Center for Independent Experts External Independent Peer Review Report

AFSC Approaches to Survey Biomass-based Stock Assessments

Colin Millar

April 2023

Table of Contents

Executive Summary

The review focuses on a method for estimating biomass from multiple surveys (absolute and relative) for management advice for Tier 3, 4 and 5 stocks and for regional apportionment of acceptable biological catch (ABC). The review is not an in depth review of applications of this method to specific datasets and the model settings, data choices and options used therin.

The assessment model is capable of combining data from multiple surveys to estimate trends in biomass for components of a stock complex defined in terms of strata based on region and, for example, depth. The model provides estimates of uncertainty for model parameters and produces confidence intervals for trends in biomass and apportionment quantities, however, as far as my understanding extends, uncertainty is not used in determining management advice for Tier 4 and 5 assessments, or in the apportionment in Tier 3 assessments, unless, that is, bias correction is performed.

The underling random walk model used to model biomass is a choice used widely in fisheries stock assessment to model population and fishing trends over time, and seems well suited to the purpose of modelling trends in biomass while providing stable estimates of biomass in the terminal years.

The log normal distributional assumption is likewise well used in fisheries, however, care should be taken if moving to another distribution that the same aspect of the distribution of biomass is being modelled (i.e. the median versus the mean).

Given the variety of implementation of this method in the past, important aspects to consider are, a central location for development ideally with version control, the usability of the new method to encourage uptake, the ease with which developers and practitioners can collaborate on, explore and test new developments.

Background

The National Marine Fisheries Service (NMFS) is tasked with the responsibility of preserving, safeguarding, and regulating the USA's marine ecosystem through the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act, using the best available scientific data. As a result of the often controversial nature of NMFS science products, they require strict scientific peer reviews that are independent of any external influence. The credibility of the agency's scientific products is ensured through a formal external process for independent expert reviews, which is critical to reinforcing scientific quality assurance for fishery conservation and management actions.

The assessment of fish stocks in Alaska federal waters, when age or length-based stock assessment data is insufficient, relies on expoitable biomass estimates obtained from surveys and estimates of natural mortality. In the past, there has been wide variation in catch advice based on such estimates, but a "random effects" model has been recommended for use by the NPFMC since 2015. This model has been tested through simulations and expanded to other aspects of managing fish stocks. As this model has significant implications for fisheries management, it is important that it is based on sound statistical methods and represents the best available science.

Given that this is a review of a method and not of its application, the review will not assess specific dataset issues, but look at whether the model and modeling framework is appropriate for the datasets it is developed for, and will comment on the useability and utility of the software proposed.

Role of Reviewer

My role as a CIE independent reviewer was to conduct an impartial and independent peer review in accordance with the SoW (Appendix 2) and the predefined ToRs (Annex 2).

On the 24th of January background documents for the AFSC CIE review were made available via a google drive. The documents made available are listed in the Appendices. I reviewed the documents I was provided. An online webinar was held on the 13th of February with the primary purpose to discuss the ToRs (provided in Annex 2) and the review process. At this meeting, additional discussions took place on some of the issues raised in the background documents such as the prevalence of zeros in the survey data used in the assessments, that confidence intervals for the estimates of the biomass trend are not currently used in advice, and some explanation was given on the decisions behind the formation of stock complexes.

In reponse to questions at the webinar the following was provided:

- A list of stock assessments, which model was used, and for what purpose, a long with details of 'characteristics' of the stock assessmemnt. Also included was the links to the reports for the stock assessment. This was a very an useful resource.
- A list of surveys and relavent details for each
- Details on apportionment
- Information on zero observations, where they are most prevalent and links to the background documents where these are discussed.
- Links to variance of biomass estimates and the P* approach
- A link to a detailed explanation of the tier system was provided

The report is formated according to my understanding of the Peer Review Report Requirements given in Annex 1.

Summary of Findings

This section presents a summary of findings for a review the current data-moderate stock assessment methods used in the North Pacific, specifically related to survey averaging methods.

ToR 1. Evaluate survey-averaging methods used to support survey biomass-based approaches for data-moderate stocks in the North Pacific.

The “random effects” model that is the subject of this review, can be described in its simplest form as smoothing through the annual exploitable biomass estimates / observations, taking an estimate of the error in those estimates as known. Therefore, this simplest form of the model produces a time series, or trend, of exploitable biomass and an estimate of the variability of this trend with respect to time. In 2015 several ideas were considered for how to model the underlying biomass trend, with the simple first order random walk being the model that has been taken forward and is the basis of the proposal under review.

Additional complexity has been added to this base model:

- multiple strata can be modeled at one time for a single stock, each having its own estimated biomass trend and associated variability.
- surveys that to not measure absolute biomass can be incorporated, to provide additional information on the biomass trend. In this case a ‘catchability’ parameter is estimated (by strata if appropriate).
- Often, models with fixed observation variance, can result in overfitting to the data, and to alleviate this problem an additional observation error can be estimated.
- In some surveys the presence of zero observations can cause issues, and the proposal introduces the idea of modelling the zeros explicitly.

In total, to summarise, the full model can be written approximately in an alternative concise form as a linear (random effect / latent variable) model:

```
log_index ~ s(log_biomass):strata + strata:I_index=cpue + re(strata:index)
```

where log_observation is the observation of biomass, either an absolute estimate from a survey or an estimate of CPUE. `s()` is a random walk, `I_index=cpue` is an indicator variable for whether the observed index is a CPUE index or not. `:` indicates an interaction, so that `s():strata` denotes a different smooth for each strata, and, `strata:I_cpue` is a catchability parameter by strata that is zero for absolute survey indices. All observations can be thought of as being inverse variance weighted with overall residual variance fixed to 1. If additional variance is required this can be thought of as adding a random effect `re()`, potentially with a different variance for each strata and survey.

Setting out the model in this form, does hint at utilising alternative modelling routes, such as GLMMs, GAMMS (Wood 2017), or so called STAR (structured additive regression) (Fahrmeir et al. 2022), which can be fitted in a multitude of ways in packages such as mgcv, glmTMB, INLA, and so on. The additional benefit of using these other packages is that distributional assumptions and investigations into structural model assumptions would be easier to explore. One such example is the switch from modelling log survey index to modelling the survey index directly but using the log link, and then investigating the use of the Tweedie distribution.

Strengths

The model is a pragmatic mix of simple and complex and addresses most of the data issues raised in the background documents.

- Multiple strata (typically area and depth) are dealt with by estimating a separate biomass trend for each, optionally forcing all biomass trends to have the same year to year variability.
- By fitting these in the same model, correlations between biomass estimates can be taken into account when combining biomass trends when calculating total biomass, and FMP area specific biomasses for apportionment purposes. This is relavent if bias correction is to be done, as this requires an estimate of the variance.
- If there is concern that the observation cv has been uniformly underestimated, resulting in year to year variability that is greater than would be expected given the life history of the species being modeled, it is possible to add an additional CV component.
- In the case that absolute biomass survey data is lacking in some years (particularly if this occurs in the most recent years), trend information can be taken from a relative biomass index (such as a CPUE index) as long as the relative biomass index tracks the exploitable biomass trend. This is implemented by adding a second (or more) observations and estimating a catchability constant.
- Although zero observations are not explicitly dealt with at present, development work is ongoing to explore the use of observation models that allow for and model the occurence of zeros.

Weaknesses

The modelling zero observations is not explicitly dealt with in the model and still open to user choice. Given that the model is written in TMB and that the datasets tend to be quite small there are a range of approaches that could be implemented that would not be too computationally intensive. The Tweedie distribution has been identified as a potential solution, however I have some issues with this.

1. As written the background documents, the Tweedie is treated as an additive error term. I don't think this is correct. It should be that the observed biomass in a given year and strata is Tweedie distributed with some mean, dispersion and shape parameter.
2. It is known that inherent in the definition of the distribution, the probability of a zero is correlated with the mean of the distribution. This may or may not be desirable. If a zero is due to low density, this obviously is a good thing. But if the

zero is due to high patchyness, or a distribution shift, then missing at random would be better.

3. Care should be taken to model the same quantity when switching distributional assumptions. In the case of modeling log observed biomass with normal errors, it is the median that is being modelled, but if using a Tweedie or Gamma on the log link, then it is the mean that is modelled.

parameterisation of additional noise The parameterisation of additional noise results in a higher down-weighting of observations with low CV than those with high CVs in terms of the standard error of the observation used in the observation likelihood. Shown below is a plot where the increase of the observation SE resulting from adding an extra CV of 0.2 is added to CVs ranging from 0.1 to 0.6. The equation used to calculate the 'inflated' observation SE is

$$\sqrt{\log(CV_{obs}^2 + CV_{extra}^2 + 1)}$$

This may be appropriate, but it should be noted that if observation errors were incorporated as inverse variance weights these would be applied multiplicatively to the common (residual) variance, and hence the effect would be constant across the range of CVs (Figure 1)

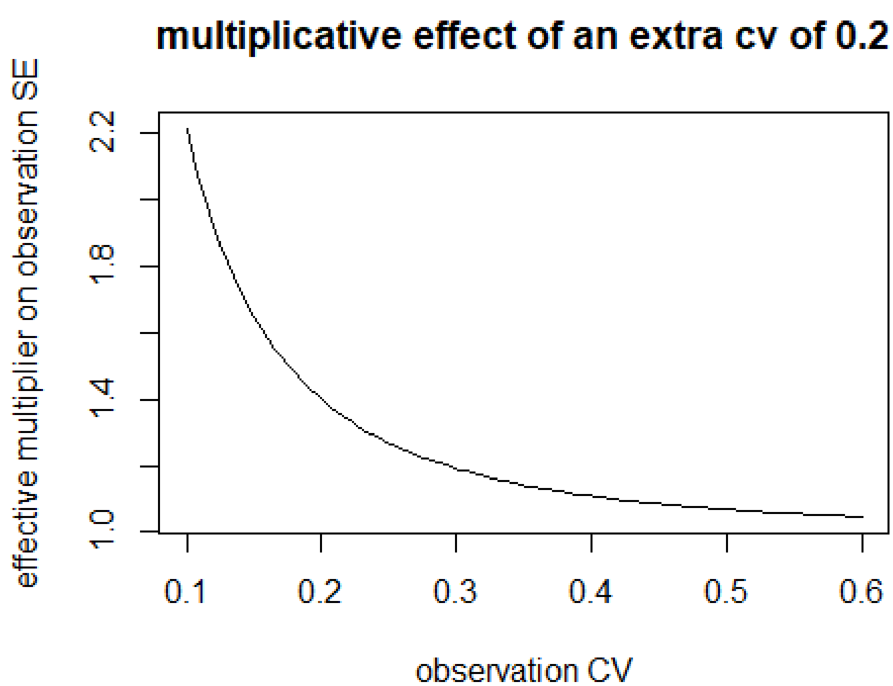

Figure 1: Multiplicative effect of an extra CV

Different stratifications between surveys and cpue indices is probably a minor issue, and not one easy to fix without modelling biomass on a finer spatial scale. And depending on the prevalence of this issue probably not worth additional effort.

Independent trends between strata is hard wired into the model, as I understand it. It is possible to share the process error terms across strata, but not any information about the trends themselves. Is it reasonable to assume that trends in neighbouring areas are independent, or is there utility in incorporating a correlation term across areas within years?

Lots of model settings options could mean that there is a danger of non-consistent application of options.

- For example, were the Tweedie distribution prove to be a useful option in practice, it can still be used to model data that has no zero observations, in this case it would tend to a gamma distribution. How this relates to using logged data with normal errors is something that should be explored, otherwise the risk of a change in model settings, which should give exchangeable results, may give different estimates for technical rather than structural reasons.
- The same line of thinking applies to allowing additional CV on observations. It would be good if there was a rule of thumb that could be applied when the biomass trend appears to be too great relative to the observation error.

ToR 2. Given the available data, evaluate the "random-effects" model now used for many assessments

i. Is this time-series/Kalman filter approach the best estimate of current biomass for management?

Strengths

There are a variety of models that could be used to model a trend in time, for example additive models like regression splines (Wood 2017), Gaussian Markov random fields (Rue and Held 2005) and a combination of these and other ideas (Fahrmeir et al. 2022). A discussed in Rue and Held (2005) random walk models are closely related to regression splines and show an equivalence between the second order random walk GMRF and the thin plate regression spline, the reduced rank version of which is the default in the widely used R package `mgcv`. So random walk models are firmly seated in statistical applications.

It is most common, at lease in my experience, to think of a regression spline as having a penalty on its second derivative. Equivalently, on a local scale values are shrunk towards the average of their neighbours, i.e. spikes and troughs are penalised as they exhibit large changes in the 2nd derivative, and overall the trend is shrunk towards a constant slope. This means that consistent increases (or decreases) from year to year are not penalised at all, and (in the absence of boundary conditions) projections exhibit the slope of the trend present in the data at the boundary.

First order random walks (and e.g. thin plate splines of order 1), on the other hand, penalise first differences. That is, the change from year to year is restricted, regardless of direction, and so estimated trends shrink towards a constant value with no slope. This also means that for projections the trend will continue at a constant value based on the level present at the boundary of the data.

It is possible in practice, for modelling biomass trends, that a combination of these features (shrinking towards the average of neighbours and penalising to a constant value) would be beneficial, but restricted to a choice of one or the other, it is my opinion that the first order random walk has better properties. The effect at the boundary are more stable for the first order random walk, and this is of critical importance for stock assessments. Furthermore, because second order models have the property that an estimate in a year tends towards the average of its neighbours, this implies that a new year of data is likely to impact terminal year estimates more so than for a first order random walk.

Weaknesses

One weakness of the approach, and this is not a big weakness, is that it is not possible to explore closely related models such as the second order random walk. Alternatives that may be interesting to try would be a second or third order random walk with optional additional random error (process error residuals around the smoother biomass trend). However, with the process error residuals this would mean a doubling of the numbers of latent variables as the realised biomass would be given by the smooth biomass trend plus some noise. And it would also mean an additional variance parameter, and may also result in linear trends for many applications. The second order random walk is also likely to have undesirable effects at the boundaries, so is unlikely to perform better than the first order random walk.

If large changes in biomass from year to year are suspected to occur due to climactic, ecological or through known human impacts, and these are measured, it would be beneficial to include these in the model. These could be included as weights that would allow a bigger step change from one specific year to the next, and may have the effect of reducing the estimate of process error, and improving the precision of biomass estimates. Potential reasons could be large recruitments, short term shifts in fishery distribution, changes in fish distribution due to cold pool effects, and so on.

ii. Are the distributional assumptions appropriate (e.g., lognormal) and how should process error estimation be handled?

Strengths

The log normal distribution, or as I prefer to state it, modelling the log observations with normal error, is a standard approach and works well in practice. It is well known that strictly positive quantities like biomass tend to have higher variability when the value is larger, that is, that strictly positive quantities tend to have a constant coefficient of variation. Hence distributions like the lognormal or the Gamma are typically employed.

Estimation of process error should be done within TMB as is the case for many stock assessmentent models. there is nothing controversial as far as I can see with the way process error is estimated within the REMA model. It is a strength that when modelling several strata, the same process error variance can be applied to all biomass trends.

Weaknesses

The downside to using strictly positive distributions are when zero observations occur. These can be difficult to handle as the reason for why the zero occured impacts on how best to deal with it.

It is worth noting that currently the model explicitly uses the normal distribution and operates on log scale data and by modelling in this way, parameter estimates are modelling the median biomass rather than the mean. This is not anything to be concerned about as this is the case for many stock assessment models, but is worth noting as when the Tweedie distribution is discussed along side the use of the lognormal, it requires some explanation to make clear whether the Tweedie distribution is applied to biomass or log-transformed biomass, and if biomass (as I think is the appropriate way) then is the log link being used? It was not clear to me from the background documents that the Tweedie distribution was being used appropriately.

Since the Tweedie distribution is being investigated with parameter, p, between 1 and 2, there is a strong link to the distributions, the Poisson and the Gamma, which are often modelled on the log link. Modelling in this way targets the mean of the observation, rather than the median (as modelling log biomass does). In terms of consistency, or for a more direct comparison, it is suggested that an exploration of modelling biomass using a Gamma distribution on the log link, and applying weights to the dispersion parameter of the Gamma, would be useful.

iii. Can multiple surveys be combined appropriately and what about catchability and selectivity?

Strengths

By modelling biomass trend within the same model, potentially sharing, where appropriate, process error variance, extra CV, and catchability parameter estimates accross strata, estimates of strata level biomass should improve. Additionally, combining biomasses accross regions within TMB allows for correlations between biomass estimates to be incorporated into the estimates of uncertainty of sums of biomasses.

Weaknesses

Another point to note on the modelling of the median is the combination of the estimates of biomass across strata. How is this done? If done within TMB the value returned, I assume, is the median of the disribution of the estimate of total biomass. But, if estimates are summed outside of TMB without bias correction being applied, then the total biomass is the sum of the estimates of median biomass.

By fixing q over years, there is an assumption that selectivity is constant. If selectivity was to change over time, say in relation to body condition, where a younger / smaller fish become less catchable, then the assumption of constant catchability in the model would not be correct. However, if there was strong evidence that this was occuring, and there is available data on the suspected cause (body condition), then either, a covariate could be included in the model to estimate the effect of body condition on catchability, or this effect could be included in the estimation of the RPW prior to use in the model

ToR 3. Evaluate use of models for biomass for stock complexes

i. How to estimate total biomass (i.e., multiple models, or run together, or haul level)

For stock complexes there appears to be several options:

1. Estimate biomass by strata separately for each species, or
2. estimate total biomass by strata for each species.

The question of whether to fit several independent models seems redundant, as it is clearly advantageous that biomass trends that will eventually be combined to be fitted concurrently and the total biomass estimated within TMB.

As to whether the separate biomass trends should be estimated separately for each species, or for all species combined, is case specific. An argument for combining across species and modelling the total biomass would be if there were years of zero observations for some species, where combining would reduce or remove the zero obervations.

Weaknesses

If a model based estimation procedure is used to produce the estimates of observed biomass and their observation CVs, it is possible that there exists correlation between the estimates of biomass within a year between a group of species, for example, if those species are commonly caught together at the haul level. If this is the case, then it may be useful to incorporate this in the observation equations which may improve the estimates of species level biomass trends.

ii. How to estimate aggregate M for ABC/OFL

The question of how to estimate M for stock complexes depends on whether the complex was formed based on a grouping of species with similar life histories and hence likely to have similar values of M, or whether the stock complex is a group of species for which there is limited information and do not nesisarily have similar values of M.

Strengths

The model allows the stock assessor to group species, estimate a single biomass trend and apply a single M value to calculate ABC / OFL. It also allows for seperate estimates of biomass by species, and in the case that these species have different Ms, it becomes possible to estimate a weighted M, based on relative biomasses by species (equivalent to applying a species specific M to species level biomass values and summing to give a total ABC/OFL for the complex)

Weaknesses

Since the REMA model does not estimate M, as this is an externally estimated quantity, it is not clear what weaknesses could be raised.

ToR 4. Evaluate use of random effects models for apportionment

Strengths

These models provide an estimate of biomass by strata (depth and area) and allow the calculation of biomass for a combination of areas such as an FMP area.

Weaknesses

In the estimation of independent trends by strata and area. If the area trends are highly correlated, then much better estimates of region and subregion biomass could be calculated.

It is my understanding that the estimates of biomass used are the median of the lognormal distribution. My instinct would be to use the estimate of the mean of biomass (i.e. the bias corrected values) for apportionment. But I do not have a strong opinion. However, it should be clearly stated that the proportions are calculated as the relative proportions of the median biomasses to the median of the total biomass, if this is indeed the case.

ToR 5. Are other methods more appropriate and make recommendations for improvements (i.e., simple moving averages, ARIMA models, spatial-temporal models)

Strengths

As discussed in ToR 1, the first order random walk model for biomass trends is appropriate and represents best available paragmatic approach.

Weaknesses

Several extensions could be made to the current model, most of which have been discussed above. As a summary potential improvements are as follows:

Inclusion of covariates

- depth effect on catchability could allow catchability to vary by depth accross area
- region effect on catchability could allow catchability to vary by region
- fish condition? - numbers may be stable but biomass may vary due to condition, if this is the case, a fish condition effect in the observation equation could provide more accurate biomass trends.

Use of spatial structure

- random walks on lattices for spatial structure, if there are many regions in a model, a spatial effect could be used to model correlated random walks, or introduce correlated catchability parameters for CPUE indices.
- correlation between biomass trends in general could be introduced where biomass trends between strata are allowed to be positively correlated. Negative correlations between regions could indicate movement of biomass from one strata to another.
- common trend in biomass with small deviations for regions would be one way to bring in correlation between strata. Ideally the common trend would get the largest process error, while the region trends would be smaller. It may be possible to estimate the effective degrees of freedom of these trends and compare this statistic accross models.

Alternative trend models

- 2nd order random walk with year random effect, this would provide a smoother estimate of underlying biomass with a deviation from this smooth trend that gives the actual biomass.
- incorporate season if surveys conducted at different times, a season effect could be included to allow for changes in biomass or catchability at different times of year

Conclusions and Recomendations

ToR 1. Evaluate survey-averaging methods used to support survey biomass-based approaches for data-moderate stocks in the North Pacific.

Conclusions

- The method provided is sound and demonstrates a good balance of simplicity and complexity given the available data.
- The model descrption could be simplified or generalised if couched as a linear model, but this depends on the prior experience of the user.
- The assessment model provides flexibility to incorporate both absolute survey indices of biomass and relative estimates of biomass from CPUE indices.
- The method of incorporation of additional observation variation appears to down-weight observations with low CV more than observations with high CV

Recommendations

- Consider thinking of the model as a generalised additive random effects model and how this could be fitted in alternative modelling frameworks such as GLMMTMB, INLA, mgcv, etc. It is likely that development of the rema package is the best way forward, but looking at the model from a different perspective could allow investigation of model structure and distributional assumtions prior to implmentation in the core rema code. Additionally, the linear model perspective may provide a route for more standardised and simple statement of the model structure when described in stock assessment reports.
- Investigate the impact of using the mean and the median estimate of biomass when caculating total biomass and area proportions of biomass.
- Consideration should be given to whether covariates may help in improving the estimates or interpretation of regional estimates, either through the catchability coefficients, or in the trends themselves. The main benefit in this would be when dealing with missing data, and if catchability varied through time based on some auxilliary variable like fish condition, or cold pool extent.

ToR 2. Given the available data, evaluate the "random-effects" model now used for many assessments

i. Is this time-series/Kalman filter approach the best estimate of current biomass for management?

Conclusions

- The random effect model used, the first order random walk, is well suited to modelling a single biomass trend.

- Due to the fact that the first order random walk penalises towards a constant value, terminal year estimates of biomass used in management are more stable than those from alternative models such as a second order random walk.

Recommendations

- Consider incorporating correlation between biomass trends accross combinations of strata.

ii. Are the distributional assumptions appropriate (e.g., lognormal) and how should process error estimation be handled?

Conclusions

- log normal error is appropriate when zero observations are few and can be treated as missing at random.
- When there are several zeros and these are associated with low biomass then an alternative observation model should be sought.
- Process error should be estimated as is done in the model currently.
- The Tweedie distribution is promising for dealing with data sets of survey observations with many zeros.

Recommendations

- be clear that it is median biomass that is being modelled and reported from the model.
- Clarify the use of the Tweedie distribution and whether the median or the mean of biomass is being modelled. Part of this could be to compare the use of a Gamma error distribution to the log normal distriution for existing applications.

iii. Can multiple surveys be combined appropriately and what about catchability and selectivity?

Conclusions

- multiple surveys can be combined appropriately and thier standard error of the combined estimate also, allowing for appropriate bias correction on the combined biomass to be applied if required.
- As long as selectivity and availability, etc, are constant over time, it is appropriate to treat CPUE indices as relative abundance, and the model is correctly set up to do this.

Recommendations

- If there is evidence for changes in selectivity / catchability over time and there is data available to inform this, the model should be extended to allow covariates in the CPUE observation equation.

ToR 3. Evaluate use of models for biomass for stock complexes

i. How to estimate total biomass (i.e., multiple models, or run together, or haul level)

Conclusions

- Estimation of biomass trends within a single model is better than fitting multiple models. Benefits include sharing parameters, such as catchability, or process error, and utilising the correlation between biomass estimates when computing total biomass and its standard error.

Recommendations

- run together and share parameters where appropriate. The return summed biomass taking into account correlations for the estimation of confidence intervals and standard errors and bias correction.

ii. How to estimate aggregate M for ABC/OFL

Conclusions

- How to estimate aggregate M depends on the reasons why the species were grouped. Some species are grouped because they have similar life histories, others because there is limited information and they cannot be assessed as a single species.
- If species within a group have different values for M, it is possible with the REMA model to estimate species level biomass trends and apply species level Ms to species level biomass and then accumulate to provide a an aggregate ABC and OFL.

Recommendations

- Since the M is estimated outside of REMA for all stock assessments, it may be useful to have a single working group explore and provide estimates of M for all stocks using REMA.

ToR 4. Evaluate use of random effects models for apportionment

Conclusions

- The proposed model is suitable for apportionment of biomass as long as the assumptions of the model are met: that each survey provides an unbiased estimate of regional biomass (up to a multiplicative constant in the case of CPUE indices).
- If bias correction is to be done prior to the apportionment calculations, it is important to fit all biomass trends within the same REMA model to allow for an appropriate estimate of standard error.

Recommendations

- Clarify the effect of using the estimate of the median or the mean when calculating apportionment.

ToR 5. Are other methods more appropriate and make recommendations for improvements (i.e., simple moving averages, ARIMA models, spatial-temporal models)

Conclusions

- The first order random walk model for biomass trends is appropriate and represents best available paragmatic approach

Recommendations

A variety of extensions could be made to the current model, listed below are set of possible extensions

- inlcusion of covariates
  - depth effect on catchability
  - region effect on catchability
  - fish condition - numbers may be stable but biomass may vary due to condition.
- use of spatial structure - possiblities include
  - random walks on lattices for spatial structure
  - correlation between biomass trends
  - common biomass trend with deviations by region.
- Alternative trend models:
  - 2nd order random walk with year random effect
  - incorporate season if surveys conducted at different times and expliotable biomass varies by season within an area

General points

Recommendations

1. rema package already exists on CRAN

The R package 'rema' unfortunately already exists on the Central R Archive Network (CRAN) which is the standard for hosting R packages. This means that if someone tries to install the 'rema' package using the 'install.packages()' function from base R, i.e.

```
install.packages("rema")
```

the user will install the wrong package (https://cran.r-project.org/web/packages/rema/index.html) a package designed for rare event meta analyses. The consequence then is that the user will have to manually remove this package and install from the github repository https://github.com/afsc-assessments/rema, by running

```r
# install.packages("devtools")
# also install R build tools for the appropriate R version
library(remotes)
install_github("afsc-assessments/rema", dependencies = TRUE)
```

Ideally packages intended for public use should not share names with existing R packages on CRAN.

2. Make rema easier to install for windows and mac users

There are useful free systems available for github hosted R packages that provide a system to check and build R packages periodically and when changes are made to the codebase. Such a system is https://r-universe.dev/search/. and is simple to set up and has the benefit of providing a means for users to install (without having to complile) via the standard `install.packages()` command, for example if set up, installation could be done via

```r
install.packages("rema", repo = "https://afsc-assessments.r-universe.dev")
```

and would not require the windows or mac user to have R build tools, or the `devtools` / `remotes` package installed.

An additional benefit, is that if the user adds 'https://afsc-assessments.r-universe.dev' to the list of repositories in thier global options through thier `.Rprofile` file, then a standard call to `update.packages()` would fetch the most recent version of the `rema` package.

3. Consider future development in RTMB.

A new R package is developing called `RTMB` (https://github.com/kaskr/RTMB) which provides bindings from R code to a subset of TMB. This allows the developers to write all thier code in 'plain' R code, allowing easier debugging and potentially increased collaboration from outside the core rema development team. Another benefit is that no compilation is required when modifying the model code, and so development and exploration should be more efficient.

An example of the simplest form of the rema model is given below and gives exactly the same estimates of biomass and process error as the rema package function:

```r
library(RTMB)
library(rema)

# use rema to get a sample dataset
aisr <- rema::read_admb_re(
  filename = system.file("example_data/aisr_rwout.rep", package = "rema")
)
```

```r
# RTMB model definition
jnll <- function(params) {
  getAll(params, data)

  ADREPORT(exp(2 * logSigmaPE))

  jnll <- -sum(dnorm(log(biomass), logB[idx], sqrt(log(cv^2 + 1)), TRUE))
  for (i in 2:length(logB)) {
    jnll <- jnll - dnorm(logB[i], logB[i - 1], exp(logSigmaPE), TRUE)
  }

  jnll
}

# get the data
data <- aisr$biomass_dat
data$idx <- sapply(data$year, function(x) which(x == aisr$model_yrs))

# initialise parameters
parameters <- list(
  logB = rep(0, length(aisr$model_yrs)),
  logSigmaPE = 0
)


# fit the model
obj <- MakeADFun(jnll, parameters, random = c("logB"), silent = TRUE)
opt <- nlminb(obj$par, obj$fn, obj$gr)

# summarise
sdr <- sdreport(obj)
pl <- as.list(sdr, "Est")

# using rema package
input <- prepare_rema_input(admb_re = aisr)
rema_fit <- fit_rema(input)
output <- tidy_rema(rema_model = rema_fit)

# compare estimates:
#   maximum difference of estimated biomass: 1.231783e-07
max(abs(output$total_predicted_biomass$pred - exp(pl$logB)))
```

References

Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang, and Brian D Marx. 2022. "Regression Models." In *Regression: Models, Methods and Applications*, 23–84. Springer.

Rue, Havard, and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. CRC press.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with r*. CRC press.

Appendices

Appendix 1. Bibliography of Materials Provided

Background Papers

- Survey_Averaging_Working_Group_Report_2013.docx This is the original working group report on survey averaging investigations of methods for averaging survey biomass estimates. This report was presented at the 2013 NPFMC Groundfish Joint Plan Teams meeting.

- Hulson et al. Tech Memo Jan2021.pdf This is a tech memo published on application of the random effects model using multiple indices for biomass and apportionment estimation.

- Monnahan et al. 2021 paper.pdf This is a paper presented to the NPFMC review bodies in September, 2021, that summarizes approaches and methods for how the random effects model was developed for the NPFMC.

- Sullivan et al. 2022 REMA paper.pdf This is a paper presented to the NPFMC review bodies in September, 2022, that summarizes the REMA model, the consensus version of the random effects model used by the NPFMC.

- This is a website for the rema package as discussed in the Sullivan et al. 2022 paper (above). In addition to hosting background information, code examples and documentation, it also has a living version of the underlying methodology from the main report. This version may be better to use (instead of the report) because it's updated every time there is a typo or if a new method is added: https://afsc-assessments.github.io/rema/articles/rema_equations.html

Background Presentations

- Monnahan et al 2021 JT PT presentation.pdf - This is the oral presentation provided to the NPFMC Joint Groundfish Plan Teams (JT PT) in support of the Monnahan et al 2021 paper.pdf

- Spencer-Evaluation-Statistical-Models.pdf - This was a presentation to the 2015 Lowell Wakefield Symposium. It contains a comparison of a few different models, and ideas for developing process error priors based on life history. The estimation of process errors is an ongoing question.

- Survey_Average_wg_2013 - This is a presentation given to the North Pacific Fishery Management Council advisory bodies. Some of this information is summarized in the Spencer-Evaluation-Statistical-Models.pdf.

- Survey_average_wg_2015_final - This is a presentation given to the North Pacific Fishery Management Council advisory bodies. Some of this information is summarized in the Spencer-Evaluation-Statistical-Models.pdf.

- Sullivan et al 2022 REMA JT PT presentation.pdf - This is the oral presentation provided to the NPFMC Joint Groundfish Plan Teams (JT PT) in support of the Sullivan et al 2022 REMA paper.pdf

Background Review Comments

- JT PT Minutes Survey Averaging Working Group Sept2013.pdf - These are the NPFMC Joint Groundfish Plan Team (JT PT) minutes review comments to the Survey Working Group Report 2013.
- JT PT Minutes RE Model Sept 2021.pdf - These are the NPFMC Joint Groundfish Plan Team (JT PT) minutes review comments to Monnahan et al 2021 paper.
- SSC Report RE Model Sept 2021.pdf - These are the NPFMC Science and Statistical Committee (SSC) review comments to Monnahan et al 2021 paper.
- JT PT Minutes REMA Sept2022.pdf - These are the NPFMC Joint Groundfish Plan Team (JT PT) minutes review comments to Sullivan et al 2022 REMA paper.
- SSC Report REMA Oct2022.pdf - These are the NPFMC Science and Statistical Committee (SSC) review comments to Sullivan et al 2022 paper.

Appendix 2. CIE Performance Work Statement

External Independent Peer Review

AFSC Approaches to Survey Biomass-based Stock Assessments

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, 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 . Further information on the Center for Independent Experts (CIE) program may be obtained from https://www.ciereviews.org.

Scope

Stocks in Alaska federal waters that have reliable survey biomass estimates, but lack sufficient data for an age or length based stock assessment, are assessed using only of exploitable biomass estimates and an estimate of natural mortality. Historically, exploitable biomass estimates for catch advice varied widely, from the most recent survey estimate, to a variety of moving averages. Starting in 2015, a working group tested via simulations a "random effects" model, which essentially is a close approximation of the univariate Kalman Filter model against a variety of alternatives and recommended it for use by the NPFMC. The method has been expanded to stock complexes, multiple survey indices, and regional apportionment of Acceptable Biological Catch (ABC).

Given that the adoption of this approach has direct implications on management advice, it is important that the methods represent the best available science and are statistically sound. Therefore, the CIE reviewers will conduct a peer review of the current methods based on the Terms of Reference (TORs) referenced below. Given the direct impacts to

Alaska fisheries, it will be important for NMFS to have a transparent and independent review process of the model used in these assessments.

Requirements

NMFS requires two reviewers to conduct an impartial and independent desk review in accordance with this Performance Work Statement (PWS), OMB Guidelines, and the ToRs below. The reviewers shall have working knowledge and recent expertise in the application of fish stock assessment methods, particularly survey-based data-moderate assessments. The CIE reviewers shall have expertise in random effects models, and times-series approaches such as Kalman filter methods. In addition, the CIE reviewers should understand design-based and model-based survey estimation methods. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Tasks for reviewers

Each CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

1. *Pre-review Background Documents*: Review the following background materials and reports prior to the review:

The following document details recent advances in the methodology. Monnahan et al. 2021. Improving the consistency and transparency of Tier 4/5 assessments. https://meetings.npfmc.org/CommentReview/DownloadFile?p=86098951-a0ed-4021-a4e1-95abe5a357fe.pdf&fileName=Tiers%204%20and%205%20assessment%20considerations.pdf

This document is summarized in a presentation at the September 2021, North Pacific Fishery Management Council Groundfish Plan Team meetings. https://meetings.npfmc.org/CommentReview/DownloadFile?p=02281578-6fca-4f7b-a33b-0f129152a7e4.pdf&fileName=PRESENTATION_Tier%204%20and%205%20Consideration.pdf

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 addition to the documents cited above, the Project Contact will provide pertinent case study example stock assessments that highlight the different use of the methods. 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:

2. *Webinar*: Additionally, approximately two weeks prior to the peer review, the CIE reviewers will participate in a webinar with the NMFS Project Contact and other staff to address any clarifications that the reviewers may have regarding the ToRs or

the review process. The NMFS Project Contact will provide the information for the arrangements for this webinar.

3. *Desk Review*: Each CIE reviewer shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. Modifications to the PWS and ToRs can not be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor.

4. *Contract Deliverables - Independent CIE Peer Review Reports*: Each CIE reviewer shall complete an independent peer review report in accordance with the PWS. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Place of Performance

Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Period of Performance

The period of performance shall be from the time of award through May 2022. Each reviewer's duties shall not exceed 10 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 reviewers
Two weeks prior to the review	Contractor provides the pre-review documents to the reviewers. Reviewers participate in Webinar.
March 2022	Each reviewer conducts an independent peer review as a desk review
Within two weeks after review	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

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

Travel

Since this is a desk review travel is neither required nor authorized for this contract.

Restricted or Limited Use of Data

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

NMFS Project Contacts:

Sandra Lowe/Chris Lunsford Supervisory Fish Biologists NOAA/NMFS/AFSC
Sandra.lowe@noaa.gov, Chris.Lunsford@noaa.gov

Annexes

Annex 1: Peer Review Report Requirements

- The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.
- 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.
- The reviewer report shall include the following appendices:
    - Appendix 1: Bibliography of materials provided for review
    - Appendix 2: A copy of the CIE Performance Work Statement

Annex 2: Terms of Reference for the Peer Review

AFSC approaches to biomass based stock assessments

CIE reviewers are contracted to complete their independent peer review based on the ToRs. Therefore, the CIE-NMFS review and approval process is based on whether the CIE independent reports addressed each ToRs. The AFSC requests a desk review in March 2022 to review the current data-moderate stock assessment methods used in the North Pacific, specifically related to survey averaging methods. CIE reviewers shall address the following Terms of Reference (ToR) during the peer review and in the CIE reports.

- Evaluate survey-averaging methods used to support survey biomass-based approaches for data-moderate stocks in the North Pacific.
- Given the available data, evaluate the "random-effects" model now used for many assessments
  - Is this time-series/Kalman filter approach the best estimate of current biomass for management?
  - Are the distributional assumptions appropriate (e.g., lognormal) and how should process error estimation be handled?
  - Can multiple surveys be combined appropriately and what about catchability and selectivity?
- Evaluate use of models for biomass for stock complexes
  - How to estimate total biomass (i.e., multiple models, or run together, or haul level)
  - How to estimate aggregate M for ABC/OFL
- Evaluate use of random effects models for apportionment
- Are other methods more appropriate and make recommendations for improvements (i.e., simple moving averages, ARIMA models, spatial-temporal models)