Improving the consistency and transparency of Tier 4/5 assessments

Cole Monnahan, Jane Sullivan, Cindy Tribuzio,

Grant Thompson, Pete Hulson

Introduction

A group of AFSC assessment scientists was tasked to evaluate Tiers 4 and 5 assessment methods in response to assessment approach differences highlighted during the 2020 assessment cycle. Differences in estimating ABC exist, particularly with how data sets were combined and also with the model code itself. We propose adopting modern tools to assist with versioning and ensuring for smoother transitioning to new authors, accurate reviewing of assessments, and increased transparency and reproducibility.

A questionnaire was sent to all groundfish and crab assessment authors to collect information on current practices including what input data, and "random effects" (RE) model was used for Tiers 4 and 5 stocks and complexes (referred to as Tier 4/5 stocks throughout for brevity). This report details the variety of code versions used (i.e., .tpl files), how zero observations have been dealt with in some cases, and how multiple areas and species (in complexes) have been combined. We also discuss how model weaknesses can be exacerbated given the potential that future surveys may be dropped.

The efforts reported below are a continuation of the work by the Survey Average Working Group (2012-2013). That WG explored the statistical properties of several estimation methods in the context of our Tier 4/5 assessments and recommended the random walk model in ADMB-RE, which was adopted. The original RE model was extended to a multivariate version in 2015. The Groundfish JPT provided a framework for implementation in the September 2015 minutes, which laid out both single and multispecies approaches for groundfish assessments. This framework has not been universally implemented, and through the course of multiple assessment cycles, inconsistencies and questions have arisen.

The goals of this work is to:

- Collate and summarize information on different model software versions used and how zeroes are dealt with implicitly or explicitly
- Examine the range of approaches for combining estimates when there is more than one survey for a single species assessment, or a complex with more than one species.
- Examine the way uncertainty is calculated given multiple surveys/species, and how increasing survey reductions would increase uncertainty and the implications for a P* approach to explicitly account for it.

Development of the RE model

As noted above, the survey average working group activities developed the random-walk model (along with a variety of others including Kalman filter). These were written in ADMB and subsequently the code was named "re.tpl". Over time the RE model was adapted to specific stock assessments and sometimes modified (e.g., to incorporate multiple survey indices). The source code and sometimes the executable was rarely tracked for changes and such modifications without any consistent versioning. From a poll, of

the 20 respondents, there were three different versions and a variety of minor variants within those. These are summarized in the following sections.

Single-survey model (RE)

For one version of the RE model, there were some minor differences in output formats, mceval statements (which appear to be unused), and in what phase logSdLam was estimated, despite being the only estimated parameter. These differences will not affect the resulting estimates of the model and are not discussed further. The initial draft repository (posted in September 2013) was experimental with a goal to improve collaboration and transparency. However, this was premature and failed to be regularly maintained. In 2019 a version was posted that differed and persisted in 2020[1]. This version included an inadvertent (and undocumented; it was not the code used in the simulation testing) informative prior (penalty) on the log of the process error (logSdLam):

```
jnll += square(logSdLam+1.5);// Modest penalty to keep process error from getting too large..
```

This penalty will affect biomass estimates depending on the data. For context, this is equivalent to a N(-1.5, SD=sqrt(0.5)) prior on the log of the process error parameter (Fig. 1).
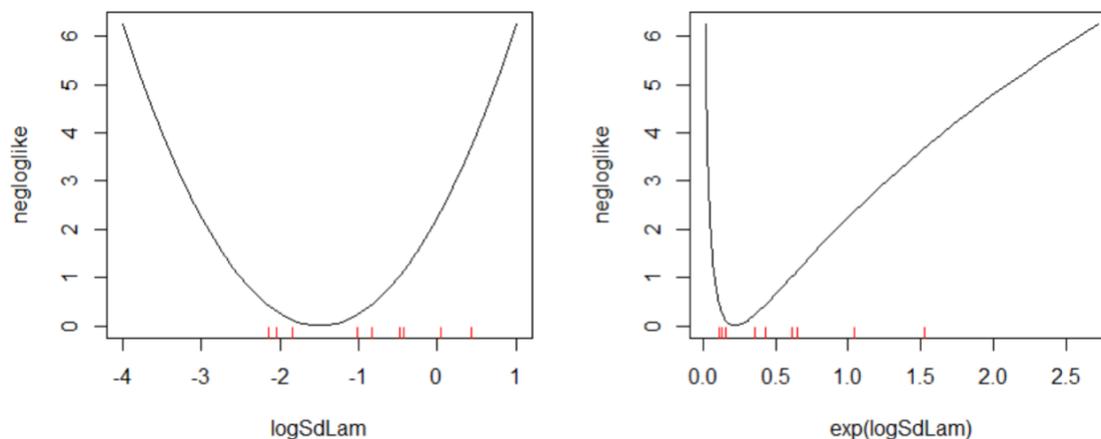
Figure 1. The prior on the log of the process error parameter (logSdLam) used in some versions of the RE model. Line shows the unscaled negative log density, and the red points on the *x*-axis are estimated values of logSdLam from recent assessments from the respondents of the poll (see below for more details). The first plot shows the prior on the parameter, while the second shows it in natural space.

Including a prior on the process error (or the ratio of process and observation error) has been an interesting topic and warrants further study. Relative to what was recently in the repository, this should be checked carefully and ensure that if a penalty is used, it is clearly evaluated and documented.

Multi-area model (REM)

A multivariate extension of the RE model (REM) was developed in 2015, which was meant to simultaneously estimate multiple biomass trends in more than one area, with potential pooling or explicit estimation of parameters determining the amount of regional process error at the discretion of the user. This development was implemented into the RE model in order to increase efficiency in determining apportionment across regions (e.g., instead of running three separate RE models for the three regions in the GOA, the model could run these simultaneously). The REM model also added a feature that allowed flexibility to the process errors if users had reason to believe that productivity of species was different among regions (e.g., in the EGOA where there is no trawling compared to the remainder of the GOA).

---

[1] https://github.com/afsc-assessments/SurveyAverageRandomEffects/blob/master/src/re.tpl

Investigations also revealed that estimating total biomass as the sum of regional biomass in the REM model aided in stability of point estimates as well as uncertainty estimates compared to estimating total biomass alone in the RE model. The REM was used in GOA rockfish and spiny dogfish assessments. We report two noteworthy differences from the univariate models above. First, in the REM version, the logSdLam parameter is bounded between -5 and 2. These bounds are extreme and unlikely to impact estimates but may stabilize optimization. Using a bounded transformed parameter vs. leaving it unbounded but with a prior/penalty may be a topic for further research.

A second concern is how in the REM version, observations of zero biomass are treated compared to missing surveys. In one version there is an 'if' statement that equates to ignoring zero biomass:

```
if(srv_est(i,j)>0) obs(biom(yrs_srv(i),j),i,j);
```

In this statement, if a '0' is defined in the data file then it is subsequently ignored, effectively treating it like a missing survey. In several other versions, this statement is modified to:

```
if(srv_est(i,j)>-1) obs(biom(yrs_srv(i),j),i,j);
```

Where a '0' biomass can be passed to the model and the user must specify the SD, while adding the flexibility to define missing surveys as a negative number in the data file that will be subsequently ignored. In versions where a zero biomass was passed to the model, the likelihood was adjusted by adding a small constant in order to avoid taking the log of a zero value with the statement:

```
jnll+=BTS_wt*0.5*(yconst(i,j) + square(biom-log(srv_est(i,j)+0.0001))/yvar(i,j))
```

This has led to some confusion about how zeroes were dealt with (see below), because this behavior is only apparent by looking at the source code.

REM with additional longline survey (REMA)

An extension of the REM model was developed to include additional index data (REMA; Hulson et al. 2021), in which a scaling coefficient between the bottom trawl survey and the additional index data is estimated. There are two specific versions, one for shortraker and one for shortspine thornyhead. It should be noted that the REMA model is still under development for the species that it is being applied to, and effort to collate the code has been limited.

Versioning summary

In summary, the original univariate version of the RE model was extended to include a wide set of trivial changes and one instance of an informative prior on process error, all of which do not allow zero observations inputs (i.e., the model will crash). It also was extended to the REM model, which, in some cases ignores zero observations, but in others converts them to a small value, and further into two versions that include additional index data. Because *.tpl files are often not included among the model files, it is impossible for authors to know the underlying source code to be able to understand implications for zero observations and other key differences.

Versioning recommendations

Untraceable divergences among model versions used for NPFMC stock assessments is a quality control and transparency issue, which requires a long-term solution as staff change and assessments change authorship. Recently a version was put online into a modern version control system (https://github.com/afsc-assessments/SurveyAverageRandomEffects), and we recommend that:

- This repository be adopted as the official version of the model.
- The RE and REM models be merged into a single one that has unified and consistent inputs and outputs that are accessible via a spreadsheet/text editor, or programmatically via R.
- All models should also be version controlled.
- Updating the naming conventions used in the input and output files

The advantages of this approach are that it will unify the modelling approach for most Tier 4/5 assessments and apportionment. The official version will be unambiguous and that everyone has easy access to it, and that any future changes are documented, traceable, and easily disseminated to the whole group. Finally, it will provide a starting place to bridge to alternative or updated models in the future.

Zero biomass observations

A zero biomass observation occurs when the species of interest is not observed in a survey, resulting in an estimated biomass equal to zero and no associated uncertainty estimate. Zero biomass observations occur for low abundance or rare species, species with patchy or localized distributions (e.g. spiny dogfish in the western GOA; Tribuzio et al. 2020), and species that aren't well-sampled by the survey (e.g. harlequin rockfish in the GOA and BSAI that are associated with untrawlable habitat; Tribuzio and Echave, 2019; Sullivan et al., 2020). It is important to contrast a zero biomass observation with a missing survey, where no sampling occurs in an area. For example, this occurred in 2001 when the survey did not sample in the EGOA, and has occurred in several years when the 501-700 m and 701-1,000 m depth strata were not sampled in the GOA trawl survey.

Zero biomass observations are problematic in the context of the RE model because the model estimates biomass in log-space, and therefore does not accept zeros. The RE model also requires inputting a CV value, which is not available for a zero biomass observation. All authors with zero biomass observations reported that they treat these observations as NAs, either to maintain continuity with past methodology, or because they considered zeros a "failed survey" (i.e., they do not believe there is truly zero biomass; Table 1). Due to the inconsistency in the treatment of zeros in the RE tpl versions, authors arrived at these NAs in different ways. In the univariate RE tpl, zeros are not filtered out, and the model will return a NaN when it takes log(0). In contrast, the REM tpl filters out zero observations internally, with the option to ignore zero biomass or include it with a manually defined variance, which could lead to confusion about the treatment of zeros for authors who strictly use the executable file.

While all stocks with zeros currently treat zeros as NAs, by either removing them manually or using a model which filtered them out, alternative methods could be used. In the 2018 BSAI Other Rockfish assessment, there were 12 zero biomass observations of non-shortspine thornyheads (non-SST) in the EBS shelf survey. Spies et al. (2018) explored different methods for accounting for these zeros, including: (1) treating them as NAs, (2) replacing zeros with biomass = 0.1 t and variance = 0.001, and (3) replacing zeros with biomass = 0.01 t and variance = 100. The RE-estimated biomass in the terminal year was 720 t for scenario 1, and greater than 1,500 t in scenarios 2 and 3. Ultimately, the stock assessment authors recommended replacing zeros with NAs, in accordance with other Tier 4/5 assessments, which was endorsed by the review bodies.

Zeroes recommendations

The results of the April 2021 questionnaire showed that all Tier 4/5 assessments that have zero biomass observations treat these as NAs. However, this fact was not clear to some authors because it was occurring in the tpl without their knowledge. With the aim of increasing transparency on this issue, we recommend that:

- Authors use standardized RE software that does not filter out zeros automatically. Under this scenario, the software will produce an error if zeros are used in the input, which will compel authors to be explicit about the treatment of those data.
- Authors include in their SAFE chapters information about the zero biomass observations (e.g. when and where they occurred) and the method used to handle this zero (e.g. input as NA value).

Though outside of the scope of this report, we acknowledge that alternative modeling frameworks exist that more naturally handle zero values (e.g., delta-models like VAST). To our knowledge, only one Tier 5 assessment, eastern GOA walleye pollock, has attempted to apply the VAST model to survey biomass

observations[2]. We support further exploration of these alternative methods for Tier 4/5 assessments but note that considerations for errors in ABC estimates be explicitly considered in the diagnostics.

Table 1. Tier 4/5 assessment author responses to the April 2021 questionnaire about the treatment of zero biomass observations.

How are zero biomass observations treated?		
No zeros	**Remove zeros (i.e. treat them as NAs)**	**Other (e.g. replace zeros with small value)**
Pribilof Islands red king crab Pribilof Islands blue king crab Pribilof Islands golden king crab AI Pacific cod GOA shallow water flatfish EGOA pollock BSAI Bogoslof Island pollock BSAI shortraker GOA shortraker GOA thornyheads	BSAI other rockfish BSAI other flatfish BSAI skates GOA skates GOA sharks[3] GOA other rockfish	None

Approaches to fitting to multiple surveys

Many Tier 4/5 assessments in the GOA and BSAI use multiple surveys or survey regions; however, the methods used across assessments use different RE model versions (e.g. univariate vs. multivariate) and survey time series. For the assessments that reported using multiple surveys, three fit each survey/survey region index in separate RE model runs using the univariate RE tpl, and five fit multiple survey/survey region indices simultaneously in the same RE model run using either the REM model or its extension that includes the longline survey (REMA; Table 2). The implications of these different approaches arise when calculating the variance of the summed biomass estimates. This issue is further explored in the following section, "Calculating the variance of summed biomass estimates."

---

[2]https://meetings.npfmc.org/CommentReview/DownloadFile?p=7c7874b1-2da3-4e02-88a9-90dd9291a71b.pdf&fileName=PRESENTATION_pollock%20eastern%20GOA_draft_assessment_PTSept2020.pdf

[3] The only Tier 5 species in the GOA sharks complex is spiny dogfish, the rest are Tier 6.

Table 2. Tier 4/5 assessment author responses to the April 2021 questionnaire about the approach used when fitting to multiple surveys or survey areas.

If multiple surveys or survey areas are used, which best describes your workflow?	
Fit each index in separate RE run (univariate RE)	**Fit multiple indices in same RE run (multivariate RE)**
BSAI other flatfish	BSAI other rockfish
BSAI skates	GOA thornyheads
GOA skates	GOA shortraker
	GOA sharks
	GOA other rockfish

Responses from the questionnaire also showed differences in the survey time series used across assessments, although a review of the rationales behind these disparate time series was beyond the scope of this report. For assessments that use the GOA trawl survey, five begin the time series in 1984 and two begin in 1990 (Table 3). The two assessments that fit to the longline survey, GOA thornyheads and GOA shortraker, begin the longline time series in 1992 (Table 3). Survey time series used from the BSAI region varied considerably. Authors reported four different starting years for the AI trawl survey (1980, 1986, 1991, 2000) and EBS shelf trawl survey (1975, 1982, 1987, 2000), respectively (Table 4). These inconsistencies highlight a potential need to review and document the rationale behind the existing survey time series used across assessments (Tiers 1-6).

Table 3. Tier 4/5 assessment author responses to the April 2021 questionnaire about the survey time series for the GOA trawl and longline surveys.

GOA only: Which surveys are used, and what are the start and end years for each survey index (as of the 2020 assessment cycle)?		
Stock	**GOA trawl survey**	**Longline survey**
GOA shallow water flatfish	1984-2019	
EGOA pollock	1990-2019	
GOA skates	1990-2019	
GOA sharks	1984-2019	
GOA thornyheads	1984-2019	1992-2020
GOA shortraker	1984-2019	1992-2020
GOA other rockfish	1984-2019	

Table 4. Tier 4/5 assessment author responses to the April 2021 questionnaire about the survey time series for the trawl surveys in the BSAI.

Stock	AI trawl survey	EBS shelf survey	EBS slope survey	Bogoslof acoustic trawl survey
BSAI only: Which surveys are used, and what are the start and end years for each survey index (as of the 2020 assessment cycle)?				
BSAI other rockfish	1991-2018	1982-2019	2002-2016	
BSAI other flatfish	1986-2018	1987-2019	2002-2016	
BSAI skates	2000-2018	2000-2019	2002-2016	
Pribilof Islands red king crab		1982-2019		
Pribilof Islands blue king crab		1975-2018		
Pribilof Islands golden king crab			2002-2016	
AI Pacific cod	1991-2019			
BSAI Bogoslof Island pollock				1991-2020
BSAI shortraker	1980-2018		2002-2016	

Surveys recommendations

Considering the different approaches used to fit multiple survey inputs for a single stock, we recommend:

- A consistent, well documented, tested, and understood source code be adopted
- The REM model as described would be preferred because it is simpler and clearer to use (only one model run) and has expanded options for calculating uncertainty of the total biomass (see below).

Questionnaire results also found that assessments used a range of different start years for GOA, EBS shelf, and AI trawl surveys. Although outside the scope of Tier 4/5 models, the rationale for these varying start years is unclear and warrants review.

Stock complexes

There are seven groundfish complexes managed under Tier 4/5 harvest control rules in the GOA and BSAI, which represent a diverse array of flatfish, skates, and rockfish. Species contained within a single complex can be hard to distinguish from one another or have divergent life histories, making it sometimes necessary to group species but also potentially problematic.

Questionnaire results suggest that methods for lumping or splitting species when estimating biomass within a complex are as diverse as the complexes themselves. For example, authors reported that they grouped species based on similar natural mortalities (M), low biomass or low sample sizes, limitations due to species identification, or in the case of the GOA shallow water flatfish species, not grouped at all (i.e., each species' biomass is estimated separately; Table 5). These groupings can have consequences; for example, if all species within a group are trending positively except for one (e.g., BSAI other flatfish assessment; Monnahan 2020). This issue would be exacerbated if the downward-trending species had lower productivity traits (e.g., slow growing, late to mature, long-lived) than the other species in that group.

Table 5. Tier 4/5 assessment author responses to the April 2021 questionnaire about the methods used to group species within a complex. Respondents were allowed to select all that apply.

Complexes only: How do you group the species within the assessment?			
Grouped by natural morality	**Lumped due to low biomass or low sample sizes**	**Lumped together due to species ID issues**	**Not grouped (all species estimated separately)**
BSAI other rockfish BSAI other flatfish GOA other rockfish	BSAI other rockfish BSAI skates GOA thornyheads[4] GOA skates	BSAI other rockfish BSAI skates[5] GOA skates	GOA shallow water flatfish

During our review of complexes, we also found that authors used different methods to calculate input biomass variances and reference points. This finding is best illustrated by contrasting the assessment workflows for GOA and BSAI other rockfish assessments (Tribuzio and Echave 2019, Sullivan et al. 2020; Table 6). First, the GOA other rockfish authors calculate new input variances at the species M-group level using haul-level data from the GOA trawl survey. This method replicates the design-based estimator used by the Groundfish Assessment Program (GAP) but treats each species within a group as the same species. The RE model is then fit to these custom design-based estimates of survey biomass for each species group. These RE biomass estimates are used to weight the natural mortalities of each species group, thus creating a "weighted M" for the entire complex. Next, the haul-level data is used to create a time series of survey biomass estimates and their variances for the entire complex for each tier, combining all species within the complex and same tier. The RE model is fit one final time to the new complex-wide survey biomass. Using the RE output of this model run, the complex-wide weighted M is used to calculate ABCs/OFLs under the Tier 4/5 harvest control rules.

The GOA other rockfish assessment is the only assessment utilizing this approach. Based on the results of the questionnaire, the other complexes that group multiple species within the assessment (BSAI other rockfish, BSAI other flatfish, BSAI skates, GOA skates) do so by summing the design-based biomass and variance estimates provided by GAP (Table 6). Reference points are calculated for each species group using species group Ms, and then these results are summed to the complex-wide level (Table 6).

---

[4] Only shortspine thornyhead data are included in the stock assessment

[5] The RE model is run separately for each species except for a "minor skates" group that are rare and have a patchy time series

Table 6. A generalized comparison of the workflows used in the GOA and BSAI other rockfish assessments to calculate input variances, Acceptable Biological Catches (ABC), and Overfishing Limits (OFL). A contrived example is used with species grouped based on natural mortality. Note that this workflow is simplified and does not account for the spatial complexities of each assessment.

Fake example Tier 5 complex has 5 species split into 2 species groups based on natural mortality (M)	
Group A: M=0.1 Species 1 Species 2 Species 3	**Group B: M=0.2** Species 4 Species 5
GOA OTHER ROCKFISH WORKFLOW	**BSAI OTHER ROCKFISH WORKFLOW**
Step 1: Calculate new group-level variances using haul-level data (i.e., treat all species within a group as the same species and calculate new variance using design-based estimator/GAP method) Step 2: Run RE model for each group using variances from Step 1 Step 3: Calculate a weighted-average of M based on RE model biomass results from Step 2 Step 4: Calculate new variance for the entire complex using haul-level data (i.e. repeat Step 1 combining Group A and Group B) Step 5: Run RE model for the entire complex combined using variance from Step 4 Step 6. Calculate final ABCs and OFLs for the entire complex using the weighted-average M from Step 3 and the RE biomass results from Step 5	Step 1: Sum species-level GAP variances to get group-level biomass and variances Step 2: Run RE model for each group Step 3: Calculate ABCs and OFLs for each group using their respective Ms and RE biomass estimates from Step 2 Step 4: Calculate final ABCs and OFLs for the complex by summing group-level ABCs and OFLs from Step 3

Species complex recommendations

Despite the complexity of the GOA other rockfish assessment workflow, one clear advantage is that it allows the inclusion of rarely observed species that have no GAP-estimated biomass (e.g. blackgill and chilipepper rockfish in the GOA). While there is clear merit in this approach, and it follows the JPT framework laid out in 2015, the contrast in methodologies highlights a lack of consistency across assessments. We recommend further analysis of the implications of the alternative approaches outlined in Table 6. Specific objectives of this analysis include the following:

1. Verify that the custom design-based estimator applied in the GOA other rockfish assessment is the same as GAP's design-based estimator.
2. Evaluate the differences in variance estimates between the summed GAP and custom design-based estimator.
3. Quantify the differences in estimates between the GOA other rockfish's weighted-M approach and the method of summing species group ABCs to the complex level.

We recommend that the results of objectives 2 and 3 be tested across multiple stock complexes, and use the results to suggest a unified approach.

Calculating the variance of summed biomass estimates

The uncertainty estimates from the RE model are not currently used in Tier 4/5 groundfish assessments as a means to adjust any management recommendations. However, there are some potential issues with calculating the variance for summed biomasses, which would also apply to Tier 3 and other situations. The ultimate goal is to calculate confidence intervals for the total biomass (i.e., summed across areas). In development of the REM model, the complication of estimating variance for the sum of lognormal random variables was initially identified, as this model estimates total biomass as the sum of regional biomass estimates.

Briefly, the REM model estimates $X_1(y),..,X_n(y)$ log-biomasses in year y (e.g., n=4 if using 4 survey areas). We assume that $X_y \sim N(mu_y, sd_y) = N(biomsd(y), biomsd.sd(y))$. Note that $B_y = \exp(X_y) = biomass$ in year y follows a lognormal distribution. Ignoring the year, what is needed is the distribution of the log of total biomass so that an assertion of normality can be made and confidence intervals calculated, which are then exponentiated to get uncertainty for total biomass. Specifically,

```
Var(log total biomass)=Var( log(B₁(y)+..+Bₙ(y)) )
```

Because of the non-linearities caused by exponentiation, summation and logging, there is no analytical way to calculate this. There are at least four ways to approximate this variance. The first is from Marlow (1967), obtained by matching the moments to another lognormal distribution (see https://en.wikipedia.org/wiki/Log-normal_distribution#Related_distributions for details), namely:

$$\sigma_Z^2 = \ln\left[\frac{\sum e^{2\mu_j + \sigma_j^2}(e^{\sigma_j^2} - 1)}{(\sum e^{\mu_j + \sigma_j^2/2})^2} + 1 \right]$$

This is the approach used by the REM model currently, implemented as

```
dvar vector SD numer =
  rowsum(elem_prod(exp(2*biomsd+square(biomsd.sd)),(exp(square(biomsd.sd))-1)));

dvar_vector SD_denom = square(rowsum(exp(biomsd+0.5*square(biomsd.sd))));

dvar_vector SD_biom_TOT = sqrt(log(elem_div(SD_numer,SD_denom)+1));
```

The second approximation is via the standard Delta method implemented in ADMB by using an sdreport_number. The third is to skip this variance calculation and instead sum the lower and upper limits of confidence intervals of each of the *n* areas/surveys (referred to below as the "additive" method). Finally, the fourth is to use integration via MCMC, which is an approximation only in the sense of using a finite number of samples for estimation, but whose approximation error goes to zero as the number of posterior samples goes to infinity. These approximations are demonstrated with the 2020 BSAI other rockfish non-SST assessment (Fig. 2).
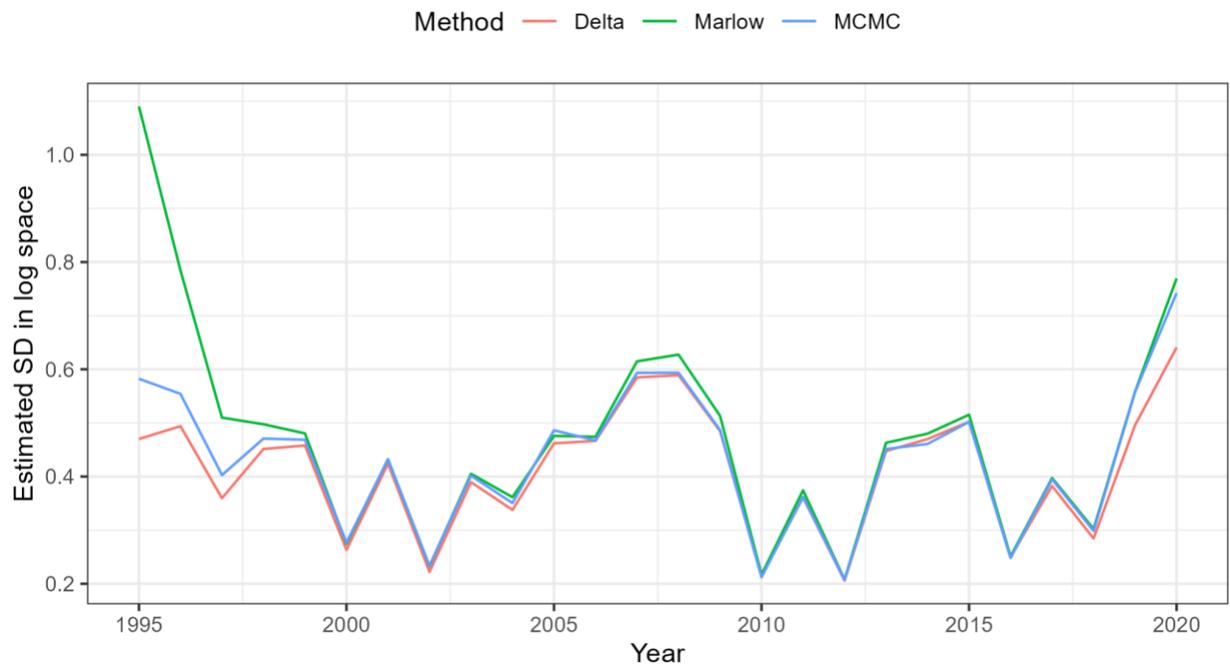
Figure 2. Results from different methods for approximating the SD of total biomass in log space in the RE model for BSAI other rockfish non-SST. The model starts in 1982 but early years are left off for visual clarity.

These discrepancies are exaggerated for estimates before 1995 and left off for visual clarity in Fig. 1. For this example, the Marlow estimates are higher and the Delta-Method estimates lower than MCMC respectively. It is not clear whether this would be the general case. The magnitude of difference between methods appears to depend on the relative sizes of the uncertainties by the areas, which are greater in the earlier years and 2020 (Fig. 3).
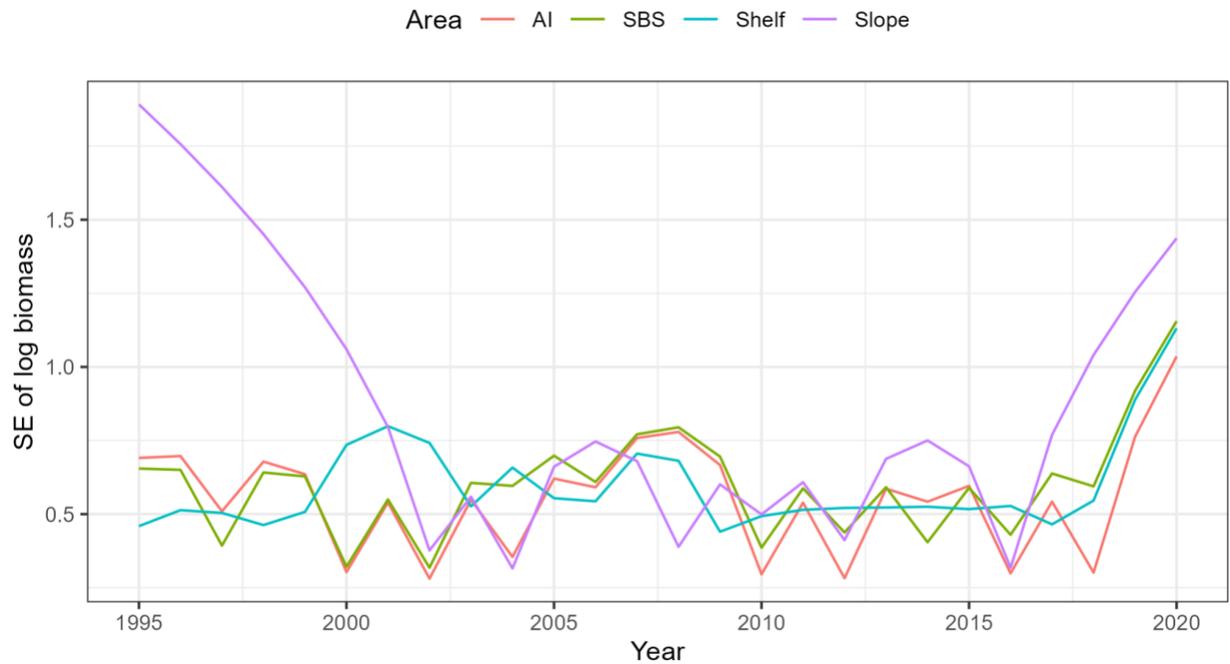
Figure 3. Estimated SE from design-based estimates of individual areas used as inputs to the REM model for the BSAI other rockfish non-SST model. The model starts in 1982 but early years are left off for visual clarity.

Even if the variance estimates are accurate, there is a second important assumption to consider, namely that log total biomass is normally distributed. This assumption allows calculation of symmetric confidence intervals and which are then exponentiated to get asymmetric confidence intervals for the total biomass (natural scale). This is not necessary for Bayesian credible intervals, which are calculated from quantiles without any distributional assumption. Thus, comparing confidence and credible intervals gives clues to differences among these methods (Fig. 4).
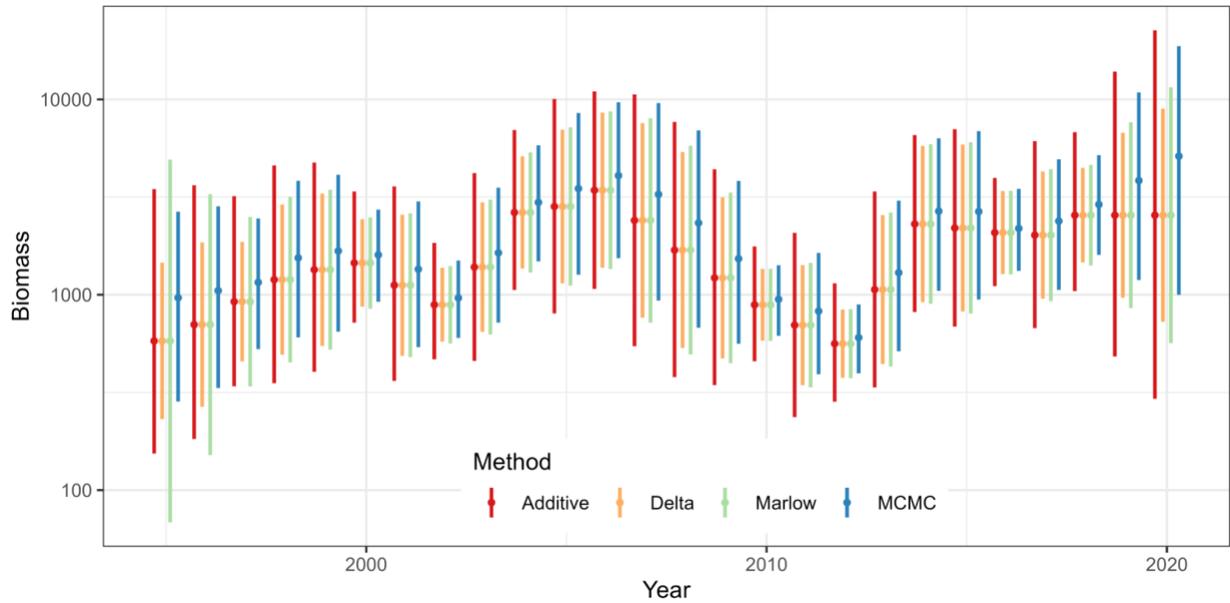
Figure 4. Uncertainty estimates of total biomass (summed across areas) from BSAI other rockfish non-SST using different approximation methods. Vertical lines show the estimated 95% confidence (Additive, Delta and Marlow methods) and 95% credible intervals (MCMC method) with means (points). Note the log10 *y*-axis. The model starts in 1982 but early years are left off for visual clarity.

A more direct way to check for normality is to use QQ-plots on the posterior draws from MCMC for log of total biomass for different years (Fig. 5). All the distributions have longer right tails than expected, while 1982 also has a thinner left tail. Therefore, it appears that for this example the distributions are not normal, so it is no surprise that the confidence and credible intervals are different.
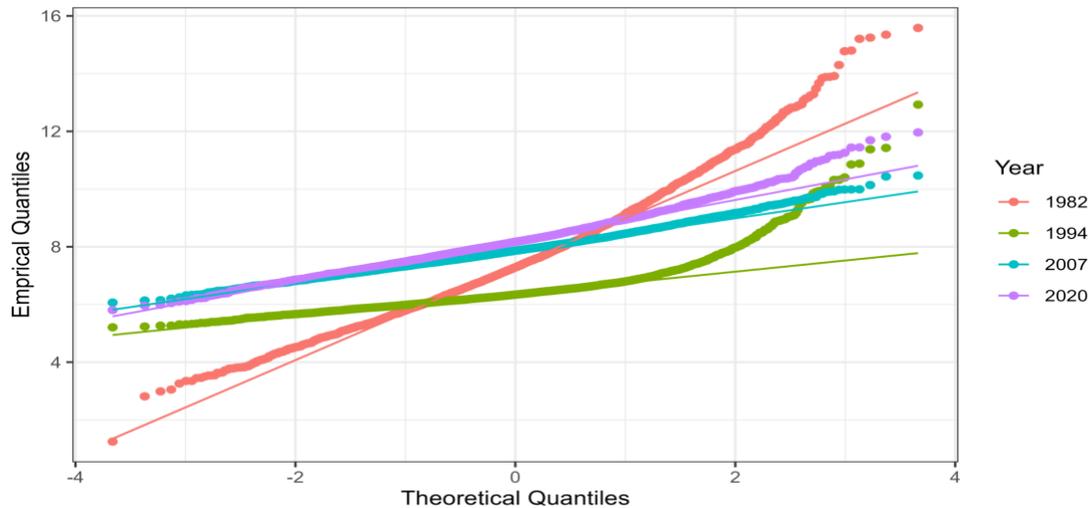
Figure 5. Quantile-quantile (QQ) plots of the posterior distribution of log total biomass for four different years. The points are the observed relationship, while the line shows the expectation if the posterior were truly normal.

In summary, the REM model forces an explicit acknowledgement of how to calculate confidence intervals for the total biomass. The approximation of Marlow (1964) is like the Delta method and MCMC for many years, but not all. The additive approach of summing lower and upper confidence limits appears to lead to the largest confidence intervals in most years. It is unclear which approach would be preferred in general.

The more important finding here is that there are some meaningful violations of the assumption of normality when summing biomasses and then logging them. In general, the **sums of lognormal random variables are not lognormal.** They have no closed-form distribution as reported on in the statistical literature. Studies on their properties is an active area of research (e.g., Torres-Hurtado et al. 2021). In many circumstances, the distribution will be approximately normal and close enough for our purposes. However, only one model was examined and clear evidence was found that, in some years, the distribution was far from normal. For Tier 4/5 assessments, and other applications of the REM model, this is unlikely to be a concern because uncertainty is not presently factored into ABC advice.

We note that summing lognormal random variables is common in our assessment processes, but is less obvious than in the REM model. **This will also occur in any model that sums design-based estimators** (say between areas or species) with a stock assessment model that assumes lognormal likelihoods (presumably all Tier 1 and 3 models). The distribution of these indices will not be strictly lognormal, by definition. Further, the design-based estimators are calculated by summing estimates and variances from distinct areas, which again is potentially problematic. It is unclear when this would matter statistically (and thus to management), but seems a potentially worrisome, albeit technical issue.

Variance Recommendations

We recommend further investigation into the statistical properties of summed lognormal variables and the potential implications and solutions for use in assessments. This could be done in collaboration with RACE staff involved in the design-based calculations.

Accounting for missing and suspended surveys with a P* approach

Survey cancellations (e.g., COVID-19) and planned reductions will cause increasing uncertainty in biomass estimates as time goes on, and the current approach does not account for this. When no data are available, the RE model will predict a stable biomass, and thus reference points, but with increasing uncertainty in each year. This uncertainty may need to be a consideration in the future. Other management councils and the crab plan team incorporate a P* approach to implementing a buffer based on uncertainty to reduce catch. This approach explicitly acknowledges the increasing probability of overfishing with increasing uncertainty, and counteracts this by specifying a larger buffer (e.g., Prager et al. 2011).

Quantifying potential effects to reductions

Two different scenarios were used to investigate the potential effects of using a P* approach for Tier 4/5 stocks. The first was a forecast to see how management advice would change under different P* values assuming no future surveys for 10 years, under a fixed set of P* values (Fig. 6). Although a fixed set of P* values were used in the calculations, an alternative would be to use species-specific P* values, set equal to the probability of overfishing associated with the standard 25% buffer between OFL and maxABC.
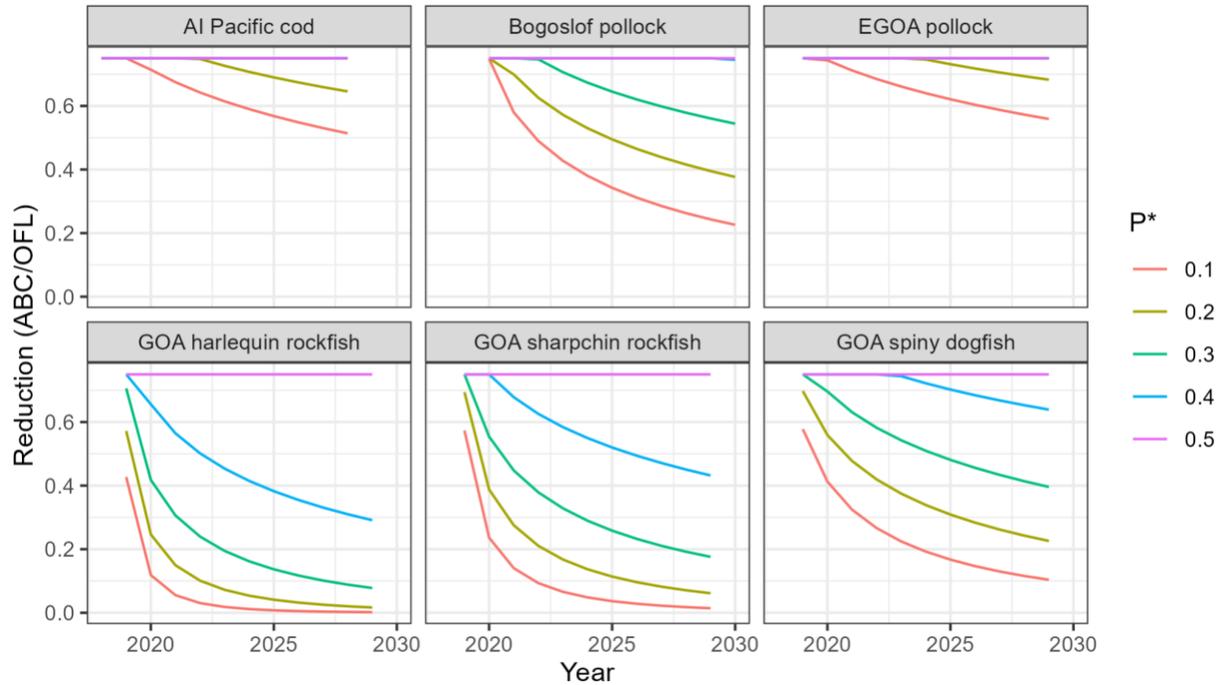
Figure 6. The reduction under different fixed P* values (colors) assuming no future surveys for select species. Projections were done using provided estimates of uncertainty from the most recent assessments.

While unrealistic, this scenario quantifies the upper extreme of buffers that could be applied. In general, the effects on recommended ABC would depend highly on the specific stock. For these stocks a P* level of 0.5 provides no change in buffer in this scenario, and for AI Pacific cod and the two pollock stocks even a P* of 0.4 has no effect. In contrast, some of the stocks would experience extreme reductions in ABC at P* = 0.4 or 0.3.

The above analysis examined the effect on single-survey, single-species stocks. However, the multi-survey assessments can have various combinations of reductions and resulting effects on the buffer. Thus an additional, and more nuanced, analysis was done by taking some specific multi-survey stocks and dropping only some surveys (e.g., the EBS slope) into the future. BSAI other rockfish (non-SST) and BSAI other flatfish were used as case studies, where only the former has appreciable biomass on the slope. The EBS slope survey was dropped from 2010 onward and the estimated biomasses were compared to the original model with the EBS slope data (Fig. 7).
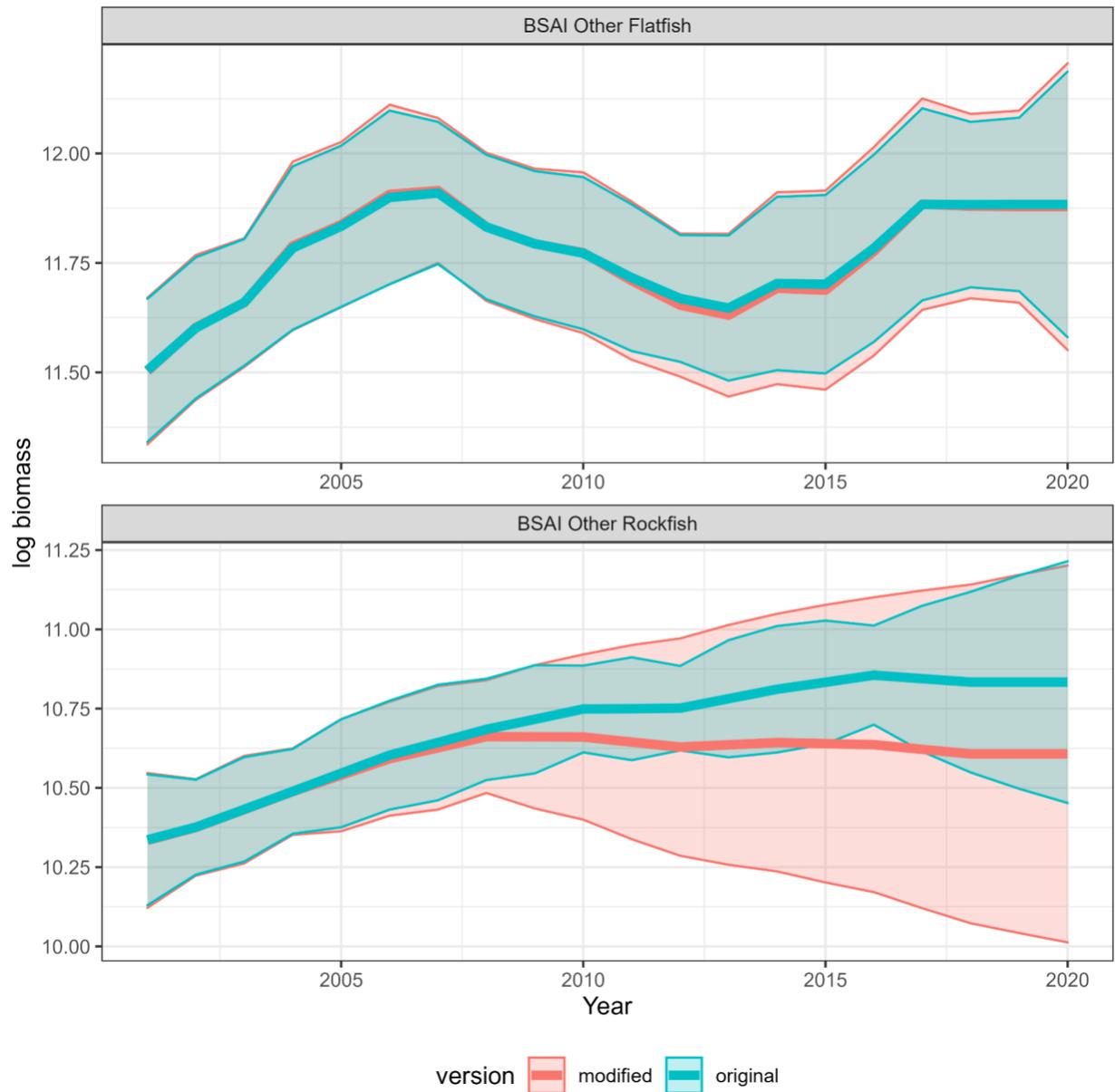
Figure 7. The effect on estimated biomass when dropping 10 years of EBS slope survey data ('modified') compared to the model with full data ('original'). Estimates (thick lines) and 95% confidence intervals (shaded regions) are denoted by color.

Here the effect of dropping ten years of slope data has the expected effect: strong trends on other rockfish and a minimal effect on other flatfish. In this case there is no change to the buffer for P* levels above 0.2.

Conclusions

We explored the potential effects of implementing a P* approach to adjusting the buffer due to uncertainty in the estimated biomass. We found that for most stocks, a level of P* > 0.3 would have a minimal impact on the buffer even under extreme conditions of survey loss. The effect on single-survey stock assessments is easy to explore, but multi-survey ones are more nuanced and should be further explored.

One outcome of this approach is that, for planned biennial surveys, the uncertainty is expected to oscillate up and down. If an equivalent oscillation in the buffer, and hence ABC, is deemed undesirable, the approach could be modified in various ways; for example, by using it only in the event that the amount of time between surveys exceeds some specified number of years.

P* recommendations

The P* approach is already used for crab stocks, and could be a good way to formally incorporate this uncertainty into management advice for groundfish stocks as well. A discussion of the implications of reduced surveys on Tier 4/5 assessments, and whether further analyses on this approach should be investigated, is thus strongly recommended. More thorough analyses of the effects on management could be explored in the future and help guide discussion and implementation of this approach, if deemed appropriate or of potential interest.

Overall conclusions

The questionnaire and analyses conducted to date have highlighted important topics for continued discussions as well as critical modeling differences amongst assessments implementing random effects models. Certainly, there are differences among the assessments, but many of these differences could be mitigated. A more consistent, transparent, and reproducible approach across groundfish Tier 4/5 and crab Tier 4 assessments would streamline the assessment process, benefit current and future authors, and make reviews more productive and helpful. The recommendations outlined in this report will help in the development of an expansion of the 2015 random effects protocol, which addresses questions that have arisen since its implementation, as well as investigations into larger questions applicable to the incorporation of uncertainty in lower tier assessments.

References

Echave, K. B. and P. J. F. Hulson. 2019. Assessment of the Shortraker Rockfish in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2020. North Pacific Fishery Management Council, 605 W. 4th. Avenue, Suite 306, Anchorage, AK 99501.

Echave, K. B., P. J. F. Hulson, and K. A. Siwicke. 2020. Assessment of the Thornyhead stock complex in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2021. North Pacific Fishery Management Council, 605 W. 4th. Avenue, Suite 306, Anchorage, AK 99501.

Hulson, P.-J.F., Echave, K.B., Spencer, P.D., and Ianelli, J.N. 2021. Using multiple indices for biomass and apportionment estimation of Alaska groundfish stocks. U.S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-414, 28 p.

Marlow, NA. (Nov 1967). "A normal limit theorem for power sums of independent normal random variables". *Bell System Technical Journal*. **46** (9): 2081–2089. doi:10.1002/j.1538-7305.1967.tb04244.x.

Monnahan, C.C. 2020. Assessment of the Other Flatfish stock complex in the Bering Sea/Aleutian Islands. In: Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands as projected for 2021. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Prager, M. H., C. E. Porch, K. W. Shertzer, and J. F. Caddy. 2003. Targets and limits for management of fisheries: a simple probability-based approach. North American Journal of Fisheries Management 23:349-361. https://doi.org/10.1577/1548-8675(2003)023<0349:TALFMO>2.0.CO;2

Spies, I. B., TenBrink, T., Spencer, P. D., Palsson, W., and Nichol, D. 2018. Assessment of the Other Rockfish stock complex in the Bering Sea/Aleutian Islands. In: Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands as projected for 2019. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Sullivan, J., Spies, I., Spencer, P., Kingham, A., TenBrink, T., and W. Palsson. 2020. Assessment of the Other Rockfish stock complex in the Bering Sea/Aleutian Islands. In: Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian Islands as projected for 2021. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Torres Hurtado J. G., Bustamante R., & Caicedo C. E. 2021. A fast method to seek the mean and variance of the summation of lognormal variables. Revista Facultad De Ingeniería Universidad De Antioquia. Retrieved from https://revistas.udea.edu.co/index.php/ingenieria/article/view/343943

Tribuzio, C. A. Matta, M. E., Echave, K., Rodgveller, C. 2020. Assessment of the shark stock complex in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2021. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Tribuzio, C. A. and Echave, K. B. 2019. Assessment of the Other Rockfish stock complex in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2020. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Turncock, B. J., M. Bryan, and T. K. Wilderbuer. 2017. Assessment of the shallow-water flatfish stock complex in the Gulf of Alaska. In: Stock Assessment and Fishery Evaluation Report for Groundfish Resources in the Gulf of Alaska. North Pacific Fishery Management Council, Anchorage, AK, USA.