Methodology Review Panel Report: Catch Estimation Methodology Review

National Marine Fisheries Service (NMFS)
Southwest Fisheries Science Center (SWFSC)
Santa Cruz, California
28-29 March 2018

Plus supplemental review webinar held on 31 July 2018

Methodology Review Panel Members:

David Sampson (Panel Chair), Scientific and Statistical Committee (SSC), Oregon State University

Noel Cadigan, Memorial University, Newfoundland, Center for Independent Experts (CIE)

Owen Hamel, SSC, National Marine Fisheries Service (NMFS), Northwest Fisheries Science Center (NWFSC)

Ole Shelton, SSC, NMFS, NWFSC (absent from the webinar)

Catch Estimation Methodology Team:

Nick Grunloh, University of California Santa Cruz

John Field, NMFS, Southwest Fisheries Science Center (SWFSC)

E.J. Dick, NMFS, SWFSC

Don Pearson, NMFS, SWFSC

Others Present:

Will Satterthwaite, SSC, NMFS, SWFSC

Cameron Speir, SSC, NMFS, SWFSC

Melissa Monk, NMFS, SWFSC

John DeVore, Pacific Fishery Management Council

Overview

This report describes a review of the catch estimation methodology being developed by staff at the National Marine Fisheries Service's Southwest Fisheries Science Center (SWFSC) laboratory in Santa Cruz, California. The Methodology Review Panel (the Panel) and the SWFSC team of analysts met in Santa Cruz during 28-29 March 2018. A number of interested parties joined the proceedings by means of listen-only webinar technology. The review process adhered to the Council's Terms of Reference for Stock Assessment Methodology Reviews (June 2016). The Panel did not include representation from the Groundfish Management Team or the Groundfish Advisory Panel.

The SWFSC catch estimation methodology addresses the problem of estimating time-series of landed catch by species from situations where landings data are available for "market categories", which are the labels fish buyers use for reporting their fish purchases (e.g., 100 pounds of *small rockfish* for $75). The landings data are separated into "strata" (e.g., *Port*, *Quarter*, *Market Category*) and fishery agents at fishing ports along the US West Coast routinely collect samples from the various strata and determine estimates of the stratum-level *Market Category* species composition by weight. However, only a small proportion of the landings (by market category) are directly sampled for their species compositions, which means that estimates of species compositions (*%Species*) must be developed for application to un-sampled landings in un-sampled strata. It is standard practice to derive these estimated *%Species* values for a given landing by borrowing information available from sampled landings that are in close proximity in time and space. The SWFSC catch estimation methodology, which takes a formal Bayesian approach for "borrowing" information, has several advantages over current practice: it has a solid basis in statistical theory; it predicts species compositions for un-sampled strata and domains of interest; and it will estimate the uncertainty associated with the derived landings time-series.

The draft agenda from the methodology review is attached as Appendix A, a statement of the goals and objectives for the review and the terms of reference are attached as Appendix B, and a list of remote participants connected via the webinar link is attached as Appendix C.

The report of the day-long webinar held on 31 July 2018 is included as Appendix D. At that webinar the Panel received responses by the SWFSC team to some of the short-term requests formulated during the March review.

Meeting summary

The two-day methodology review began with a welcome by the Panel chair, Dr. David Sampson (Oregon State University), followed by a round of self-introductions from the attendees, a brief review of the agenda, and assignment of reporting duties. The primary document on which the review focused (Grunloh et al., 2018) was provided several weeks in advance of the review; copies of the slideshows presented during the review were made available on an ftp site.

Day 1: Overview of California's groundfish fisheries and port sampling program

Dr. John Field (SWFSC) presented an *Overview of the California Current Ecosystem and California Groundfish Fisheries*. Landings by commercial fisheries in California (CA) have been tracked since 1928 using a "fish ticket" data system, but the information for some species groups (e.g., rockfish) was reported in an aggregated form. There was little sampling for the species compositions of the aggregated landings until 1978. Accounting for rockfish landings in CA by species is particularly challenging because there is greater diversity of *Sebastes* species in the waters off CA than elsewhere in the North Pacific. The 2010 landings reconstruction for CA

(Ralston et al., 2010) applied species composition from the late 1970s and early 1980s (when species composition sample data are available) to landings by market category to derive estimates of annual landings by species. This approach uses static *%Species* values by market category based on samples collected when the trawl fishery operated across a broad range of depths. Because the approach does not account for the gradual shift of the trawl fisheries to deeper waters during the historical time period, it likely overestimates the landings of slope species and under- underestimates the landings of shelf species.

Mr. Don Pearson presented an *Overview of the Existing Sampling Program and Catch Estimation Methodology*. In California the collection of commercial groundfish fishery data falls under the auspices of the California Cooperative Groundfish Survey (CCGS), which was established in 1978 to coordinate groundfish data collection and analysis activities involving staff from the California Department of Fish and Wildlife (CDFW, logistical support), the Pacific States Marine Fisheries Commission (PSMFC, port samplers), and the National Marine Fisheries Service (NMFS, database management and technical support). Data collected under the auspices of the CCGS are housed in a database called CalCOM, along with estimates of species compositions based on the current "borrowing" algorithm. Copies of the data and estimates are provided regularly to the Pacific Fisheries Information Network (PacFIN), a program administered by the PSMFC for storing and disseminating commercial fisheries information covering the US West Coast and Alaska. To derive estimates of landed catch by species, PacFIN multiplies the species composition estimates from CALCOM to fish ticket data uploaded from CDFW to the PacFIN database.

The CCGS data most relevant to the current methodology review are the species composition samples collected by the PSMFC port samplers. From those landings selected opportunistically for sampling, the port samplers typically take two 50-pound "cluster samples" that they identify to species and weight by species to generate sample estimates of the *%Species*. The information is post-stratified by *Year*, *Quarter*, *Market Category*, *Gear Group*, *Condition* (live/dead), and *Port Complex*. Commercial fishers in CA are not legally required to allow their fish to be sampled. A large percentage of the "strata" (i.e., domains of interest) with landings reported on fish tickets do not have any direct sample observations of the *%Species*. To process these un-sampled strata, the data processing system in CalCOM borrows data from sampled strata to impute (fill in) the missing *%Species*. The data borrowing rules allow borrowing across *Quarters* and some *Port Complexes*, but not across *Gear Groups*, *Market Categories*, or between ports on either side of Point Conception. The data borrowing rules, which were established when the CCGS began (1978), were based on expert opinion and not on any formal statistical analysis of the sample observations of the *%Species*. For data processing, the observed *%Species* are expanded to the landings by market category and then summed across sampled landings, i.e., the *%Species* values are weighted by the market category landings to generate the sample average *%Species*. Note that the applications of the SWFSC catch estimation methodology presented at the review did not weight the *%Species* values by the market category landings.

Day 1: Conceptual basis for the proposed model

Dr. E.J. Dick (SWFSC) gave a presentation on *Model-Based Catch Estimation: Assessment Context and Conceptual Basis* that laid out the motivation underlying the SWFSC team's approach. Arguments favoring a model-based approach rather than data-borrowing (status quo) are that the model-based approach provides a flexible framework for: sharing information among strata; quantifying uncertainty; objectively comparing alternative model structures (e.g., how to pool information); and evaluating variance components to identify simpler designs for collecting sample data. Although the current work is focused on improving estimates of commercial landings

from 1978 to the present, the model-based approach could be used to estimate species compositions for use in historical reconstructions (pre-1978), including estimates of uncertainty.

Day 1: Description of the model-based approach and its performance

Mr. Nick Grunloh (University of California Santa Cruz) gave a presentation entitled *Improving Catch Estimation Methods in Sparsely Sampled, Mixed Stock Fisheries* that laid out the basic statistical modeling approach and example applications to CA fish ticket landings by market category and associated species compositions sample data by market category from two time periods, 1978 to 1982 and 1983 to 1990. Full details of the approach were included in the draft manuscript (Grunloh et al., 2018) provided in advance of the methodology review. The statistical modeling approach, which is based on a Bayesian hierarchical modeling framework, builds on the approach that Shelton et al. (2012) used in a pilot study.

The two focal time periods for the applications to data (early: 1978- 1982; late: 1983-1990) are from the early years of the available data series when there were many fewer market categories than in recent years. Despite this, species compositions from the early and late periods clearly indicated marked changes in the number of market categories having appreciable landings (e.g., six market categories in the early period versus 12 in the late period) and marked changes in the species compositions for certain generic categories (e.g., market category 250, "unspecified rockfish").

The analytical team (the Team) considered a suite of statistical models that treat the weights of species in a sample as being integer-valued and the integer weight of the j^{th} species in the i^{th} sample as being distributed according to four possible distributions: Poisson, binomial, negative binomial, or beta-binomial. The pilot study by Shelton et al. (2012) only explored the Poisson model. Based on results from an application to data for market category 250 in the early period, the Team favored the beta binomial model, which is sufficiently flexible (relative to the three other distributions) to account for overdispersion in the data relative to the Poisson assumption, including a bimodal form, with high densities at 0 and 1, as observed in some of the data.

There was considerable discussion amongst the Panel members and with the Team regarding the statistical framework. A number of issues were discussed.

- The four probability distributions that were explored are all integer (counting) models, but the actual observed data are rounded fish weights (by species). It might be more accurate to model the underlying stochastic process as a counting process to generate the number of fish by species combined with continuous processes that generate the weights of the individual fish by species. However, it is unclear if the extra complexity of such a model would be beneficial.
- The overdispersion parameter in the beta binomial portion of the model (described as ρ [*rho*] in Grunloh et al. 2018) is treated as a single parameter across species for each time period (i.e., no subscript). This parameter is constant within a given market category / time period, even though the mean may change a great deal. Because the model variance is a function of the overdispersion parameter and the mean, and the compositions are bounded by 0 and 1, posterior predictive species compositions with large variances may be bimodal (e.g., Figure 8 in the Appendix to Grunloh et al. 2018). Trying to include a subscript on *rho* would likely get quite complicated and may not be doable in the modeling software the Team is using (INLA).
- It was noted that using Poisson distributions for each species when combined together results in a multinomial distribution. The species compositions (as proportions) have to sum to 1. A similar transformation can be applied to posterior predictive draws from all four of the

distribution models considered by the Team. Although there are ways to generate overdispersed multinomials, the Team explored only the above four distribution models.

There was consensus among the Panel that it would be beneficial in general (and going forward) if the Team could produce additional verification that the basic statistical model is performing adequately. Several options were discussed.

- Conduct leave-one-out cross-validation tests. Using WAIC (Widely Applicable Information Criterion) is an approximation of this. Cross-validation tests are simpler to understand and more convincing in general.
- Residual plots would be helpful diagnostics and might provide compelling evidence that the model is performing as expected and not producing results that are biased. For example, standardized residuals could show if there are appreciable differences in variability among port complexes.
- It should be possible to evaluate the variance structure by looking at the replicate clusters within each sample. Although the Team pooled the clusters for simplicity, the between-cluster differences provide information about the variance. One could also subsample randomly among clusters (within samples) to see how variable the results are.

A small issue was identified in the notation in the Grunloh et al. manuscript. In the equation for the mean value function (μ [*mu*]) the term for the observed aggregate cluster size for each sample (*n*) should not have a subscript *j* (for the species) because *n* represents the pounds in a sample aggregated across species. For most samples there are two clusters of about 50 lb. each so *n* should be about 100 in general.

After the Day 1 lunch break Nick Grunloh continued with his presentation of the Team's statistical modeling framework. Using the beta binomial model the Team explored six approaches for modeling temporal variation (by *Year* and *Quarter*) in the species compositions, ranging from a model with fixed additive effects for *Year* and *Quarter* to models with random *Year:Quarter* interactions having different forms of partial pooling controlled by the hierarchical prior variances. The Team used model comparison techniques to select among the six model structures. Inclusion of the *Year:Quarter* structure was demonstrably important. The Team elected to proceed with the model structure that had random *Year*:*Quarter* interactions constrained by a single variance parameter (denoted as model M4). Detailed results from the model fitting were shown for market category 250 for each of the two time periods (1978-1982 and 1983-1990).

Nick also presented results of an exercise that used Bayesian model averaging to explore models (based on M4) that used different forms of spatial pooling of the species composition data, ranging from a model with entirely separate *Ports* to models that pooled data from combinations of *Ports*. To reduce the number of possible combinations, the set of *Port* combinations was restricted to combinations having adjacent *Ports* and no more than three *Ports* in a combination, producing results that pooled data for 274 different *Port* combinations. The final results are based on averaging the 274 models using weights proportional to the posterior probability.

The Panel members raised various points with the Team concerning the model and the results.

- If there is seasonality in the species compositions there might be value to adding time-series components for the beta coefficients. The Team countered that using *Year*:*Quarter* interaction terms should capture any important seasonal trends.
- The Panel expressed concern about estimation stability in the model (given its complexity and the large number of parameters). The Team assuaged the concerns.

- There was a question about the practical consequences of differences in model structure for the predictive distributions from estimates. It would be informative to examine plots comparing the predictive fits from the six structural forms. Nick indicated that he had looked at plots of some of the different model outputs and that differences were generally small.
- E.J. discussed the feasibility of creating some test examples and propagating fits through the process to look at real world consequences of differences among model fits. This was agreed to be a good idea.
- There was discussion about the Team's use of crossed main-effects in the model structure as opposed to nested effects. The framework in Shelton et al. (2012) nests *Port* within *Species* and has other nesting properties not considered by the Team. There was a suggestion that the Team could consider a model that crossed coefficients for *Species* by *Port* to account for port-to-port differences in species composition.
- An example of single quarter hindcasts of the posterior predictive species compositions for market category 250 by *Port* and *Gear* indicated strikingly similar dispersed predictions for all *Port* and *Gear* combinations. This prompted discussion that the model structure may be accommodating wide dispersion at the expense of reduced differences in the predicted mean values.
- There was a question about the calculation of the posterior predictive compositions and whether the sample size for the total cluster weight (*n*) influences the posterior predictions.
- There was a question about weighting the different species composition samples as a function of the landing weight as the CalCOM system currently does in deriving average *%Species* values. Nick and E.J. showed an example model formulation that includes offsets for incorporating weighting by landing size. Although this formulation was not incorporated in the model results presented to the Review Panel, the model structure is sufficiently flexible to incorporate this additional feature.
- The point was made that if certain kinds of mistakes are bad and others are not, it would be sensible to focus on producing diagnostic plots that are informative for the kinds of mistakes that we are trying to avoid. This lead to discussion about the value of the general model-based catch estimation approach in the context of its potential value to management and the quality of the approach relative to the status quo (data borrowing) approach.
- To gauge predictive accuracy of the model the Team evaluated the species composition posterior predictive distributions using highest density intervals (HDI) at three levels containing 68%, 95% and 99% of the posterior predictive probability. Unlike other accuracy metrics, the HDI metric can accommodate the multimodal distributions that occur in the observations and the model predictions.
- There was a substantial series of questions about how one might compare the predictions from a model based approach against the predictions from a borrowing based (status quo) approach. There was limited agreement about what might be possible for these types of comparisons.

Following an afternoon break the Panel considered possible tasks the Team could consider overnight and respond to during the second day of the review. After much discussion the Panel agreed not to file any formal requests for the Team to consider that evening.

Day 2: Comparison of model results to status quo catch estimates for selected species

The second day of the review began with E.J. giving a presentation on *ComX Landings Distributions by Species/Gear/Year, and Species/Year* that applied the Team's model based approach (called ComX) to generate estimates of landings by species and compare them with

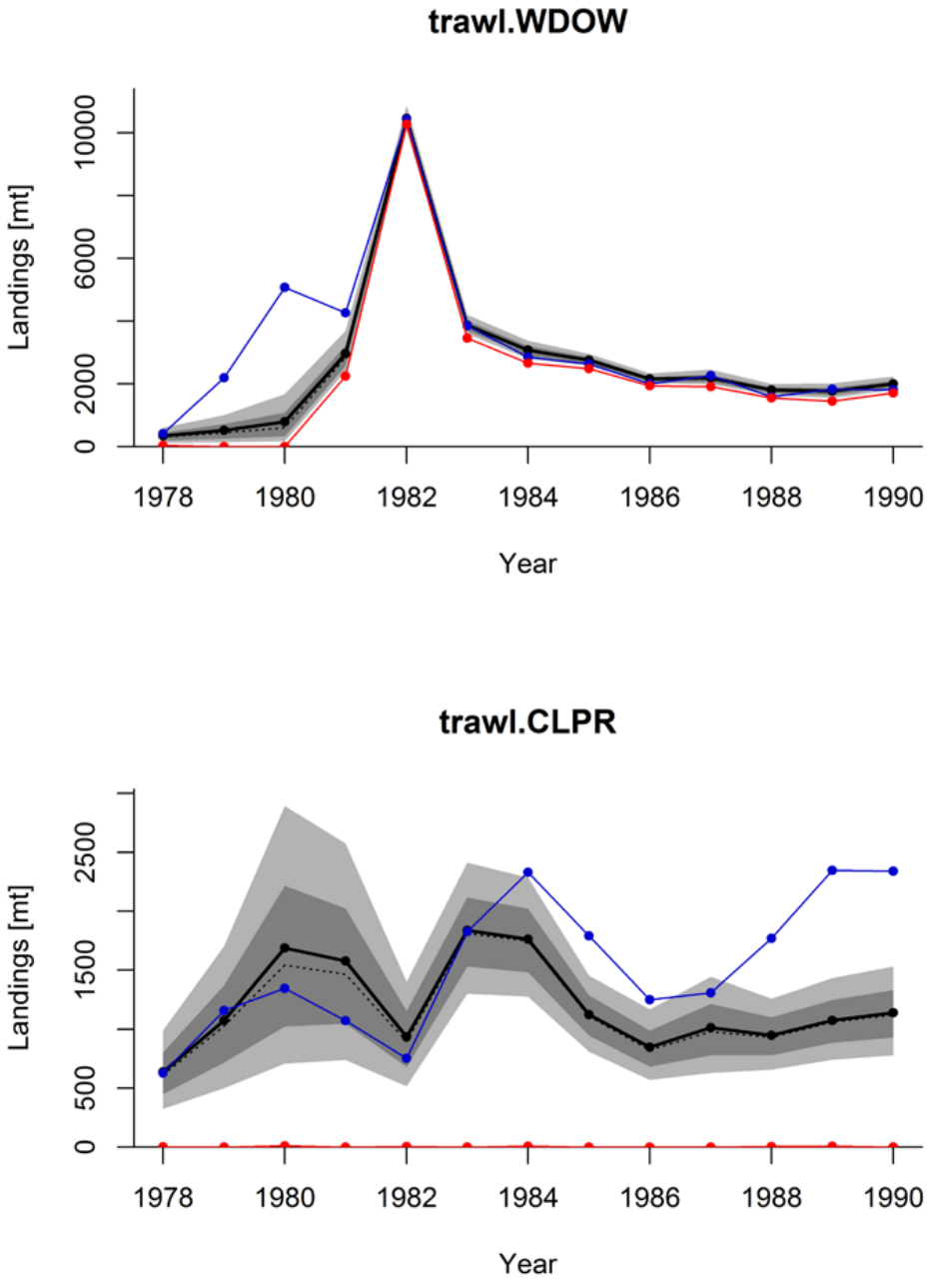
corresponding estimates based on COMLANDS (the table in CalCOM that uses the status quo, data borrowing approach to estimate species landings). The presentation focused on the top nine rockfish species, which accounted for about 90% of the landings during the focal period (e.g., widow rockfish, bocaccio, chilipepper rockfish), and an illustrative set of five minor rockfish species (e.g., cowcod, Pacific ocean perch). The comparison plots showed ComX and COMLANDS estimates of species landings by *Gear* and across all gear types for the period 1978 to 1990. E.J. noted that the results excluded strata for which ComX did not produce *%Species* estimates due to insufficient data; these same strata were excluded from COMLANDS to maintain strict apples-to-apples comparability.

Although there was reasonable agreement between the ComX and COMLANDS estimates in many of the plots, there were also examples of fairly large discrepancies in one or a few years, with the COMLANDS point estimates falling outside of the ComX 10^{th} to 90^{th} percentiles. Examples are shown below for trawl landings of widow rockfish and chilipepper rockfish.

There was discussion of the causes of the discrepancies between the ComX and COMLANDS estimates and speculation that the discrepancies could have several sources. One source may be the different forms of weighting in the two approaches. When combining *%Species* from different samples ComX does not weight by the market category landings whereas COMLANDS does. There was considerable discussion (and no consensus) about whether the *%Species* sample values should be weighted or treated as all being equivalent. The Team proposed that weighting by landings would be appropriate if species compositions vary systematically as a function of landed weight within a stratum. A second cause of discrepancies could be that COMLANDS includes forms of interaction (e.g., *Species* by *Port*, *Species* by *Gear*) that are not included in ComX. If suitable diagnostic plots were available, they might assist in locating (or eliminating from consideration) potential sources of the discrepancies. A third source for some of the discrepancies could be the abrupt changes that sometimes occurred in the sets of market categories. To aid interpretation it would have been helpful to have had a table of species and what market categories they appeared in each year.

One very useful output associated with the ComX estimates of species annual landings were tables of coefficients of variation (CVs) associated with the estimates. These CVs gave clear indications of species and time periods that had reasonable amounts of information to support the estimates and distinguish these from species and time periods for which there was poor information support.

Figure. Example comparison plots of estimated trawl landings of widow rockfish (upper panel) and chilipepper rockfish (lower panel). The blue points (connected by blue lines) are the COMLANDS estimates and the black points (connected by black lines) are the ComX median estimates. The lighter shaded regions indicate the 10th to 90th percentiles of the ComX estimates; the darker shaded regions indicate the 25th to 50th percentiles. The red points (connected by red lines) are the nominal market category landings.



Day 2: Potential means of serving data and variance estimates

Don Pearson gave a presentation on *Potential Means of Serving Data and Variance Estimates*, which outlined some of the anticipated challenges of getting ComX data and estimates to end users. Because ComX uses Bayesian methods, its estimates are distributions (rather than point estimates) and the number of landings estimates is greatly magnified compared to more traditional approach such as COMLANDS or in PacFIN. For each stratum and species there will be 10,000 posterior draws and the full ComX data base will likely need to house about 20 billion records. Don illustrated the system with some live queries from a test dataset of 5.6 billion records that included landings estimates for rockfish for the period 1978 to 1990; he demonstrated that the system was capable of very rapid information retrieval.

It was noted that the variances of strata estimates of landings from ComX are not additive as in traditional estimates based on stratified sampling schemes, which assume independence among strata. The non-additivity of the ComX estimates arises because the model-based approach estimates covariances among parameters and some parameters are shared across strata in the Bayesian statistical model.

Day 2: Futures efforts

After the Day 2 lunch break the group discussed additional issues that will arise if the Team's model-based catch estimation approach is applied more broadly than the focus of the current study (California rockfish commercial landings for the period 1978 to 1990).

Reconstructing CA landed catches during the "pre-data" period

The Team's analysis to date has focused on the relatively recent time periods, which are data-rich in terms of having contemporaneous species composition samples. The hope is that the Team's model-based methodology will provide a general framework that could be used to hind-cast estimates of *%Species* and derive estimates of species-level landings (and corresponding uncertainty estimates) for earlier years (pre-1978 in CA). Related to this topic is the information presented on Day1 that illustrated gradual transitions over the years to deeper fishing locations, implying that the *%Species* for the generic rockfish market categories should gradually shift over time to having greater proportions of slope rockfish species and reduced proportions of shelf rockfish species. Team members mentioned two possible ideas for hind-casting: (1) filter the earliest available species composition samples based on the fish ticket blocks and associated depths to generate appropriate species composition sample data sets and (2) slowly weed out the slope species from the earliest species composition samples. The Panel suggested an approach that filtered draws from the model to gradually thin out the slope species and obtain reasonable mean values through time. To offset shrinkage in the uncertainty (as species are removed) a constant CV could be used. Random draws could be used to get the slope of the time trend.

There was some discussion of how the Team's model would perform when applied to data from after 1990, when there are many more market categories and little year-to-year stability in the set of market categories. Would it be better to handle the changes in market categories using time-blocks or as a time-series process?

Reconstructing WA and OR landed catches

Although the port sampling programs in WA and OR share some features with the port sampling program in CA, there are also some important structural differences (and changes over time within WA and OR) that will need to be considered if the Team's model is to be applied to data from WA or OR. In principle there seem to be no good reasons why the modelling approach would not be

appropriate for WA and OR. There was general consensus that there should not be pooling of information across state boundaries because market categories are unlikely to be consistent from state to state.

Reconstructing landed catches for other species groups

There was discussion of whether the Team's model could be applied to other non-rockfish market categories that are comprised of mixtures of species (e.g., flatfish, skates, and sharks). Estimating species-level landings of skate is an immediate focus because the Council is interested in having stock assessments produced in 2019 for bignose skate and longnose skate. Species composition sampling for skates started in 2009 in CA, which means that any landed catch reconstructions for CA will have limited information available to support the estimates. Estimating species-level landings for un-specified flatfish in CA will (as with skates) have limited information available to support the estimates. Species composition sampling of the flatfish market categories started in 2004 in CA. These relatively recent species composition samples are likely to provide a poor reflection of conditions in earlier years when the trawl fisheries were unconstrained by the Rockfish Conservation Area closures.

Reconstructing recreational landed catches

The Team's model-based approach could probably be usefully applied to the available recreational data on reported landings of unspecified rockfish, but would likely involve a slightly different model. Species composition sampling in the recreational fisheries does not conform to the standardized cluster sampling protocols used in sampling the commercial fisheries. Also, the recreational data are generally recorded as catch rates in numbers of fish by species per unit effort and expanded using estimates of total effort rather than by percent composition of fish weights.

Review panel deliberations leading to requests and recommendations

Following the discussion of future efforts, the Panel, with feedback from the Team, discussed the following ideas, with the aim of formulating requests for additional work by the Team.

- The Panel members were in general agreement that the proposed model-based framework is superior to the ad hoc structure of the status quo data borrowing approach. However, the model-based framework is still being developed and is not yet ready for the Panel to provide with a definitive thumbs-up or thumbs-down evaluation.
- The Panel members were also in general agreement that it would be very helpful if the Team could provide a set of diagnostic plots to facilitate visual assessments of the model fits. It would be particularly useful if there were diagnostic plots available to (a) help evaluate the importance of fitting the over dispersion versus fitting the main effects and (b) help identify how the main effects might be modified.
- Also, a structure or agreed process is needed for determining how to whittle down from the very large set of possible models to something more reasonable and manageable. There was discussion of the pros and cons of using DIC, WAIC, and other statistical model selection approaches.
- A series of self-test simulations would provide important support for the validity of the statistical model and the model construction approaches.
- Another step in the evaluation of the methodology should take the projections based on the statistical models and consider how they would affect management decisions.

- It is important from an institutional perspective that the process of model construction, estimation, and selection is robust and repeatable. A standardized set of procedures for analysts to follow needs to be formalized to ensure robust outcomes.
- The Panel and Team were in general agreement that a model-based approach is an improvement and that there is little to be gained from a formal comparison between the proposed model-based approach and the existing data borrowing methods.

Review Panel recommendations – short term

Because the Panel's report on this methodology review will not be reviewed by the Council until September 2018, it seemed plausible that the Team could produce additional analyses for consideration by the Panel prior to September. The Panel could then produce an Addendum to this report that describes results of the additional analyses and the Panel's conclusions regarding the new information. The Panel decided to structure its recommendations into short-term ones that the Team would try to respond to by mid-July 2018 and long-term recommendations that would need to be reviewed later by the SSC or a methodology review panel.

At a day-long webinar held on 31 July 2018 the Panel received responses by the Team to many of the short-term requests. Details of the webinar and the Team's responses are provided in Appendix D.

Short-term requests

- As a diagnostic template, for each sampled stratum compare the posterior predictive distributions at the 68^{th}, 95^{th}, and 99^{th} percentiles with the current observed species proportions (create fully stratified versions of tables 2 and 3 in the Grunloh et al. methods documentation). With each row, include sample sizes and associated landing weights with a graphical display to highlight problems and outliers (circle size proportional to landing weights).

  Rationale: The Team provided broad-scale summary metrics (e.g., WAIC, MSE and DIC) for evaluating the goodness-of-fit of the different model forms and structures. Fine-scale diagnostics are needed to help identify aspects of the data that are not adequately addressed by the different models. The diagnostic template will provide a mechanism for fine-scale exploration of goodness-of-fit.

- The diagnostic template should be developed for each of the sensitivity runs (vary across a range of plausible time models and priors and limit to the top 2-3 market categories).

  Rationale: Application of the diagnostics across a wide range of models will form a test of how well the diagnostics illustrate whether the models capture important structural features that are thought to be embedded in the data.

- Explore an alternative time block: an extension of 1983 and 1984 to the first time block.

  Rationale: The panel expressed concerns about how the model would perform when applied to shorter time periods, as will occur when the model is used with data more recent than 1990. Results from the above recommendation could be compared to the results from the current two time blocks (1978-1982; 1983-1990) to explore how fits to data from the late period degrade when the model for the late period is based on fewer years of data. Also, comparisons of the two forms of blocking serve as a sensitivity evaluation of the selection of the block boundary, which was chosen on a fairly arbitrary basis.

- Explore various two-way interactions (beyond the current explorations; e.g., *Species*:*Port* and *Species*:*Gear*).

Rationale: The Team did not have time to search across the multitude of possible interaction terms that they could have included in the model. From various anecdotal comments made during the review it seemed likely that the model would benefit from the inclusion of other interaction terms. Explorations with the diagnostic template may suggest potentially beneficial terms.

- Explore variability in sampling among clusters within samples.

  Rationale: The Team combined the data from the cluster samples, which had been drawn independently from the sampled landings. These replicate samples contain information regarding the structure of the variability in the data that may provide useful corroboration (or refutation) of the structure implicit in the statistical model.

- Redo the modeling of the early time block without southern CA ports. Explore spatially and temporally (i.e., alternative time blocks).

  Rationale: The available dataset does not have any sample data in the early time block from the southern CA ports. It was unclear how this lack of data influenced the model results. The requested analysis will clarify the situation.

- Compare alternative ComX outputs and the current time series of estimated catches.

  Rationale: It would be informative to see the landings estimates corresponding to the additional models developed in response to the above requests. The landings estimates can be generated for a small set of illustrative species and do not need to be comprehensive.

- Provide a summary table of species' sample sizes in each market category by time block.

  Rationale: The requested information will assist in understanding where there are gaps in the available data that the model is filling in by means of its pooling structure.

- Provide self-test documentation (simulated data) for example models.

  Rationale: Results from this analysis will provide a demonstration of model performance under best-case scenarios, where the data being analyzed exactly conform to the assumptions of the statistical model. The analysis will serve to verify (or refute) that the model performs as expected.

Technical merits and/or deficiencies of the methodology

The model-based catch estimation framework developed by the SWFSC Team is superior to the ad hoc structure of the status quo data borrowing approach in that it has a firm theoretical basis and it produces objective estimates of landings and associated measures of uncertainty. That said, there are still aspects of the approach that are underdeveloped; the method is not yet ready for direct application to existing data sets for generating estimated landings series for use in stock assessments.

Areas of disagreement regarding panel recommendations

Areas of disagreement among panel members

There were no areas of disagreement among the Panel members.

Areas of disagreement between the panel and proponents

There were no areas of disagreement between the Panel and the Team.

Unresolved problems and major uncertainties

- Given that the species composition data are indexed by seven distinct dimensions (*Year*, *Quarter*, *Market Category*, *Species*, *Port*, *Gear*, *and Condition*), there are a very large number of potential interaction terms that could be incorporated into the model structure. However, the observational data are quite scarce and will not support a highly parameterized model. The challenge is to develop an approach to model building that results in a parsimonious model, meaning the model is simple, but not too simple. Further, there needs to be evidence leading one to conclude (a) that the model's predictions mimic all the important systematic features of the data or (b) that any instances of systematic lack of fit have no important consequences for stock assessments and management.
- The model-based approach (by design) does not produce estimates for very poorly sampled strata. The Team elected to "refrain from modeling any period where the minimum possible number of effective parameters exceeds the number of samples for the modeled period." The landings associated with poorly sampled strata will continue to be reported either as the nominal species (e.g., nominal blue rockfish) or as a generic category (e.g., unspecified rockfish). In general there are small landings associated with poorly sampled strata, which mitigates the impact of having poorly sampled strata.

Issues raised by the public and others during the review

There were no issues raised by the public or others during the review. For the most part the only people attending the review meeting in person were members of the Panel or the Team.

Review Panel recommendations – long term

The following items are likely to require considerable time and effort by the Team to explore and resolve. However, the Panel considers that first two items will need resolution before the methodology is used to produce landings estimates for stock assessments.

- Explore weighting by landings using gear-specific and species-specific models (conditional on results of the explorations outlined in the short-term recommendations).

  <u>Rationale</u>*: ComX and COMLANDS take fundamentally different approaches to deriving the %Species values that are applied to the market category landings. The ComX approach (no weighting of the sample %Species by the size of the sampled landings) has the advantage of reducing the potential high influence of %Species from large landings (because all sample %Species are treated as equivalent). The COMLANDS approach (weight the sample %Species by the size of the sampled landings) has the advantage of producing estimates that are "consistent" in that if all the landings were fully sampled the resulting %Species values would be exactly equivalent to the actual species compositions.*

- Explore over-dispersion in the ρ [*rho*] parameter to allow greater differences among species.

  <u>Rationale</u>*: The model underlying the results presented to the Panel had a single over-dispersion parameter across species for each time period (i.e., no subscript). This appeared to have the effect of artificially inducing bimodality in the posterior predictive species compositions. It would be helpful to understand the consequences of this structural assumption (on the goodness-of-fit and on the estimated landings) by comparing results of the current model with results from models that allow species-to-species variation in their over-dispersion.*

- Investigate discrepancies between ComX catch series and current catch series.

<u>Rationale</u>: It is important to understand the source(s) of discrepancies between the ComX and COMLANDS landings estimates to better inform any conclusions regarding which approach is "better".

Research and data needs

The Panel and Team did not identify any research and data needs beyond those described above in sections *Review Panel recommendations – short term* and *Review Panel recommendations – long term*.

References

Grunloh, N., Dick, E., Pearson, D. Field, J. and Mangel, M. 2018. Improving catch estimation methods in sparsely sampled mixed-stock fisheries. Unpublished manuscript, draft version dated 13 March 2018.

Ralston, S, Pearson, D.E., Field, J.C., and Key, M. 2010. Documentation of the California catch reconstruction project. NOAA Tech. Memo. NMFS, NOAA-TM-NMFS-SWFSC-461.

Shelton, A.O., Dick, E.J., Pearson, D.E., Ralston, S., and Mangel, M. 2012. Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Canadian Journal of Fisheries and Aquatic Sciences 69: 231–246.

Appendix A.

PROPOSED AGENDA
Catch Estimation Methodology Review

Pacific Fishery Management Council
March 28: NMFS Southwest Fisheries Science Center
Santa Cruz Laboratory
110 McAllister Way
Santa Cruz, CA 95060

March 29: Center for Ocean Health Library
Ocean Health Building
Room 201
University of California Santa Cruz
115 McAllister Way
Santa Cruz, CA 95060

Telephone: 831-420-3900

March 28-29, 2018

This is a meeting of a Pacific Fishery Management Council-sponsored methodology review with remote listen-only attendance via webinar (see webinar information below). Public comments will be accepted at the discretion of the Chair.

A suggestion for the amount of time each agenda item should take is provided. All times are approximate and subject to change. At the time the agenda is approved, priorities can be set and these times revised.

To Attend the GoToWebinar:

1. Use this link: http://www.gotomeeting.com/online/webinar/join-webinar
2. Click "Join a Webinar" in the top right of page.
3. Enter the Webinar ID: 942-468-499
4. Please enter your name and email address (required)
5. You must use your telephone for the audio portion of the meeting by dialing this TOLL number 1-415-930-5321
6. Enter the Attendee phone audio access code 580-006-830
7. Enter your audio phone pin (shown on screen after joining the webinar)

System Requirements

- PC-based attendees: Required: Windows® 7, Vista, or XP
- Mac®-based attendees: Required: Mac OS® X 10.5 or newer
- Mobile attendees: Required: iPhone®, iPad®, Android™ phone or Android tablet (See the GoToMeeting Webinar Apps)

WEDNESDAY, MARCH 28, 2018 – 8:30 AM

A. Call to Order-GFSC Administrative Matters

1. Call to Order and Introductions — Dave Sampson
2. Approve Agenda
3. Rapporteur Assignments
4. Overview of Terms of Reference
   (8:30 a.m., 0.5 hours)

B. Overview

1. Overview of California Fisheries and Methodology Objective — John Field
   (9 a.m.; 0.5 hours)

2. Overview of the Existing Sampling Program and Catch Estimation Methodology — Don Pearson
   (9:30 a.m.; 0.5 hours)

BREAK (10-10:15 a.m.)

C. Basis for the Proposed Model — E.J. Dick

1. Assessment and Management Implications
2. Review of Theoretical Basis for Model-based Approach
3. Hierarchical Model Structure
   (10:15 a.m.; 0.75 hours)

D. Description of the Proposed Model — Nick Grunloh

1. ComX Model Development
2. Model Description
3. Model Validation
   (11 a.m.; 1 hour)

LUNCH (12-1:30 p.m.)

E. Model Performance and Model Averaging Approach — Nick Grunloh
(1:30 p.m.; 1.5 hours)

BREAK (3-3:15 p.m.)

F. Discussion by Reviewers

1. Q&A with Model Proponents
2. Methodology Discussion
   a. Historical Catch Reconstruction Issues
   b. Impediments to More Recent Periods
   c. Sample Weighting Issues
3. Consider Potential Model Explorations

(4 p.m.; 1 hour)

THURSDAY, MARCH 29, 2018 – 8:30 AM

G. Comparison of Model Results to Current Catch Estimates for Select Species E.J. Dick
(8:30 a.m.; 1.5 hours)

BREAK (10-10:15 a.m.)

H. Response to Reviewers' Questions and Comments from Day 1 and Discussion of Potential Next Steps
(10:15 a.m.; 1.75 hours)

LUNCH (12-1:30 p.m.)

I. Potential Means of Serving Data and Variance Estimates Don Pearson
(1:30 p.m.; 1 hour)

J. Future Efforts

1. Efforts for Reconstructing Catches During the "Pre-Data" Period
2. Reconstructing Historical WA and OR Catches
3. Reconstructing Catches for Other Species Groups
4. Reconstructing Recreational Catches

(2:30 p.m.; 1 hour)

BREAK (3:30-3:45 p.m.)

K. Reviewer Panel Recommendations
(3:45 p.m.; 1.25 hours)

ADJOURN

PFMC
03/05/18

Appendix B. Goals and objectives and terms of reference

Goals and Objectives

The goals and objectives of the review of the new SWFSC methodology are to:

1. Evaluate the theoretical basis of the SWFSC's new Bayesian hierarchical modeling approach for estimating species-level landings and associated estimates of uncertainty;
2. Compare the Bayesian model-based landings estimates from the application to California data with the existing estimates (based on the data-borrowing rules) and establish which of the two estimation methods produces the best available species-level landings data for use in stock assessments;
3. Identify potential impediments to the application of the new methodology to California data from other time periods, which likely have other sets of market categories and different fishery characteristics;
4. Evaluate the feasibility and utility of applying the new methodology (or extensions of it) to California data for historical periods for which species-composition samples are even sparser (or altogether absent); and
5. Identify potential advantages and impediments to the application of the new methodology to data from Oregon and Washington, including the incorporation and routine maintenance of species-level estimates of landings in the regional data repository.

Terms of Reference (for the Review Panelists)

1. Become familiar with the draft report describing the proposed methodology, the analytical model underlying the methodology, the example application of the methodology to data from the California fisheries for groundfish, and the analytical model along with other pertinent information prior to review panel meeting.
2. Discuss the technical merits and deficiencies of the analytical method and the input data during the open methodology review meeting.
3. Evaluate model assumptions, estimates, and major sources of uncertainty.
4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified.
5. Determine whether the science reviewed is considered to be the best scientific information available.
6. When possible, provide specific suggestions for future improvements in any relevant aspects of data collection and treatment, modeling approaches and technical issues, differentiating between the short-term and longer-term time frame.

Appendix C. Remote participants at (1) the March review meeting and (2) the July webinar.

1. March review meeting

Last Name	First Name
Ainsworth	Justin
Budrick	John
Errigo	Mike
Erwin	Brenda
Jannot	Jason
Larinto	Traci
Lyons	James
Richter	Gerry
Satterthwaite	Will
Tsou	Theresa
Wargo	Lorna
Weyland	Phil
Whitman	Alison

2. July supplemental webinar

Last Name	First Name
Armeli Minicante	Enrico
Brock	Tara
Buell	Troy
Cadigan	Noel
Cahalan	Jen
Carter	Al
DeVore	John
Edwards	Jason
Field	John
Grunloh	Nicholas
Hamel	Owen
John	Budrick
Krause	Sandra
Larinto	Traci
Mandrup	M
Mattes	Lynn
Niles	Corey
Paulling	Marissa

Last Name	First Name
Petersen	Joe
Reed	Heather
Sampson	David
Satterthwaite	Will
Waterhouse	Lynn
Whitman	Alison

Appendix D. Supplemental Catch Estimation Methodology Review Report.

To be completed.

Appendix E.

PROPOSED AGENDA
Supplemental Catch Estimation Methodology Review

Pacific Fishery Management Council
Large Conference Room
7700 N.E. Ambassador Place, Suite 101
Portland, OR 97220
Online Webinar
Telephone: 503-820-2280

July 31, 2018

This is a meeting of a Pacific Fishery Management Council-sponsored methodology review with remote attendance via webinar (see webinar information below). There will also be a public listening station at the Pacific Council office (address listed above). This meeting is open to the public and public comments will be accepted at the discretion of the Chair.

A suggestion for the amount of time each agenda item should take is provided. All times are approximate and subject to change. At the time the agenda is approved, priorities can be set and these times revised. Discussion leaders should determine whether more or less time is required, and request the agenda be amended.

To Attend the GoToWebinar:

8. Use this link: https://www.gotomeeting.com/webinar
9. Click "Join a Webinar" in the top right of page.
10. Enter the Webinar ID: 531-002-459
11. Please enter your name and email address (required)
12. You must use your telephone for the audio portion of the meeting by dialing this TOLL number 1-914-614-3221
13. Enter the Attendee phone audio access code 953-706-939
14. Enter your audio phone pin (shown on screen after joining the webinar)

System Requirements

- PC-based attendees: Required: Windows® 7, Vista, or XP
- Mac®-based attendees: Required: Mac OS® X 10.5 or newer
- Mobile attendees: Required: iPhone®, iPad®, Android™ phone or Android tablet (See the GoToMeeting Webinar Apps)

TUESDAY, July 31, 2018 – 8:30 A.M.

A. Call to Order

1. Call to Order and Introductions — Dave Sampson, Panel Chair
2. Approve Agenda
3. Rapporteur Assignments
4. Summary of March Review — Dave Sampson
4. Review Terms of Reference — Dave Sampson
   (8:30 a.m., 0.5 hours)

B. Presentation of Diagnostic Templates

1. Diagnostics for Each of the Sensitivity Runs (Requests 1 and 2) [1/] — Nick Grunloh and E.J. Dick
2. Discussion
   (9 a.m., 0.75 hours)

C. Presentation of Summary Tables of Sample Sizes in Each Market Category by Time Block

1. Summary Tables Presentation (Request 7) [a/] — Don Pearson and John Field
2. Discussion
   (10:15 a.m., 0.25 hours)

Break (10:30 a.m., 0.25 hours)

D. Exploration of Alternative Time Blocks

1. Revised Modelling of the Early Time Block Without Southern California Ports (Requests 3 and 5) [a/] — Nick Grunloh and E.J. Dick
2. Discussion
   (10:45 a.m., 1.25 hours)

Lunch (12 p.m., 1 hour)

E. Exploration of Various Two-Way Interactions Beyond the Previous Explorations

1. Two-Way Interactions (e.g., Species:Port, Species:Gear) (Request 4) [a/] — Nick Grunloh and E.J. Dick
2. Discussion
   (1 p.m., 1 hour)

[1/] The request numbers in the topics refer to the numbered list of requests specified in the terms of reference for the supplemental methodology review.

F. Compare Alternative ComX Outputs and the Current Time Series of Estimated Catches

1. Comparison of Outputs (Request 6) [a/] — Nick Grunloh and E.J. Dick
2. Discussion
   (2 p.m., 1 hour)

Break (3 p.m., 0.25 hours)

G. Example of Simulation Self-Test from Past Work

1. If Possible, Provide a Simple Example (Request 8) [a/] — Nick Grunloh and E.J. Dick
2. Discussion
   (3:15 p.m., 0.75 hours)

H. Panel Discussion

(4 p.m., 1 hour)

ADJOURN

PFMC
07/18/18