**Synopsis of Yelloweye Biomass estimation from ROV survey data**

8/3/21 – Phil Joy

Methods outline:

1. Imported ROV data and ran script to format data for distance analysis
   1. Kellii checked transect lengths with GIS and they were off from the ones I calculated in R by ~2.5%. Pretty minor, but reran data with GIS lengths and need to work on R code to try and figure out the discrepancy. Chatting with Kellii we both agreed that GIS probably more accurate. R transect lengths were longer than GIS, so density estimates are a tad higher when using the GIS transect lengths.
2. Ran distance analysis on data to estimate density of YE/ km2.
   1. Ran different combinations of detection functions and adjustments. Uniform with various adjustments were the best, but several others were competitive based on AIC, including some using either depth or life stage as a covariate.
   2. The current operational plan calls for getting rid of some fish from the distance analysis based on length, but that seemed faulty to me. For one, not all fish have length measurements, so dealing with that issue gets pretty problematic. Secondly, a poorly measured yelloweye is still a yelloweye and should count towards the density estimate. I think the best approach is to examine the length distribution of the decently measured ROV fish and apply the proportion measured below a designated cutoff value to adjust the biomass estimate on the back end.
   3. Models in general had good fit, but without truncation some models failed to converge on occasion.
   4. Truncated distances at 5% and the models ran well and displayed improved fit. Plots look great and P-values on the GOF tests were very high (> 0.8 in general…). 5% truncation is cited as being standard in distance modelling.
   5. Based on those results I did not try any binning.
   6. Several models had AICc scores within deltaAIC of less than 4.
   7. All models produced similar estimates of D (figure 1).
   8. Model averaged all models with deltAIC < 4 (7 models) using bootstrap procedure as recommended (Table 1).
3. Examined length and weight data from port samples.
   1. Cleaned data and used average of last 3 years to generate average weights by subdistrict. Because of covid, there was not enough data in ’20 or ’21 to do much with, but 3 years produced robust sample sizes (see spreadsheets). Kellii had this on her SAFE to-do list as well, so these weights may get updated.
   2. Removed some outlier values. During weight-length exam (see below) it was clear that there were some excessively and suspiciously very fat and very skinny YE. I calculated condition (K=W/L^3) and removed those in the 0.1 and 99.9th percentiles to get rid of the weirdos (figure 2). This might be too restrive and could be dialed back, but I wouldn’t expect any difference in biomass calculations. Average weights and L:W regression taken from this data.
   3. Used SSEO port data from last 3 years to produce linear model for back-transforming ROV lengths into weight estimates.
4. ROV length data
   1. Lots of very imprecise estimates from the stereo cameras!
   2. Compared length point estimates to CF samples and ROV fish seem a little smaller on average, but more very big and very small fish relative to harvests (Figure 3).
   3. To get mean weights of ROV measured fish I ran a bootstrap
      1. First removed really bad length estimates; culled all fish where the point estimate minus 3 standard deviations went negative. These turned out to be the fish that were the tiniest according to the point estimates.
      2. Randomly sample that set of lengths, with replacement. Ignored the imprecision in the length measurements and assumed that although imprecise they were unbiased. Resampling the point estimates should be adequate in this situation.
      3. For each rep, the new set of fish lengths were back transformed to weights using the regression derived above.
      4. Measured 3 things in each rep:
         1. The proportion of fish greater than 270 mm (I used that as the cut off)
         2. The average weight of all fish
         3. The average weight of fish bigger than 270 mm
5. Biomass calculations (YE\_biomass\_est\_2020 for 20201\_pj.xls): I used the spreadsheets that were provided by the groundfish team and will walk through the tabs here. I know Andrew is familiar with this spreadsheet and apologize to Chris if my descriptions below are hard to follow. Green tabs used GIS calculated transect lengths while blue tabs used R calculated transect lengths.
   1. Biomass calcs 2020: I did not touch this tab and left it for reference…
   2. Biomass calcs 2021 Best Mod GIS: This tab used the density estimate from the best model (lowers AIC) to get biomass.
      1. This spreadsheet was set up to calculate confidence intervals using the cvs of the estimator components. I wasn’t familiar with those methods but would like to know where they came from since combining error on multiple variables is usually accomplished through the variances. I calculated the confidence intervals using the variances of the components, which are highlighted in tan and are in columns J-U. My calculations produced similar variances, but slightly lower estimates.
         1. Rows 15-18 use the weights from the commercial fishery
         2. Rows 44-47 use the estimated ROV weights to do the same thing with the SSEO data. (Table 1)
   3. Biomass calcs 2021 Best pgt GIS: This tab is similar to the last one, but now I’ve included uncertainty in the proportion of the ROV fish that are greater than the cutoff length of 270. You’ll see that in column K and L where I use that to modify the biomass estimate of SSEO fish. The proportion estimate comes from the ROV bootstrapping. 99% of the fish are greater than the cutoff value of 270 that I used so this turns out to be pretty trivial (Table 1). However, this is set up to easily consider other cutoff values and at least gives us methodology for dealing with this piece of uncertainty.
   4. Biomass calcs 2021 Model Avg. This tab is similar to the last one, but I replaced the Density estimate from the best model with the one produced in the model averaging bootstrap. The density estimate is slightly higher with the model average and thus biomass is higher.
   5. Finally, I ran the model averaging and ROV bootstrap stuff together to directly produce biomass estimates, and estimates of the lower 90% CI, within the bootstrap. Figure 4 and Table 1
      1. Model uncertainty
      2. Uncertainty in ROV lengths
      3. Uncertainty in the proportion greater than 270’
      4. Uncertainty in harvest weights
   6. I haven’t done anything to deal with the major variance component missing from the biomass calculation: the YE habitat area in the subdistrict. This is mentioned in the SAFEs so it’s an established unknown and using the lower 90% CI as the biomass is the current means of compensating for this. For what it’s worth, this is the kind of thing that could be pulled into a Bayesian analysis to try and deal with some of that unknown. But I suspect that moving towards an ASA model is the higher priority at this point.
6. Going forward:
   1. All of the modelling produced similar results, so I think we can feel good about whichever particular model we want to use going forward.
   2. For the SAFE report things are organized for Kellii to take the density estimate and derive biomass in the spreadsheet. I’ve updated the spreadsheets with my calculations, but also don’t want to step on Kellii’s (or anyone else’s) toes. I thought we should clarify how to proceed before sitting down with Rhea and Kellii.
   3. I have the spreadsheet calculations coded in R and can clean that up for easier use in the future.
   4. Cleaning up my code and getting it onto github
   5. Plots for report/ pubs/ meetings…
   6. Do you want me to delve into the historical data? Do we want to reevaluate past estimates?
   7. We could run some simulations to look for potential biases in our estimates. I’m not sure it’s necessary given that the distance sampling methodology is well established, and the bootstrap procedure should eliminate any bias resulting from nonnormality in the data and error structure. However, I thought I would throw it out there as a potential option.

**Table 1.** Summary of yelloweye density and biomass estimates for SSEO from three sources of weight information (port samples, and back calculated weights from ROV lengths for all ROV fish and ROV fish greater than 270 mm) and from four modelling constructs: The best model (lowest AIC), the best model incorporating estimates of the number of ROV fish greater than 270 mm, an average of the best 7 models, and a full bootstrap where biomass is estimated in the bootstrap and thus combines the variability associated with model uncertainty, sampling error in lengths and weights, and uncertainty about the proportion of fish greater than 270 mm.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Transect length source | Weight Source | Model method | Density (YE/km2)  Dhat (cv) | SSEO Biomass (mt) | 10% quantile SSEO Biomass (mt) |
|  | Port samples | * Best Model | 1,907(0.15) | 7,041 | 5,350 |
| R calculated | * Best Model +   prop ROV > 270mm | 1,907(0.15) | 6,784 | 5,140 |
| * Model Avg. | 1,926 (0.15) | 6,848 | 5,204 |
| ROV all fish | * Best Model | 1,907(0.15) | 7,759 | 5,530 |
| * Best Model + prop ROV > 270mm | 1,907(0.15) | 7,482 | 5,320 |
| * Model Avg. | 1,926 (0.15) | 7,553 | 5,385 |
| ROV > 270 | * Best Model | 1,907(0.15) | 8,043 | 5,742 |
| * Best Model + prop ROV > 270mm | 1,907(0.15) | 7,749 | 5,518 |
| * Model Avg. | 1,926 (0.15) | 7,823 | 5,585 |
| GIS measured | Port samples | * Best Model | 1,957 (0.14) | 7,225 | 5,525 |
| * Best Model +   prop ROV > 270mm | 1,957 (0.14) | 7,185 | 5,493 |
| * Model Avg. | 1,989 (0.15) | 7,299 | 5,524 |
| * Full Bootstrap | 1,987 (0.14) | 7,295 | 5,577 |
| ROV all fish | * Best Model | 1,957 (0.14) | 7,144 | 5,371 |
| * Best Model + prop ROV > 270mm | 1,957 (0.14) | 7,104 | 5,340 |
| * Model Avg. | 1,989 (0.15) | 7,181 | 5,347 |
| * Full Bootstrap | 1,987 (0.14) | 7,164 | 5,385 |
| ROV > 270 | * Best Model | 1,957 (0.14) | 7,108 | 5,343 |
| * Best Model + prop ROV > 270mm | 1,957 (0.14) | 7,069 | 5,312 |
| * Model Avg. | 1,989 (0.15) | 7,217 | 5,375 |
| * Full Bootstrap | 1,987 (0.14) | 7,200 | 5,413 |

Chart, box and whisker chart

Description automatically generated

Figure 1. Density (Yelloweye/km2), estimates and 95% confidence intervals for SSEO from the best models according to AIC scores.

Chart, scatter chart

Description automatically generated

Figure 2. Length and weight of port sampled Yelloweye. Red dots represent the outliers that were removed and blue dots represent the data that was retained for calculating mean weights and for calculating the L:W relationship used for modelling weights of ROV fish.

Chart, histogram

Description automatically generated

Figure 3. Length distribution of SSEO port samples and point estimates of ROV measured Yelloweye.

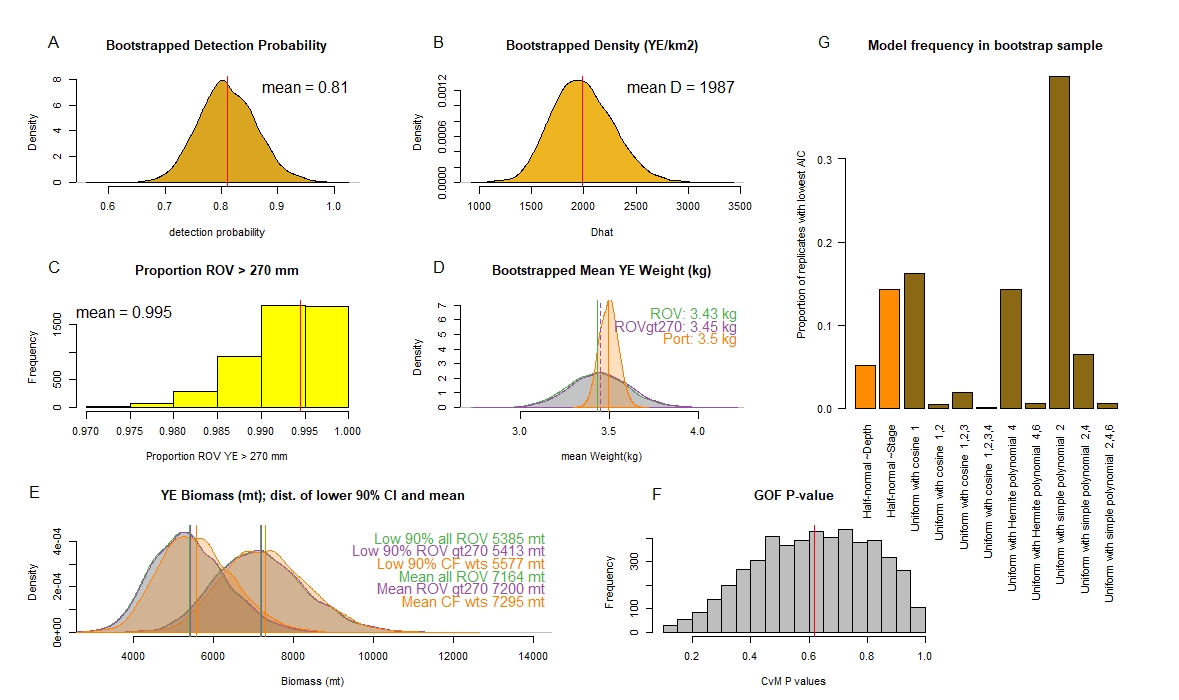


Figure 4. Summary plots of full bootstrap output showing, (A) mean detection probability, (B) mean density, (C) the proportion of ROV fish greater than 270, (D) mean Yelloweye weights from port sampling (CF) and ROV measured fish, (E) resultant biomass estimates (estimates of the mean biomass and estimates of the lower 90% CI), (F) the GOF values (P < 0.1 were culled; we could be more stringent if we want), and (G) the frequency of the various models that were selected during the bootstrap procedure.