

Center for Independent Experts (CIE) External Independent Peer Review
Independent Peer Review of Catch Estimation Methods in Sparsely Sampled Mixed Stock
Fisheries.

March 28-29, 2018

Dr. Noel Cadigan

Centre for Fisheries Ecosystems Research
Marine Institute of Memorial University of Newfoundland
St. John's, NL. Canada

Executive Summary

The proposed catch estimation method for sparsely sampled mixed stock fisheries was extensive, ambitious, and has a high potential to be a substantial improvement in methodology. The catch estimation methodology is well advanced; however, more research is yet required to demonstrate model robustness and reliability, and to demonstrate why estimates based on the new method are improvements over current data-borrowing estimates, particularly when there are substantial differences between the two procedures. Additional research is required on model diagnostics to demonstrate reliability, and on model assumptions (e.g. no interactions between species compositions, ports, and gears) to demonstrate robustness. A more appropriate multivariate statistical distribution should be used to model port-sampling data, and this could have impacts on model selection, catch estimation, and uncertainty evaluation.

Background

The National Marine Fisheries Service (NMFS) and the Pacific Fishery Management Council (PFMC) held a methodology review panel to evaluate and review a new methodology under development by the Southwest Fisheries Science Center (SWFSC) for partitioning landings reported as aggregated categories of fish into species-level estimates of landed catch. Commercial landings by market category are reported as “aggregate landings” and are considered to be a complete census. Species-level landings are produced by California, Oregon, and Washington by applying species-composition proportions from port samples to their respective sets of aggregate landings data. Each state generally provides species-level estimates of landings by strata: year, quarter, market category, port-group, gear-type, and condition (live/dead). There are many of these strata, and often the states apply ad-hoc data borrowing rules to impute species-proportions for strata that have not been sampled. The new methodology proposed by the SWFSC scientists involves a hierarchical modelling approach to estimate the species composition of catches in unsampled strata, and to quantify uncertainty in historical landings. The main focus of the peer review was to evaluate the efficacy of the new methodology as applied to California data, in light of the larger objective of whether or not to recommend that the method be used to revise historical catch estimates.

Participation of an external, independent reviewer was considered to be an essential part of the technical review process of the proposed SWFSC methodology. The Center for Independent Experts (CIE) reviewer was required to be an active and engaged participant throughout panel discussions and able to voice concerns, suggestions, and improvements while respectfully interacting with other review panel members, advisors, and stock assessment technical teams. The CIE reviewer was required to have excellent communication skills in addition to working knowledge of Bayesian hierarchical statistical modeling approaches and recent experience applying statistical analyses to compositional data (e.g., species- or length-compositions). Experience and familiarity with fishery landings data processing and management systems was

advantageous. The CIE reviewer was contracted to deliver an independent peer review report formatted with content requirements as specified in Annex 1. The Terms of Reference (ToR) for this specific methodology review are attached in Annex 2.

The review meeting was coordinated by Stacey Miller, Northwest Fisheries Science Center. The review meeting was chaired by Dr. David Sampson, Oregon State University.

Role of reviewer

I reviewed background materials and reports provided by Mr. John DeVore on March 14, 2018. These reports were made available via an ftp site provided by the Pacific Fishery Management Council. These documents are listed in Appendix 1. I participated in the review meeting and contributed substantially to the review of the proposed methodology. I recorded important results and conclusions during the meeting and provided these notes to the meeting chairperson for consideration in developing the review panel report. During and subsequent to the review meeting I conducted some simulation testing of statistical methods relevant to the proposed SWFSC catch estimation methodology (CEM), and I provide these results and hope they will be useful in further development of the model.

This report is structured according to my interpretation of the required format and content described in Annex 1 of Appendix 2.

Note: CEM - SWFSC catch estimation methodology.

Summary of findings

ToR 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.

The reported titled ‘DRAFT: Improving Catch Estimation Methods in Sparsely Sampled Mixed-Stock Fisheries’, by Grunloh et al. (2018) was provided prior to the meeting. This report described:

- The problem of incomplete species information when landings are recorded for mixed-species market categories. Also, sometimes categories that are named after a single species still contain several species.
- The current practice used to estimate species catch based on market category landings data (assumed to be a census) and port sampling of species compositions within market categories and other strata.

- Some rationale for the large number of strata used for ‘catch expansion’, and problems with sampling many strata.
- Current practice of “data borrowing” rules to deal with un-sampled strata.
- Model-based approaches to deal with un-sampled or sparsely sampled strata, and the need for methodological improvements.
- Sparsely sampled mixed-stock rockfish fisheries, and challenges for estimating species catch, particularly for California fisheries.
- Methodological extensions to the model-based framework proposed by Shelton et al. (2012): i) alternative likelihoods to address over-dispersion; ii) multiple hierarchical structures for pooling information through time; iii) integrate model predictions across uncertainties in the spatial model structure; iv) summarize a general framework for quantifying uncertainty including an efficient database design for dissemination of results at any level of aggregation.

I also read closely the Shelton et al. (2012) paper, and my main concern with that paper was that the model they proposed did not include over-dispersion, and I was pleased to see that this was addressed in the model extensions that we reviewed.

ToR 2. Discuss the technical merits and deficiencies of the analytical method and the input data during the open methodology review meeting

The problem of estimating species catch based on market category (i.e. mixed-species) landings information and often sparsely sampled catch strata is complex, and I conclude that the modelling initiatives reviewed at the panel meeting were extensive and ambitious and I think the SWFSC scientists are well advanced in extending ‘best practice’ to deal with this problem.

There are many merits to the proposed CEM, and as usual there may be alternative and better methods which I discuss below. However, the main deficiency discussed during the review meeting of the CEM, and the deficiency that I feel needs to be addressed before I could recommend that the approach be used to provide species-specific catch estimates, is that the sometimes substantial differences exist in the CEM estimates compared to those currently provided using somewhat ‘ad hoc’ data borrowing rules. Sometimes the CEM catch prediction intervals are substantially different from current estimates. We need to understand why these differences exist. It could be a problem with sparse sampling and the application of data borrowing rules in the currently used method, or it could be a problem with the proposed CEM model. I need to better understand these differences before I would feel comfortable in recommending that the CEM is a substantial improvement in terms of catch estimates. The CEM is certainly an improvement in terms of uncertainty quantification because there is no quantification of uncertainty in the current data-borrowing approach that is used.

As was pointed out in the Grunloh et al. (2018) report, the model based-approach was partially motivated from “small area estimation” model-based approaches used to analyze survey samples when the objective is to provide inferences about smaller areas than the survey was designed for. This approach can be used to deal with un-sampled strata. I find the basic model-based approach to be appropriate for species catch estimation. However, a drawback of a model-based approach is possible lack of robustness to model assumptions. It will be useful to demonstrate that CEM estimates are reasonably robust to minor changes in model assumptions, or larger changes to assumptions that are subjective (i.e. priors) and hard to verify. This is the sort of information a review panel needs to see to determine if a method is ‘robust and reliable’.

The species composition port sampling use a cluster sampling protocol for the many strata of California’s commercial fisheries. Each sample is intended to be two 50 pound clusters selected at random from a stratum. Variations in the sampling protocol may result in only a single cluster being taken, or the size of clusters may vary based on the particular challenges of sampling each stratum. Samples are recorded as integer pounds for each observed species. These data were analyzed using various count data distributions with over-dispersion, similar to Shelton et al. (2012).

Modelling integer weights using a counting process was discussed at the review meeting. This is an unusual way to model such data. Model results may be sensitive to the weight scale; for example, if species were modelled in ounces rather than pounds. Preferred models presented at the panel review meeting included over-dispersion, so I expect those results will not be very sensitive to the scale chosen, but this should be checked. Models with no over-dispersion will be sensitive to the weight-scale chosen. Alternative and perhaps more appropriate statistical models for the port-sampling data are beta-regression and Dirichlet-regression models in which the species-fraction of total sampled weight is modelled. These could be weighted by the total weight of the sampled catch.

The model was estimated based on the total catch weight by species, usually aggregated across each of two clusters selected in a stratum. Note that sometimes only one cluster was sampled due to practical limitations. Within-cluster sample differences are not accounted for in the model. No information was provided on the magnitude of the within-sample differences in species catch, so I am not sure how important this might be. This information should be provided, and if it is an important source of variation then it should be included in a revised model.

An important deficiency identified at the review meeting was a lack of diagnostics to evaluate model goodness-of-fit. The analysts were reluctant to produce residual diagnostic plots; they preferred to focus on the adequacy of coverage of prediction intervals for species catch composition. This may be fine and looking at the coverage of a range (i.e. 67 and 95%) of intervals may address mis-specification of both the model mean and variance. However, these types of diagnostics that were presented at the review meeting were aggregated too much over strata and for blocks of years within market categories to be informative about the reliability of the model at the scale (i.e. year, gear, port, and species) that is relevant to stock assessment. We discussed in

detail a diagnostic procedure to evaluate model fit. A Bayesian procedure was outlined and the model analysts will provide an example of this diagnostic plot for the review meeting summary report.

ToR 3. Evaluate model assumptions, estimates, and major sources of uncertainty

Some of the technical description of the SEM model in Grunloh et al. (2018) needs improvement. It was a little hard to follow in places because of inconsistent notation. The Appendix A: Motivating Example lacked detail. The text only described that $y_{ij} \stackrel{iid}{\sim} p(y_{ij}|\theta_j, \varphi)$ which is poor notation. Usually we write upper case $Y_{ij} \stackrel{iid}{\sim} p(y_{ij}|\theta_j, \varphi)$. More importantly, it is not clear how the total sample size of the i 'th cluster, $n_i = \sum_j y_{ij}$, is included in these distributions.

I suggest it is clearer to write $Y_{ij} \stackrel{iid}{\sim} p(y_{ij}|\lambda_i, \pi_j, \varphi)$ where π_j is the species- j proportion of catch weight ($\sum_j \pi_j = 1$), λ_i is a nuisance parameter for the size of the i 'th sample with $E(Y_{ij}) = \lambda_i \pi_j$, and φ is an over-dispersion parameter. For the Poisson distribution it is well known that $Y_i = \sum_j Y_{ij}$ also has a Poisson distribution with mean λ_i , and the MLE of this parameter is $\hat{\lambda}_i = y_i$. In this case, all of the relevant sample information about the π parameters is provided by the conditional distribution, $Y_{ij}|Y_i, j = 1, \dots$. However, such a partitioning of information does not exist if Y_{ij} are Negative Binomial (NB) random variables (see below), and in this case the MLE of λ_i is no longer y_i . I was not sure how the NB version of the Grunloh et al. (2018) model was implemented.

The Binomial and Beta-Binomial versions of the CEM model treated each species sample integer catch-weight as independent Binomial or Beta-Binomial random variables. Even if we assume that a counting process is appropriate for these data, modelling them as independent random variables is not strictly appropriate because it does not explicitly account for the sample constraint that $\sum_j Y_{ij}$ is usually fixed. The negative correlations in the data that are caused by this sampling constraint may not be properly accounted for using independent Binomial or Beta-Binomial models. I think an important consequence is less reliable goodness-of-fit (or model selection) statistics (e.g. DIC or WAIC). This will be described more under ToR4.

The models presented at the review meeting did not include interaction between gear, species compositions, and ports. Hence, the model is constrained so that species compositions are assumed to be similar across ports and gears. They could only differ by an additive amount for the logit of species proportions. We needed to see detailed model diagnostics to support this modelling assumption. However, interactions between year and quarter parameters were modelled in a variety of ways. This is good. Maybe the model could be extended to include possibly random interactions between species compositions, ports and gear. However, I recognize that model parsimony is important, and I am not advocating that such interactions should be included in the model. I am only advocating that such interactions be further explored.

The posterior predictive distribution, $BB(\dots)$, needs a sample size (n). The value used was $n=100$. Does this matter?

An issue that I thought about after the review meeting was fish length. Is this information collected during port sampling, and should it be included in the model? It may be that the length distribution of the catch has some bearing on species composition, in addition to gear type. I am not sure how I would include length information in such a model, but at least some consideration of this would be useful.

ToR 4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified

A research interest of mine involves likelihood components for compositional count data in integrated stock assessment models. I have been investigating methods to use for Negative Binomial compositional count data, which I adapted for this review in the context of the species composition samples. I focus on the NB distribution because in some survey contexts (e.g. Cadigan, 2011) it provides a good description of the variability in catches for many species.

I simulation tested the model outlined in Appendix A of the Grunloh et al. (2018) as best I could understand it. I considered $n = 10$ or $n = 50$ cluster samples from a population of 7 species, with a population vector of proportions (of numbers) $\pi = (0.01, 0.1, 0.2, 0.38, 0.25, 0.05, 0.01)$. The total sample size per cluster was randomly chosen but usually 50:

```
N = rnbinom(n,mu=60,size=25)
N[N>=50]=50.
```

This R code generated samples sizes that are usually $N_i = 50$, but some samples could be smaller. I randomly generated species counts by sub-sampling from NB random species counts:

```
n.sp = length(pi);species=1:n.sp
X = matrix(NA,nrow=n, ncol=n.sp)
for(i in 1:n){
  X[i,]=rnbinom(n.sp, mu=20*N[i]*pi, size=k)
  if(sum(X[i,])>50){
    tx = table(sample(species,50,replace=TRUE,prob=X[i,]/sum(X[i,])))
    tx1 = rep(0,n.sp)
    tx1[match(as.numeric(names(tx)), species)]=tx
    X[i,]=tx1
  }
}
```

I assumed that the catch of the j 'th species in the i 'th trip was $NB(20N_i\pi_j, k)$ with $k=1$, and that about 1/20 of the catch was port-sampled. The NB total catch of species j in trip i had mean $\mu_{ij} =$

$20N_i\pi_j$ and variance $\mu_{ij} + \mu_{ij}^2/k$. If the total catch for all species exceeded 50 in a trip, then a subsample of size 50 was selected with probabilities equal to the total catch proportions.

I estimated the species proportions using five statistical models. The first was the Multinomial distribution. I assumed $Y_{ij}|Y_i$ was Multinomial (MN) with parameters π_j . This is a mis-specified model which I investigated to demonstrate that over-dispersion results in inaccurate confidence interval coverage for π 's. I also fit a simple NB model, $Y_{ij} \stackrel{iid}{\sim} NB(\lambda_i\pi_j, k)$. This is also a mis-specified model because of the sub-sampling of catches. For Poisson data this subsampling is basically Poisson thinning and the subsampled catches are still Poisson distributed, but this is not the case for NB distributed catches. Whereas the Multinomial model can be motivated from $Y_{ij} \stackrel{iid}{\sim} Poi(\lambda_i\pi_j)$, no such partitioning of information for $\underline{\pi}$ based on $Y_{ij}|Y_i$, and for λ_i based on Y_i , exists for NB data. For each i , the distribution of $Y_{ij}|Y_i$ and Y_i both involve all of the parameters, $\underline{\pi}$, $\underline{\lambda}$, and k . Hence, a conditional NB model for port-sampling fixed sample size composition data is not possible, and this is why I used the above NB model.

An approximate conditional distribution may be derived for NB data by formulating this statistical model as a Poisson-Gamma mixture model. That is, if $Y_{ij}|\gamma_{ij} \stackrel{iid}{\sim} Poi(\lambda_i\pi_j\gamma_{ij})$ and γ_{ij} are iid Gamma random variables with mean 1 and variance k^{-1} , then the marginal distribution of $Y_{ij} \stackrel{iid}{\sim} NB(\lambda_i\pi_j, k)$. The approximate conditional NB distribution (NBac) I investigated is based on $Y_{ij}|\{Y_i, \gamma_{ij}\} \stackrel{iid}{\sim} MN(\pi_{ij}^*)$ where $\pi_{ij}^* = \pi_j\gamma_{ij} / \sum_j \pi_j\gamma_{ij}$. I used TMB to derive the marginal likelihood of $Y_{ij}|Y_i$. This is a Multinomial-gamma mixture model which is different from the Dirichlet-Multinomial (DMN) model. I also investigated this latter model which may be derived as a mixture of Multinomial counts with probabilities that have a Dirichlet distribution, and the Dirichlet probabilities may be derived from gamma random variables with means π_j and variances π_j/k ; that is, $Y_{ij}|\{Y_i, \gamma_{ij}^\dagger\} \stackrel{iid}{\sim} MN(\pi_{ij}^\dagger)$ where $\pi_{ij}^\dagger = \gamma_{ij}^\dagger / \sum_j \gamma_{ij}^\dagger$ and γ_{ij}^\dagger are Gamma random variables with mean π_j and variance π_j/k . The NBac and DMN are similar in that $E(\pi_j\gamma_{ij}) = E(\gamma_{ij}^\dagger) = \pi_j$ but these distributions are substantially different because $Var(\pi_j\gamma_{ij}) = \pi_j^2/k$ whereas $Var(\gamma_{ij}^\dagger) = \pi_j/k$. For the same value of k , DMN counts will be more variable than NBac counts, especially for species with small π_j . An issue with the NBac model is that it ignores sample information about the NB k parameter that is in the sample totals, Y_i .

The final statistical model I investigate is to model each $Y_{ij}|Y_i$ as Beta-Binomial random variables, with means $\pi_j Y_i$ and variance $\pi_j(1 - \pi_j)Y_i\{1 + (Y_i - 1)\rho\}$. This is the preferred statistical model in Grunloh et al. (2018). This is another over-dispersed count data distribution, but it does not explicitly account for negative correlations in Y_{ij} for different j when the total $Y_i = \sum_j Y_{ij}$ is fixed by the sampling design.

All models were implemented using TMB and estimated in R using `nlminb()` and the gradient function provided by TMB. There are 7 π_j parameters to estimate, but because $\sum_{j=1}^7 \pi_j = 1$ these can be estimated using only 6 free parameters which I call $\underline{\theta}$. I used a logistic model for this:

```
pfunc = function(theta){
  etheta = c(exp(theta),1)
  pi=etheta/sum(etheta)
  return(pi)
}
```

I evaluated model fit using AIC, for simplicity. This is different from Grunloh et al. (2018), but I have no experience with WAIC or DIC statistics they used, and these statistics seem to be more appropriate for Bayesian and/or hierarchical models. AIC may be sufficient for this simple simulation setting.

There is a technical issue that probably makes the AIC's not comparable for some of the models. This is because the data used in the conditional and unconditional models are in some sense different. For example, with the NB model there are $n + 7$ parameters but no constraints on the counts, whereas for the NBac model there are only 7 parameters, but the data are considered to be constrained (i.e. conditioned) so that $\sum_j Y_{ij} = Y_i$ for $i=1, \dots, n$. Hence the models are not applied to the “same” data and some adjustment of likelihoods may be required, along the lines of Appendix B in Albertsen et al. (2016)¹. However, the AIC's for the Multinomial, NBac, and the Dirichlet-Multinomial models should be comparable.

For simplicity, I used a parametric bootstrap procedure based on MLEs of $\underline{\theta}$ and its asymptotic covariance matrix to generate a multivariate bootstrap distribution for $\underline{\pi}$. I derived confidence intervals for $\underline{\pi}$ using the percentile method. Better confidence interval procedures are possible.

I generated 10 simulations data sets of species counts in samples of about 50 fish, and fit these data using the five statistical models outlined above. Results are illustrated using 1 simulation for $n=50$ and $n=10$ cluster samples in Figure 1.

¹ Albertsen, C.M., Nielsen, A. and Thygesen, U.H., 2016. Choosing the observational likelihood in state-space stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(5), pp.779-789.

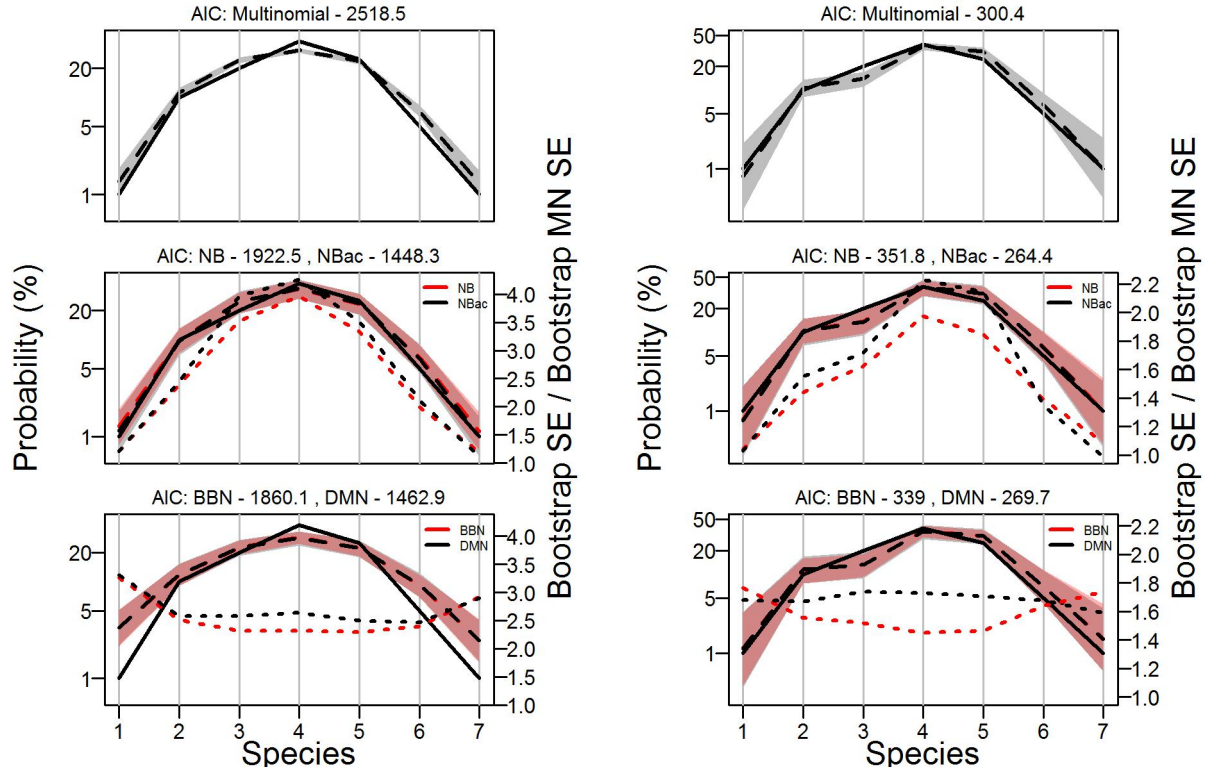


Figure 1. Species composition estimates, based on NB simulated data with $n=50$ (left panel) and $n=10$ (right panel) port-samples, each of approximate size 50. Solid black lines connect population species proportions. Dashed lines and shaded regions indicate π estimates and 95% confidence intervals. Colors correspond to statistical estimation methods which are: NB – Negative Binomial, NBac – approximate conditional NB, BBN – Beta-Binomial, DMN - Dirichlet-Multinomial. Dotted lines indicate standard errors (SE) for each estimation method relative to Multinomial standard errors, to indicate the type of over-dispersion in the estimates.

Results were qualitatively similar for the other simulations. The BBN and DMN confidence intervals usually did not cover, or barely covered, true values when the π were small and when $n=50$, but the bootstrap confidence intervals for these methods seemed to have better coverage properties when $n=10$. The NB or NBac intervals seemed to have good coverage properties for $n=10$ or $n=50$, although I admit that 10 simulations is much too low to properly evaluate confidence interval coverage properties. Also, the NBac AIC was lowest in all 10 simulations, indicating that AIC could diagnose the true data generating distribution among the conditional variants I studied (i.e. MN, NBac, BBN, and DMN). The NB AIC may not be directly comparable to the other models for reasons outlined above.

I performed another set of 10 simulations when species counts were generated from a Dirichlet-Multinomial distribution, with $n=50$ or $n=10$:

```

X = matrix(NA,nrow=n, ncol=n.sp)
for(i in 1:n){
  rpi=rgamma(n.sp, shape=k*pi, scale=1/k)
  rpi=rpi/sum(rpi)
  X[i,]=rmultinom(1,N[i],rpi)
}

```

I used $k=4$ in this simulation so that the variance of the catch for species 4 was about the same as in the NB simulation.

Seven of 10 simulation when $n=10$ generated no catch for one of the rare species, and this caused other estimation problems. When $n=50$ this did not happen, and the bootstrap confidence intervals from the BBN and DMN estimation procedures seemed to have good coverage properties (e.g. see Figure 2), whereas the NB and NBac intervals seemed less reliable. The DMN AIC was the lowest of the five methods in all 10 simulations when $n=50$.

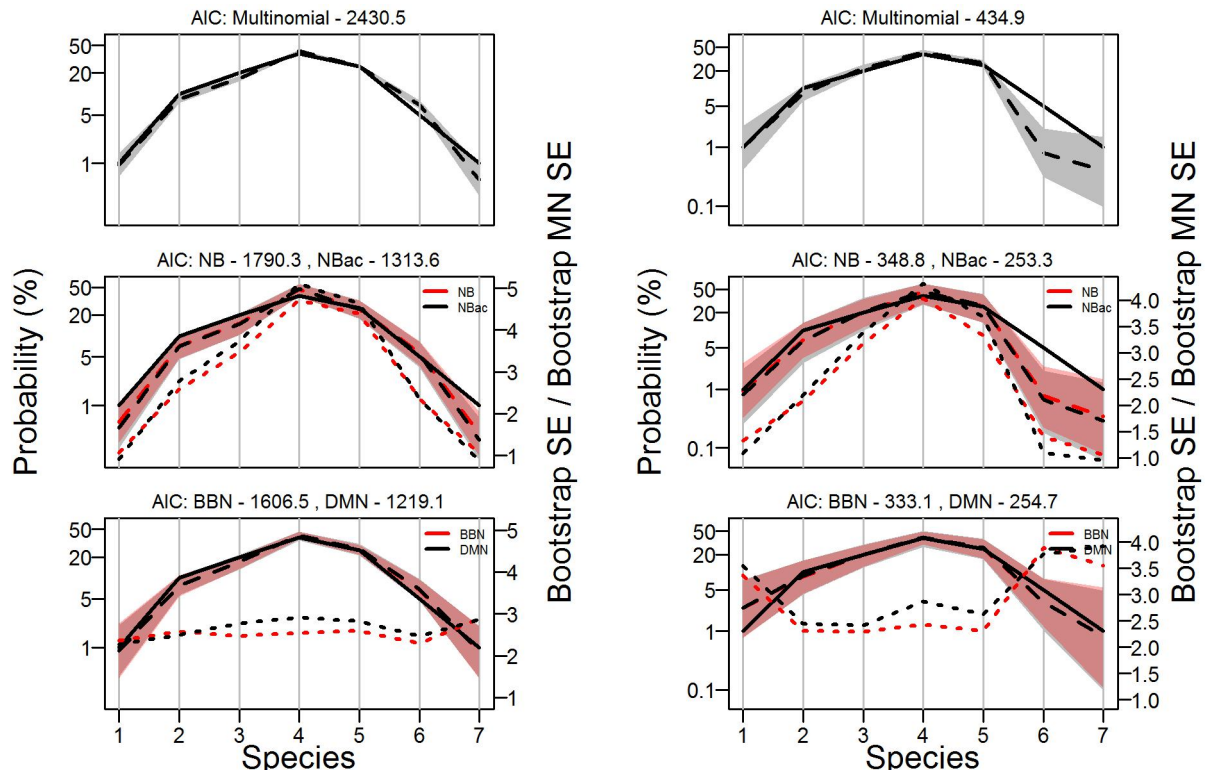


Figure 2. Species composition estimates based on DMN simulated data with $n=50$ (left panel) and $n=10$ (right panel) port-samples, each of approximate size 50. See Figure 1 for more description of this figure.

I hope these simulations have demonstrated that it is important to get the type of over-dispersion right. Also:

- When there is over-dispersion in the data, then statistical inferences based on the multinomial distribution will not be reliable. This supports the choice of an over-dispersed model in Grunloh et al. (2018).
- The type of over-dispersion matters and, depending on the source, the effects of over-dispersion on the standard errors of estimates of species composition (π) may depend on the value of π and the effect may not be simply a scaling of multinomial-type standard errors.
- Predictive distributions for species compositions based on a Beta-Binomial approximation of Dirichlet-Multinomial distribution may be slightly too narrow.
- Model selection based on a Beta-Binomial approximation of the Dirichlet-Multinomial distribution may not be reliable.
- Model selection statistics need to be applied to the same data. Conditional distributions like the Dirichlet-Multinomial are not directly comparable to unconditional distributions like the NB.

I also used the simulations to explore some diagnostic procedures to check goodness-of-fit. The qqplot is commonly used for this purpose. Using estimated model parameters, I generated a large number (10 000) of species counts based on observed sample sizes to derive theoretical quantiles of species counts to compare with observed quantiles. The results are illustrated in Figure 3a,b. For NB generated species counts (Figure 3a), the BBN and DMN estimators resulted in observed count distributions that have tails that are slightly too long. The NB estimator also did not fit as well as the NBac. Conversely, for DMN simulated data the NB and NBac estimators resulted in count distributions that have tails that were too short and provided very poor fits to lower quantiles (Figure 3b). For example, the NB and NBac models under-estimated the proportions of zero counts for DMN simulated data.

Such diagnostic procedures may provide additional support for model selection compared to a single statistic like AIC, WAIC, etc. They help us to understand why a model does not fit our data well. However, such diagnostic parameters do not penalize for model complexity and therefore I suggest that they only be used as an aid to validate model selection.

In practise it may be difficult to diagnose the type of over-dispersion in data. In this case the best that we can do is try to be robust to this issue, and focus on methods that work well when our assumptions are met, and do not perform seriously bad when there are reasonable deviations or violations in our assumptions.

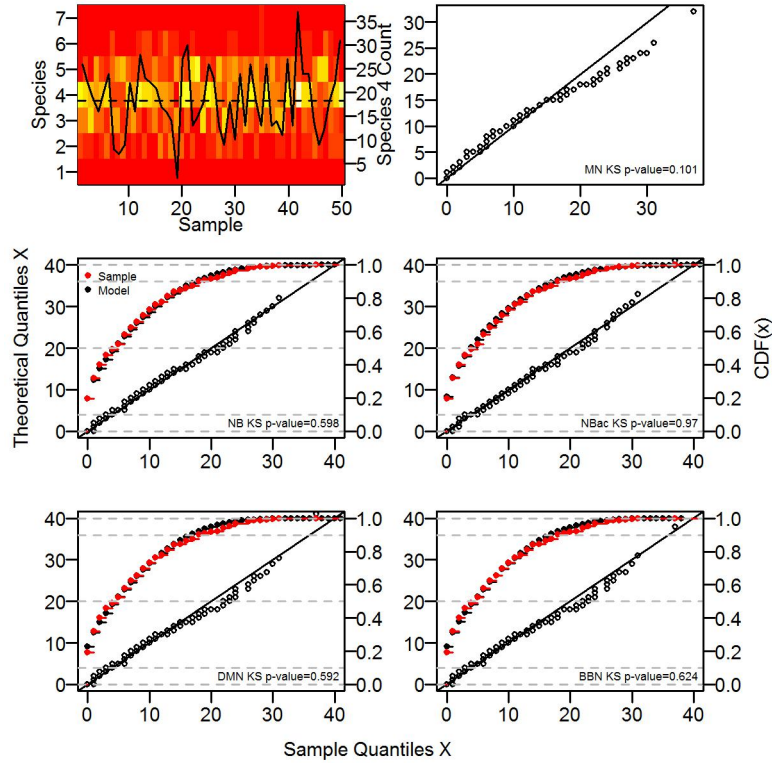


Figure 3a. Goodness of fit diagnostics for $n=50$ samples from NB species compositions. The data are illustrated in the top left panel with species counts (rows) from each sample (columns). Red indicates a low count and white indicates a high count. Furthermore, the solid lines indicate the sample counts for species 4. Observed and model predicted data quantiles are shown with open plotting symbols in other panels. Solid symbols indicate observed (red) and model predicted (black) cumulative distribution functions (CDF's) of species counts. The p-value for a Kolmogorov–Smirnov (KS) test for equality of the observed and model predicted CDF's is shown in each panel.

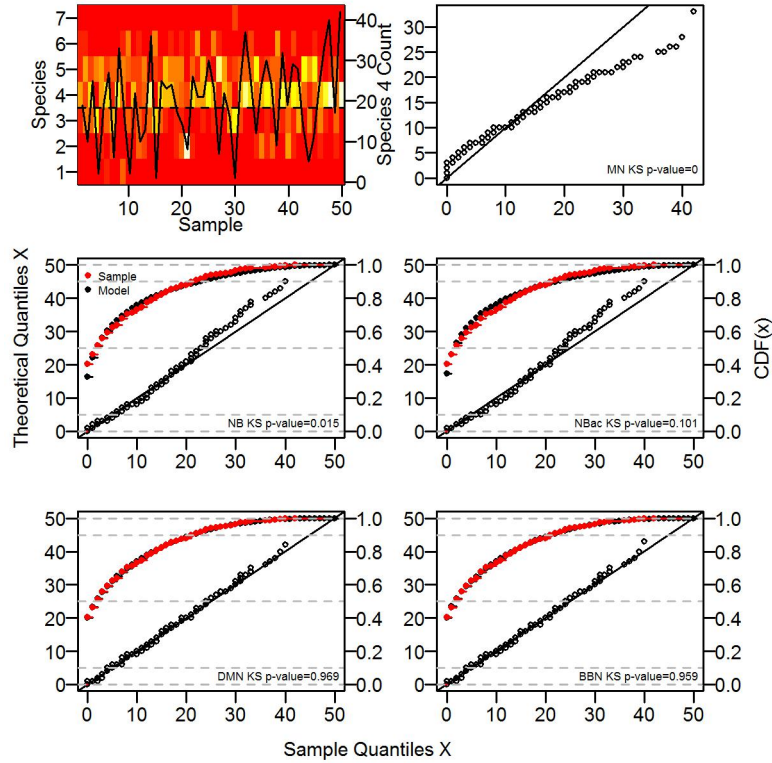


Figure 3b. Goodness of fit diagnostics for $n=50$ sampled from Dirichlet-Multinomial species compositions. See Fig. 3a for more description of this figure.

ToR 5. Determine whether the science reviewed is considered to be the best scientific information available

At present I do not conclude that the science reviewed is the best scientific information available. I did not sufficiently understand the sometimes substantial differences in CEM estimates of species catch compared with those estimates currently provided using data borrowing rules. I think the CEM model has a high potential to be a substantial improvement in methodology to estimate catch, but some more research is yet required to demonstrate model robustness and reliability, and to demonstrate why CEM estimates are improvements over current data-borrowing estimates, particularly when there are substantial differences between the two procedures. This could involve some detailed case studies where the data-borrowing decisions and subsequent species composition and catch estimates are re-examined and compared with CEM estimates. It would be especially useful if there was some ancillary information available to support one method or the other.

ToR 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.

Short-term

The CEM model should be extended to include a multivariate conditional distribution for species counts (in pounds). The Dirichlet-Multinomial or the approximate conditional Negative Binomial models are two reasonable choices and they could be evaluated for goodness-of-fit.

Provide graphical diagnostics of model fit. A procedure was outlined during the review meeting and I illustrated another approach under ToR4. Some type of qqplot is recommended in addition to what was outlined at the review meeting, and not instead of that procedure.

Consider extending the model to include possibly random interactions between species compositions, ports and gear.

Make sure that model goodness-of-fit (a.k.a. model selection) statistics are applied appropriately for conditional versus un-conditional models.

Simulation test methods – at least simplified (i.e. toy) versions of the methods.

Long Term

Modelling integer weights using a counting process is unusual. Investigate alternative and perhaps more appropriate statistical models for the port-sampling data such as beta-regression and Dirichlet-regression models in which the species-fraction of total sampled weight is modelled.

Examine and report within-cluster port-sample differences in species compositions.

I suggest a more contemporary approach to this problem involves a state-space model. One could formulate a spatiotemporal species catch-distribution model, perhaps using some software like VAST. I admit this could be complex and highly ambitious, especially if seasonal changes in species catch-distributions are to be accounted for. The observation model could involve modelling market sampling categories like fleets with a selectivity function or stochastic process applied to species abundance. Changes in the market category selectivity could be informed by changes in management practices and fishermen's behavior over time. Finally, a port-sampling procedure could be applied to model predictions of species abundance within market categories to create the state-space observation model. The data are the same (port-samples of species composition) but the approach is different. One works from the population to the data in the state-space approach, rather than the data to the population which CEM seems to do at present.

Conclusions and Recommendations

ToR 2. Discuss the technical merits and deficiencies of the analytical method and the input data during the open methodology review meeting

The modelling initiatives reviewed at the panel meeting were extensive and ambitious and I think the SWFSC scientists are well advanced in extending ‘best practice’ to deal with the problem of combined species landings. The proposed analytical method is particularly valuable in that it can provide a quantification of catch uncertainty, which is rarely done in fisheries systems I am familiar with.

The main deficiency of the analytical method that I feel needs to be addressed before I could recommend that the approach be used to provide species-specific catch estimates are the sometimes substantial differences that exist in the proposed estimates compared to those currently provided using somewhat ‘ad hoc’ data borrowing rules. Sometimes the newly proposed catch prediction intervals are substantially different from current estimates. We need to understand why these differences exist.

Modelling integer weights using a counting process is unusual.

An important deficiency was a lack of detailed diagnostics to evaluate model goodness-of-fit.

ToR 3. Evaluate model assumptions, estimates, and major sources of uncertainty

I was not sure how the NB version of the model was implemented.

The model assumed port-sampling species catches were independent Beta-Binomial random variables which is not strictly appropriate because the catches have negative correlations due to the total sample size constraint of 50 lbs. An important consequence may be less reliable goodness-of-fit (or model selection) statistics (e.g. DIC or WAIC).

The model did not include interactions between gear, species compositions, and ports. This could result in unreliable estimates of catch if such interactions are important.

ToR 4. Provide constructive suggestions for current improvements if technical deficiencies or major sources of uncertainty are identified

Modelling integer weights using a counting process is unusual. Investigate alternative and perhaps more appropriate statistical models for the port-sampling data such as beta-regression and Dirichlet-regression models in which the species-fraction of total sampled weight is modelled.

If statistical distributions based on counting processes are pursued further, then I think more appropriate multivariate models such as the Dirichlet-Multinomial should be used. However, this choice will depend on the nature of over-dispersion in the data.

ToR 5. Determine whether the science reviewed is considered to be the best scientific information available

At present, I do not conclude that the science reviewed is the best scientific information available. This because I did not sufficiently understand the sometimes substantial differences in the proposed estimates of species catch and those estimates currently provided using data borrowing rules.

The proposed method has a high potential to be a substantial improvement in methodology to estimate catch, but some more research is yet required to demonstrate model robustness and reliability, and to demonstrate why estimates based on the new method are improvements over current data-borrowing estimates, particularly when there are substantial differences between the two procedures.

ToR 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.

The CEM model should be extended to include a multivariate conditional distribution for species counts (in pounds). The Dirichlet-Multinomial or the approximate conditional Negative Binomial models are two reasonable choices and they could be evaluated for goodness-of-fit.

Provide graphical diagnostics of model fit. A procedure was outlined during the review meeting, and I illustrated another approach under ToR4. Some type of qqplot is recommended in addition to what was outlined at the review meeting, and not instead of that procedure.

Appendix 1: Bibliography of materials provided for review

Grunloh, N., E.J. Dick, D. Pearson, J.C. Field and M. Mangel. In prep. Improving Catch Estimation Methods in Sparsely Sampled Mixed-Stock Fisheries.

Miller, R.R., Field, J.C., Santora, J.A., Schroeder, I.D., Huff, D.D., Key, M., Pearson, D.E. and MacCall, A.D., 2014. A spatially distinct history of the development of California groundfish fisheries. *PloS one*, 9(6), p.e99758.

Pearson, D. E. and G. Almany. 1995. The effectiveness of California's commercial rockfish port sampling program. NOAA Technical Memorandum NMFS-SWFSC-218. 50 p.

Pearson, D. and B. Erwin. 1997. Documentation of California's Commercial Market Sampling Data Entry and Expansion Programs. NOAA Technical Memorandum NMFS, NOAA-TM-NMFS-SWFSC-240. 67 p.

Pearson, D., B. Erwin, and M. Key. 2008. Reliability of California's Groundfish Landing Estimates from 1969-2006. NOAA Technical Memorandum NMFS, NOAA-TM-NMFS-SWFSC-431. 139 p.

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

Shelton, A. O., Dick, E. J., Pearson, D. E., Ralston, S. and M. Mangel. 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(2), 231-246.

Appendix 2: CIE Statement of Work

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Review of Catch Estimation Methods in Sparsely Sampled Mixed Stock Fisheries

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act (MSFCMA), Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Project Description:

NMFS and the Pacific Fishery Management Council (PFMC) will hold a methodology review panel to evaluate and review a new methodology under development by the Southwest Fisheries Science Center (SWFSC) for partitioning landings reported as aggregated categories of fish into species-level estimates of landed catch. Existing port sampling programs collect samples of the aggregated categories (generally referred to as market categories) and determine the species composition of these samples. There are legal requirements to use "fish tickets" to report data on commercial landings by market category and these reported "aggregate landings" are considered to be a complete census. The existing data processing systems used by California, Oregon, and

Washington estimate species-level landings by applying species-composition proportions from the port samples to their respective sets of aggregate landings data. Each state generally provides species-level estimates of landings by year, quarter, market category, port-group, gear-type, and condition (live/dead). Often the states apply ad-hoc data-borrowing rules to impute species-proportions for “strata” that have not been sampled.

The new methodology proposed by the SWFSC uses a Bayesian hierarchical modeling approach based on the theory and framework laid out in Shelton et al. (2012; *Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data*; Can. J. Fish. & Aquatic Sci.; 69: 231-246). The SWFSC has further developed the Bayesian modeling approach and applied it to groundfish fisheries data from California for a period of years during which there was reasonable stability in the number of different market categories. The new approach and associated system for inference and model exploration promise to quantify uncertainty in historical landings and lead to a better understanding of the effects of having highly stratified and sparse species-composition sampling, while making the most from the available data. The theoretical basis of the SWFSC’s approach and lessons learned from its application to the California data will be the focus of the review, recognizing that a larger objective is whether or not to recommend that the method be used to revise historical catch estimates.

The general goals and objectives of PFMC methodology reviews are to:

- 1) Ensure that research surveys, data collection, data analyses and other scientific techniques in support of coastal pelagic species (CPS) and groundfish stock assessments are the best available scientific information and facilitate the use of information by the Council;
- 2) Provide recommendations regarding whether, and if so, how a particular methodology can be applied in future stock assessments;
- 3) Meet the MSFMCA and other legal requirements;
- 4) Follow a detailed calendar and fulfil explicit responsibilities for all participants to produce required outcomes and reports;
- 5) Provide an independent external review of survey and analytical methods used to develop data to inform CPS and groundfish stock assessments;
- 6) Increase understanding and acceptance of CPS and groundfish research methodologies and review by all members of the PFMC family;
- 7) Ensure that methodologies not directly related to stock assessments, such as economic analyses or ecosystem-based fishery management approaches, undergo adequate peer review, as appropriate; and

- 8) Identify research needed to improve assessments, reviews, surveys, analyses, and fishery management in the future.

The goals and objectives specific to 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.

This methodology review will likely provide the basis for making changes to how the three states (CA, OR, WA), in coordination with the Pacific Fisheries Information Network (PacFIN, <https://pacfin.psmfc.org/>), provide species-level estimates for the numerous species of groundfish managed by the PFMFC. Estimated landings are fundamental data for the stock assessments that provide the scientific basis for setting Overfishing Limits (OFLs) and Acceptable Biological Catches (ABCs) as mandated by the MSFCMA. The technical review of the proposed methodology will take place during a formal, public, multiple-day meeting of state and federal fishery statisticians and stock assessment experts. Participation of an external, independent reviewer is an essential part of the review process.

The CIE reviewer is contracted to deliver an independent peer review report be formatted with content requirements as specified in **Annex 1**. Terms of Reference for this specific methodology review are attached in **Annex 2**. The tentative agenda of this panel review meeting is listed in **Annex 3**.

Requirements for CIE Reviewer

NMFS requires one reviewer to conduct an impartial and independent peer review in accordance with the SoW, OMB Guidelines, and the Terms of Reference (ToRs) below. The CIE reviewer shall be an active and engaged participant throughout panel discussions and able to voice concerns, suggestions, and improvements while respectfully interacting with other review panel members, advisors, and stock assessment technical teams. The CIE reviewer shall have excellent communication skills in addition to working knowledge of Bayesian hierarchical statistical modeling approaches and recent experience applying statistical analyses to compositional data (e.g., species- or length-compositions). Experience and familiarity with fishery landings data processing and management systems would be advantageous.

Tasks for Reviewer

Pre-review Background Documents

Review the following background materials and reports prior to the review meeting. Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site to the CIE reviewer all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review, for example:

- The current draft report by the SWFSC describing the Bayesian modelling approach, the relevant California fisheries data*, and preliminary results from applying the Bayesian model to the California data;
- The PFMC's Scientific and Statistical Committee's Terms of Reference for the Methodology Review Process for Groundfish and Coastal Pelagic Species for 2017-2018;
- Additional supporting documents as available.
- An electronic copy of the data*, the parameters, and the software used for applying the model to the data (if requested by reviewer).

** Due to confidentiality constraints, the data made available to the reviewer and the public will need to be anonymized and aggregated.*

Panel Review Meeting

The CIE reviewer shall conduct the independent peer review in accordance with this SoW and the ToRs, and shall not serve in any other role unless specified herein. The CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel. Peer review tasks shall be focused on the ToRs as specified herein. The meeting will consist of presentations by NOAA and other scientists to facilitate the review, to provide any additional information required by the reviewer, and to answer any questions from the reviewer.

Contract Deliverable - Independent CIE Peer Review Report

The CIE reviewer shall complete an independent peer review report in accordance with the requirements specified in this SoW and OMB guidelines. The CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**.

The CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**.

Other Task – Contribution to Summary Report

The CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the ToRs. The CIE reviewer is not required to reach consensus, and should provide a brief summary of views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Foreign National Security Clearance

When a reviewer participates during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for a reviewer who is a non-U.S. citizen. For this reason, the reviewer shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for security clearance purposes. This information shall be submitted at least 40 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: <http://deemedexports.noaa.gov/> and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Location of Peer Review: For the methodology review, the CIE reviewer shall conduct an independent peer review during a panel review meeting scheduled in Santa Cruz, CA in March 2018.

Period of Performance

The period of performance shall be from the time of award through April 30, 2018. The reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewer
Approximately 2 weeks later	NMFS Project Contact provides the pre-review documents to the reviewer
March 2018	The reviewers participate and conduct an independent peer review during the panel methods review meeting
Within two weeks of panel review meeting	Contractor receives draft report
Within two weeks of receiving draft reports	Contractor submits final report to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The report shall be completed in accordance with the required formatting and content; (2) The report shall address each ToR as specified; and (3) The report shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (<http://www.gsa.gov/portal/content/104790>). International travel is authorized for this contract. Travel is not to exceed \$5,000.00.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:

Stacey Miller
National Marine Fisheries Service, 2032 SE OSU Drive
Newport, OR 97365 Stacey.Miller@noaa.gov Phone: 541-867-0535

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
 - a. Reviewer should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
 - b. Reviewer should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
 - c. Reviewer should elaborate on any points raised in the Summary Report that they feel might require further clarification.
 - d. Reviewer shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
 - e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.
3. The reviewer report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of the CIE Statement of Work

Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Review of Catch Estimation Methods in Sparsely Sampled Mixed Stock Fisheries

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.

Annex 3: Tentative Agenda

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