not available.

L. Request: Set maturity at age 1 for brown rockfish for XDB-SRA and compare to base run where age at maturity is age 4.
Rationale: To better understand the differences between XDB-SRA and exSSS. One would expect that a shorter time would reduce the proportion of rejected samples from the prior.
Response: There was little change in biomass trends when the time lag was changed. As expected, the stock depletion can go slightly lower when age at maturity is set at age one. This is because the number of pre-model rejections due to the stock going below zero is lower when the time-lag is lower (such rejections are not possible when there is no time lag.)

M. Request: Use knife-edge maturity and selectivity at age 4 for brown rockfish exSSS and compare F_{MSY}/M for both exSSS and XDBSRA.
Rationale: To better understand the differences between XDB-SRA and exSSS.
Response: The pre-data and post-model distribution for F_{MSY}/M were nearly identical. This doesn't seem to be accounting for the difference.

N. Request: Provide tables for all west coast rockfish and all west coast flatfish with the following information from the most recent stock assessment/update: a)year of the assessment/update, b) value of steepness estimated/used in the base model, c) F_{MSY}, d) %SPR at F_{MSY}, e) natural mortality (use female if different than male, use average for mature fish if age-specific), f) F_{MSY}/M (c/e).

Example	Year	Steepness	F_{MSY}	%SPR@MSY	M	F_{MSY}/M
Dover Sole	2011	0.8	0.131	0.291	0.117	1.1

Rationale: The panel wanted to learn if the high F_{MSY}/M ratios from exSSS were consistent with outcomes from other west coast assessments.
Response. The table was provided. Most estimates of F_{MSY}/M range in 0.8 to 1.0. Assumptions for steepness are variable (but are not higher than 0.8). No very high estimates of F_{MSY}/M were found even when the Beverton-Holt stock recruit relationship was assumed. There were quite a few missing values for F_{MSY}.

[2] Due to his involvement in this work, Dr. André Punt recused himself as a Panel member for this discussion.

3.3 Characterization of uncertainty

XDB-SRA and exSSS are fully Bayesian approaches. Given the number of data-moderate assessments conducted, the Panel recommends that a simple and standardized approach be used to characterize uncertainty for constructing decision tables. The Panel recommends that the posterior joint distribution for the final base model be used to evaluate uncertainty for all data-moderate assessments. To construct a decision table, the joint posterior distribution should be split using the 25% and 75% quantiles of ending year stock depletion to provide low, base, and high biomass states of nature. Stock projections should be provided using recent average catch, the ACL in the last specifications cycle, and the 40-10 rule. The GMT may provide additional scenarios for analysis. Analysts should check if the estimated standard deviation in ending biomass (sigma) is greater than the default for category 2 assessments (0.72), and if so the larger value should be used in the calculation of the P* buffer between the OFL and the ABC.

4. Assessments based on fishery-dependent indices

4.1 Brown Rockfish

The catch of brown rockfish is primarily from central area (Point Conception to Cape Mendocino) (80%), with the northern area contributing about 1% of the catch, and the southern area contributing about 19% of the catch. Four abundance indices were included in the assessment models, all of which are derived from recreational fishery data. The indices used are the central and southern California onboard CPFV indices, and the central and southern California RecFIN indices.

Results were presented for both models using several algorithms for sampling from the posterior distribution. For exSSS, results were shown for the MLE estimates, and for posteriors constructed using the AIS and MCMC algorithms, while for XDB-SRA results from the AIS algorithm were presented. For pragmatic reasons, the Panel focused on XDB-SRA for brown rockfish, as there was insufficient time to adequately review both models. The Panel compared a coastwide model with separate models for north and south of Point Conception. The southern area model showed a better fit to the indices, but there was strong updating of the prior for the productivity parameters relative to the central area model, which was considered implausible by the Panel. There is a concern that model results were being strongly driven by recent increases in indices that may be temporary, rather than reflecting true stock productivity. The coastwide model showed reasonable fits to upward trending indices, and more plausible posterior distributions for the productivity parameters than the split-area models. Overall the model was considered acceptable. The Panel recommended that the coastwide XDB-SRA model run be used as the base model, but that the OFL be apportioned spatially using the split-area assessments, since they were considered to be informative about relative abundance north and south of Point Conception.

O. Request: Develop separate XDB-SRA assessments for brown rockfish for the southern and central areas by splitting the RecFIN CPUE time series appropriately.
Rationale: The Panel was interested in learning whether split-area models for north and south of Point Conception would provide an improved characterization of population trends.
Response: Results for the requested models were presented to the Panel. The southern area model was considered marginally inadequate due to strong (and unrealistic updating) of productivity priors.

4.2 China rockfish

China rockfish is a nearshore species with catches primarily from north of Point Conception. Four abundance indices were used, all of which were derived from recreational fishery data: central California and Oregon onboard CPFV indices and coastwide RecFIN indices (separated into areas for the split-area assessments). The STAT developed XDB-SRA and exSSS models for China rockfish. As with other nearshore stocks, the Panel focused on the XDB-SRA model for China rockfish, as there was insufficient time to adequately review both models. For the XDB-SRA model, separate assessments for north and south of Cape Mendocino were presented in addition to a coastwide model.

The coastwide model for China rockfish indicated a highly uncertain biomass estimate due to evidence of different fishing effort and recreational catch trends north and south, leading to marginally acceptable fits to the index data. Sensitivity runs for the northern area assessment indicated that estimates of stock status and the OFL are reasonably robust. Sensitivity runs for the southern area assessment indicated that estimates of depletion and current fishing mortality were robust, but the overall scale of the population is relatively poorly determined. The Panel determined that the northern and southern area models satisfied the criteria for acceptability. While the coastwide model was also considered acceptable by the Panel, the split-area models fit the indices slightly better, had plausible posterior parameter distributions, and there were substantially different trends and exploitation rates by area. Therefore the Panel recommended these models as the base models for China rockfish.

4.3 Copper Rockfish

The catch of copper rockfish is mainly from south of Cape Mendocino, with only 4% of reported landings from north of Cape Mendocino. Four abundance indices were used, all of which are derived from recreational fishery data: central and southern California onboard CPFV indices, the Oregon onboard CPFV index and the coastwide RecFIN index (separated into areas for the split-area assessments). The STAT developed XDB-SRA and exSSS models for copper rockfish. As with other nearshore stocks, the Panel focused on the XDB-SRA model for copper rockfish, as there was insufficient time to adequately review both models. For the XDB-SRA model, separate assessments for north and south of Point Conception were presented in addition to a coastwide model.

The coastwide model showed marginally acceptable convergence statistics, but they were not so poor as to reject the model. The convergence statistics for the split-area models were satisfactory. Other criteria indicated adequate performance of the split-area models, although the overall scale of the population is very sensitive to which indices were included in the models. The split-area models showed evidence of differing exploitation trends spatially, which is an important feature to capture in the assessment. Because the split-area models were considered acceptable, the Panel recommended these models as the base model for copper rockfish.

P. Request: For copper rockfish, include the Oregon index in the northern XDB-SRA model and compare the differences with and without it.

Rationale: Because Oregon is included in the assessment area, the Panel wanted to evaluate the impact of including the Oregon index, even though the fraction of the total catch from Oregon is low.
Response: There were very minor changes in model results. The Panel concluded that it was appropriate to include this index in the assessment.

5. Assessments based on fishery-independent indices

5.1 English sole

English sole is assessed as a coastwide resource. This shelf species, which is caught almost exclusively in trawl fisheries, has a long history of commercial exploitation, with peak catches (catch plus discard) occasionally exceeding 4,000 t over the period 1920 to 1980. Catches have steadily declined since 1980. English sole are well represented in the trawl surveys, occurring in approximately 65% and 40% of Triennial and NWFSC tows, respectively. A full stock assessment of English sole was previously conducted in 2007.

The Panel reviewed results from the exSSS (AIS) and XDB-SRA assessments, and concluded that both models were adequate for management. The abundance indices were considered adequate for providing a signal for the assessments, and diagnostics indicated convergence of the posterior sampling algorithms for both models. Spawning stock biomass trends were similar for the XDB-SRA and exSSS (AIS) model runs, when both models were fit to the GLMM-standardized trawl survey abundance indices. Neither model was able to fit the declining trend in the NWFSC survey indices, but the 2007 full stock assessment was also not able to fit that trend.

Both models met the criteria developed by the Panel for accepting assessments, and there was no objective basis for selecting one model over the other. For pragmatic (workload) reasons, additional model runs were limited to the exSSS (AIS) model, as there was insufficient time to adequately review alternative runs from both models.

The selected English sole base model is the exSSS (AIS) model fit to the Triennial survey GLMM abundance indices (excluding 1977) split into separate early and late series, and the NWFSC abundance indices. Model results were insensitive to fitting a single or split (early/late) Triennial survey abundance series. For the base model, trawl survey q estimates were considered plausible (all slightly greater than 1), and the posterior distribution for 2000 stock depletion was updated from the post-model-pre-data distribution (SSS). Estimates of stock depletion were insensitive to alternative assumptions, with the exception of fitting the model only to the late Triennial survey abundance indices. Median estimates of spawning stock biomass were similar between the exSSS model and the 2007 full assessment, other than the most recent years where the additional data used in the current assessment had updated stock estimates.

5.2 Rex sole

Rex sole is assessed as a coastwide resource. Rex sole is a commonly occurring species in both trawl surveys and commercial fisheries, although it has not been targeted commercially in recent years. Catches peaked in the mid-1950s through mid-1980s, with annual catches up to 2,500 t. For the reasons noted above, the exSSS was the selected model for the rex sole stock assessment.

The exSSS (AIS) base model run for rex sole fits to the Triennial survey GLMM abundance indices (excluding 1977) split into separate early and late series, and the NWFSC abundance indices. This formulation was selected over using a single Triennial survey abundance series because that fit resulted in implausible catchability estimates (*q*s >5). With the split Triennial survey indices, catchability estimates for the NWFSC survey was still quite high (approximately 3). The models are incapable of fitting the relatively flat trend in the NWFSC survey indices, given the reductions in catch that have occurred.

For the base model run, the fits to the two Triennial survey abundance indices were very good, but the fits to the NWFSC survey indices were poor. The Triennial survey as a single time series is relatively informative, but much of the value of the time series is lost when split into two series. The posterior distributions were not updated much from their prior distributions. Stock depletion estimates were relatively insensitive to model assumptions and alternative data series fit in the model, but scale parameters (abundance estimates) were highly sensitive to the alternative assumptions.

The Panel concluded that the base model provides an adequate basis for management, but notes that inability to fit the NWFSC survey index implies some model mis-specification. There is considerably more confidence in stock status estimates than in the biomass scale.

Q. Request: For rex sole, run exSSS using a single triennial CPUE index (excluding 1977).
Rationale: This will be the base model run (assuming no surprises, relative to the equivalent MLE run).
Response: The STAT presented the results from the model run. There were implausible catchability estimates (*q*s >5) for this model. This led the Panel to reject this run as a potential base model.

5.3 Yellowtail Rockfish

The yellowtail rockfish assessment was conducted for a northern stock (north of Cape Mendocino). There was not adequate time to review the assessment for the southern yellowtail rockfish stock. Yellowtail rockfish are common in both commercial and recreational fisheries, although in the northern area they are taken predominantly in commercial fisheries and peak catches (catch plus discard) have exceeded 9,000 mt/year. A full assessment of northern yellowtail rockfish was conducted in 2004 (Wallace and Lai 2005).

The Panel reviewed results from XDB-SRA and exSSS, and determined that both were acceptable for providing management advice. For pragmatic (workload) reasons, the Panel focused on the exSSS run. Initial model runs fit to recreational and survey abundance indices, but the recreational fishery targets juveniles so it was decided to exclude the recreational indices from the assessment.

The exSSS (AIS) base model was fit to a single Triennial survey abundance series (1977 excluded) and the NWFSC survey abundance series. The base model fit the abundance index data reasonably well, with no patterns in residuals and minimal additional variance for the NWFSC survey indices, and moderate additional variance for the Triennial survey indices. The was some update of the 2000 depletion prior distribution relative to the post-model-pre-data

(SSS) distribution, but other model parameters were not updated. Trawl survey catchability parameters were low, considered plausible for this species due to its mid-water distributions (median *q*s of approximately 0.2 for the Triennial survey and 0.4 for the NWFSC survey). Biomass estimates were highly uncertain. Stock biomass trends were similar between the exSSS base model run and the 2004 full stock assessment.

The Panel concluded that the base model was adequate for management of yellowtail rockfish north of Cape Mendocino.

S. Request: For yellowtail rockfish (north), conduct a XDB-SRA run that excludes the fishery-dependent indices (triennial series w/o 1977).
Rationale: This run investigated sensitivity to inclusion of this CPUE index because the recreational fishery targets smaller fish. Other fishery-dependent indices were not updated for this assessment and are of limited value in monitoring recent abundance trends.
Response: The STAR and the STAT agreed to focus on exSSS runs for yellowtail rockfish for pragmatic reasons, and did not consider XDB-SRA runs further.

T. Request: For yellowtail rockfish (north), conduct an exSSS (AIS) run that excludes all fishery-dependent CPUE indices (triennial survey as single series w/o 1977).
Rationale: Investigate sensitivity to inclusion of this CPUE index because the recreational fishery targets smaller fish. Other fishery-dependent indices were not updated for this assessment and are of limited value in monitoring recent abundance trends.
Response: The biomass trends were broadly similar.

5.4 Sharpchin Rockfish

Sharpchin rockfish was assessed as a coastwide stock. This shelf species is generally not a commercial target, although they have been taken in large numbers in trawl fisheries targeting Pacific ocean perch. Commercial catches (catch plus discard) north of Cape Mendocino peaked at about 900 t/year, with little catch taken to south of this. Catches of sharpchin rockfish catches have been negligible since 2000.

The Panel reviewed results from the XDB-SRA and exSSS, and determined that both were acceptable for providing management advice. Additional model runs were limited to exSSS (AIS) for pragmatic (workload) reasons.

The exSSS base model run for sharpchin rockfish was fit to the Triennial survey GLMM abundance indices (excluding 1977) split into separate early and late series, and the NWFSC abundance indices. This formulation was selected over using a single Triennial survey abundance series because analyses based on a single Triennial index resulted in implausible catchability estimates (Triennial survey *q* slightly over 1, and NWFSC survey *q* approximately 10). Catchability for the NWFSC was still fairly extreme (~4) with the split Triennial survey indices. The models are incapable of fitting the large increase in abundance estimated for the NWFSC survey relative to those for the Triennial survey, without extreme catchability estimates.

The base model fit the abundance index data reasonably well, with no patterns in residuals and no additional variance. The model does not follow the flat to downward trend in the NWFSC survey, instead predicting an increasing biomass trend. The posterior distributions of steepness and depletion in 2000 are only updated slightly relative to their prior distributions and the post-model-pre-data (SSS) distributions, indicating that catch trends rather than index trends are primarily driving results. Sensitivity to abundance indices was moderate, with final depletion estimates ranging from about 0.5 to 0.9 for runs that differed in the data series fit in the model.

The Panel concluded that the base model results were adequate for management purposes. Scale parameters were highly uncertain, resulting in high uncertainty in OFL values.

R. Request: Run the exSSS AIS using a single triennial CPUE index (excluding 1977).
Rationale: Results from this run will be compared with an (almost) equivalent XDB-SRA run and a base run will be selected based on the plausibility of the *q* estimates.
Response: This model resulted in implausible catchability estimates, and was not considered further.

5.5 Stripetail Rockfish

Stripetail rockfish is assessed as a coastwide stock. This shelf species is found in trawl fisheries, although they are neither a target of commercial or recreational fisheries. Reported annual catches (catch plus discard) have generally been less than 50 t. With reduced trawl fishing, catches of stripetail rockfish have been negligible since 2000. Stripetail rockfish have not been previously assessed.

The XDB-SRA model was used in a sensitivity analysis to evaluate probable levels of stock status for stripetail rockfish. A similar analysis with exSSS would have yielded similar conclusions, but there was insufficient time to complete the full set of sensitivity analyses requested by the Panel. Neither model was able to obtain credible results without an informative prior to establish the scale of the population. XDB-SRA was fit to the Triennial trawl survey and to the NWFSC survey indices. Convergence to the posterior distribution appeared good, and there was little updating of parameter values from their priors (DB-SRA) except for the depletion parameter (delta in the year 2000).

The Panel noted that stripetail rockfish is rarely caught and appears to be in an essentially unfished state, as indicated by the trawl survey abundance estimates. There is little information in the trawl survey data to estimate catchability, so abundance estimates are extremely uncertain. However, over a broad range of plausible values for trawl survey catchability (*q* range from 0.22 to 4.5), stock depletion estimates were relatively consistent, ranging from 0.75 to 0.95.

The Panel recommends that status of stripetail rockfish can be estimated based on the posterior profile of *q*, but that the extreme uncertainty in abundance estimates precludes using assessment results for setting the OFL.

U. Request: Construct a likelihood profile on log(*q*) from -1.5 to 1.5 in increments of 0.5 for the stripetail rockfish XDB-SRA model.
Rationale: To evaluate the likelihood that the stock is above the target.
Response: Most of the profile was completed including the endpoints by the close of the meeting. Estimates of stock depletion are well above $B_{40\%}$ for all runs.

6. Summary of Base Models Recommended by the Panel

Stock	Base Model
Brown rockfish	XDB-SRA coastwide model. Area models should be used for OFL apportionment north and south of Point Conception.
China rockfish	XDB-SRA area models for north and south of Cape Mendocino
Copper rockfish	XDB-SRA area models for north and south of Point Conception. Separate RECFIN indices for each area. Oregon onboard observer time series included in the northern area model.
English sole	exSSS coastwide model, Triennial survey excluding 1977, split into two time periods; NWFSC survey.
Rex sole	exSSS coastwide model, Triennial survey excluding 1977, split into two time periods; NWFSC survey
Sharpchin rockfish	exSSS coastwide model, Triennial survey excluding 1977, split into two time periods; NWFSC survey
Yellowtail rockfish north of Cape Mendocino	exSSS model, exclude all fishery dependent indices, triennial survey excluding 1977, no split; NWFSC survey
Stripetail Rockfish	XDB-SRA model with triennial and NWFSC surveys treated separately. The panel concluded that there is very high probability that the stock is above the $B_{40\%}$ target level. However the Panel did not identify a preferred base model, and recommends that assessment results not be used to set an OFL.
Vermillion rockfish	The panel was unable to review the vermillion rockfish assessment due to time constraints, and can make no recommendation.
Yellowtail rockfish south of Cape Mendocino	The panel was unable to review the assessment of yellowtail rockfish south of Cape Mendocino due to time constraints, and can make no recommendation.

7. Areas of disagreement regarding STAR panel recommendations

There were no areas of disagreement between the STAR panel and the STAT, nor among STAR panel members (including concerns raised by GMT and GAP representatives).

8. Unresolved problems and major uncertainties

The unresolved problems and major uncertainties for the data-moderate assessment methods are discussed Sections 2 and 3. Here the Panel reiterates what it considers the most important issues.

- The data-moderate assessments assume known catches, but there is considerable uncertainty in historical catch reconstructions, particularly for the recreational fishery. This uncertainty has not been measured, and tools for incorporating this uncertainty in assessments are not well developed. This is an issue for all assessments.
- The comparative performance of the two assessment approaches for data-moderate assessments, XDB-SRA and exSSS, is not yet known. In most cases, however, the two models gave similar results for most stocks, and results were highly comparable when productivity parameters were constrained.
- There are fundamental differences between XDB-SRA and exSSS in how stock productivity is modeled. For exSSS, F_{MSY} increases as the ratio of B_{MSY}/B_0 decreases in a

deterministic way, while there is no prior relationship between F_{MSY} and the ratio of B_{MSY}/B_0 for XDB-SRA. It is unclear which of these assumptions is most appropriate. This is a broader issue than for just data-moderate assessments, since it questions the appropriateness of two-parameter curves such as Beverton-Holt to model the stock recruit relationship. Research to improve understanding of the relationship between the inputs of the XDB-SRA and exSSS productivity parameters is encouraged.

- Objective criteria are needed to evaluate minimum standards for model outputs to be considered "acceptable" and "preferred"
- Different priors (uniform of q / uniform on log-q) for the additional variance term were used in the two assessment models. It is unclear which performs best, and, since this term affects the weights given to each index in the model fitting, the form of the prior will influence model results, particularly when the indices are in conflict.

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

- The GMT representative expressed concern about local-scale population dynamics and exploitation patterns for nearshore species that are not captured in large-scale indices and assessment models, and recommended a more spatially-structured evaluation of CPUE data.
- The GMT representative noted that for certain nearshore species there is potential utility in using post-2003 RecFIN dockside data as well as onboard sampling data since depth restrictions have not constrained access to the adult population.
- The GMT representative also recommended expanding the analysis of CPUE data to additional sectors of the recreational fishery, such as private and rental boats. CPUE indices from these sectors may be useful in future assessments of nearshore stocks.

10. Panel Recommendations

10.1 Data input recommendations

1. The Panel strongly emphasizes the value of conducting a data workshop during which catches, indices, biology, and other data inputs are reviewed.
2. Consider developing GLMM models in which latitude and depth are treated as continuous covariates rather than as factors.
3. The historical CPFV drift-specific data should be keypunched, which should allow the algorithm for developing CPFV-based data indices to be improved.
4. Habitat maps should be developed so that structural rather than true zeros are designated using data which are independent from the data used to determine the indices.
5. Revisit the approach used to select among error models and whether to include ECE components when conducting the GLMM analyses.
6. Consider including a vessel factor (as a random effect) when developing indices for the Triennial survey.
7. Splitting the triennial survey into early and last periods became established practice without looking at the issue comprehensively or considering the loss of information from breaking a time series. A comprehensive evaluation of the issues and trade-offs is still needed.

8. Consistent residual patterns in NWFSC surveys for a number of assessments suggests there may be some unknown factor affecting survey catchability, or that some factor is affecting the productivity of multiple stocks in the same way.

10.2 Future reviews of data-moderate assessments

1. Nine stocks proved to be too many assessments to review at this STAR Panel. Reviewing a smaller number of assessments (4-8) may be more feasible goal for STAR Panel review, depending on the level of pre-STAR panel review of data inputs. If area-specific models are considered in addition to coast-wide models, additional time or fewer stocks should be scheduled. However, the first time that any assessment method or stock assessment is reviewed is always the most challenging, and future STAR Panels may find that the review goes much smoother.
2. The Panel recommends that data-moderate assessments continue to be reviewed at full STAR panels for at least for the next assessment cycle. As methods become standardized and the review process becomes more routine, it should be anticipated that review process can be streamlined somewhat.
3. Objective criteria should be developed to specify minimum standards for model outputs to be considered "acceptable" and "preferred" and included in the Terms of Reference for stock assessments. Such criteria might include minimum goodness-of-fit criteria and acceptable limits on posterior distributions.
4. While the Panel made some progress in comparing XDB-SRA and exSSS, our strategy of attempting to isolate the sources of difference between the two models ultimately proved unsuccessful, and resulted in complex requests to the STAT that were difficult to accomplish in the available time. The Panel suggests that some of the model comparison work is more appropriate outside the STAR panel review process, particularly as it involves fundamental differences in how stock productivity is modeled.
5. A standardized set of sensitivity runs, diagnostic plots, and performance statistics, such as runs tests on the residuals, should be developed to rapidly evaluate the performance of data-moderate assessments. Some pre-STAR panel planning involving the STAT and SSC to develop an analysis "package" could be helpful.
6. As with any assessment and review process, there is a trade-off between the number of data-moderate assessments and quality of the assessment and review. This trade-off should be taken into account when planning for future STAR panel reviews of data-moderate assessments.

10.3 Other Prioritized recommendations for future research and data collection

1. The MSE should be further explored to evaluate to performance of exSSS and XDB-SRA. Other potential topics include error in the catch time series, uninformative indices of abundance, and time-varying productivity. The MSE could also be used to test whether more constrained models, such as fixing steepness or B_{MSY}/B_0, results in improved model performance.

11. References

Givens, G.H. and S.E. Thompson. 1995. Alternative Bayesian Synthesis approaches to Bering-Chukchi-Beaufort Seas bowhead whale assessment: Uncertainty in historic catch and hitting with fixed MSYR. Reports of the International Whaling Commission 46: 509-530.

Ralston, S., Pearson, D., Field, J., Key, M. 2010. Documentation of the California commercial catch reconstruction project. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-461.

Stephens, A., and A. MacCall. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research 70:299-310.

Tagart, J.V. 1985. Estimated domestic trawl rockfish landings, 1963-1980. Unpublished Manuscript and Data. Washington Department of Fish and Wildlife.