MAR 580: Advanced Population Modeling

Fall 2023

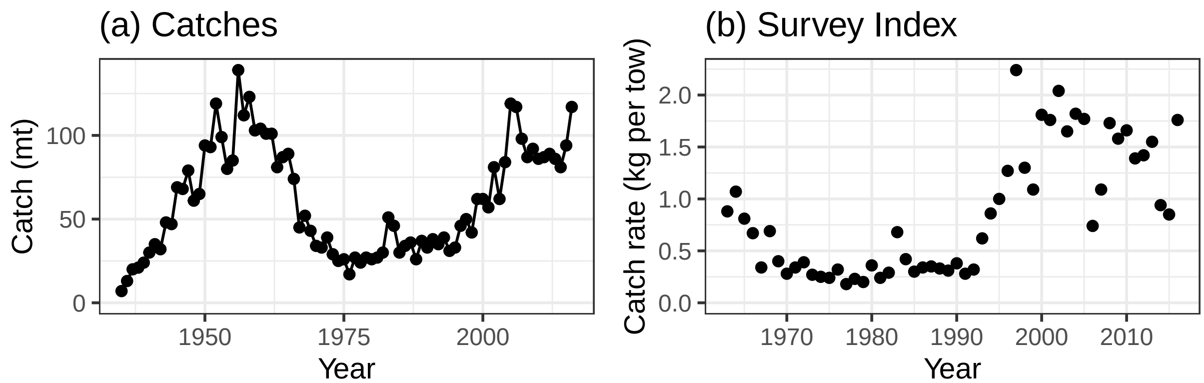
Homework Assignment 4

Statistical catch at age models

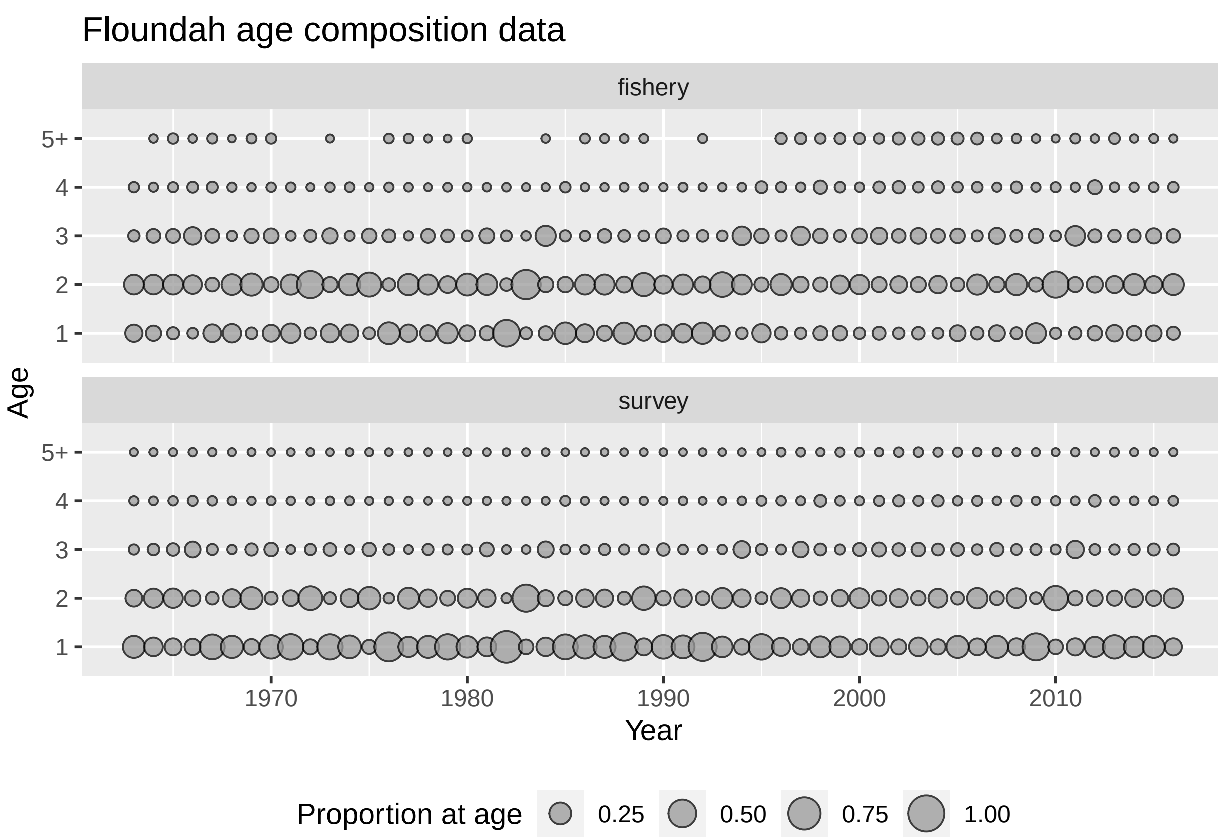
Due: 12/04/23, 9:00 am

Please provide Gavin with a brief report containing your solutions, AND also include your R script (or markdown file), .cpp, or any additional files needed to run your assignment.

For our floundah population, in addition to the time series of catches and survey index, we now have age composition information (proportions at age for ages 1, 2, 3, 4, & 5+) from both the fishery and the survey.




```
## Warning: Removed 14 rows containing missing values (`geom_point()`).
```



1. (No modeling, 20 points) Using the plots of the age data above, write a list of 5 (or more) take-homes that you infer about the population dynamics, year class strengths, and important model structure assumptions and decisions. How do your conclusions differ from previous evaluations when only confronted with the abundance and catch time series?
2. (30 points) Fit a statistical catch at age model to the floundah data, with a start year of 1963. Assume annual recruitments as deviations from a Beverton-Holt stock-recruit relationship (fix steepness at $h = 0.7$), and treat these deviations as random effects. Assume that the population age structure is in equilibrium in the first year (1963), but that the population is not in an unfished state. Parameters to estimate include the time series of fishing mortality, unfished recruitment (R_0), the initial year recruitment relative to the unfished recruitment (ϕ), the annual recruitment deviations, the standard deviation of the annual recruitment deviations (σ_R, i.e. process error variance), selectivity parameters for both the survey and the fishery assuming a logistic function of age for both (details below), and the survey observation error standard deviation, σ_I (if you do not use the analytical MLE). Assume the previously given vectors for weight at age and maturity, and that $M = 0.4yr^{-1}$.

Assume that the logs of the catches and survey index are lognormally distributed as before, and that both the annual fishery and survey catch-at-age comps are multinomial with effective sample size 100.

(Full model specification is provided below in the appendix)

3. (20 points) Produce diagnostic plots showing the fit of the model to the various data sets. Interpret the fits (and mis-fits) to the data. (Consider using the compResidual package to compute One Step Ahead (OSA) residuals for the composition data rather than Pearson residuals).
4. (15 points) Plot the estimated time series of biomass, recruitment, and fishing mortality, as well as the estimated stock-recruit relationship along with the estimated deviations. Provide commentary on the estimated population dynamics.
5. (10 points) Compare your SCAA to the Schaefer (and/or Pella-Tomlinson) model that has a starting year of 1935. What are similarities and differences regarding the estimates of stock size and exploitation pattern?
6. (5 points) Discuss what the implications might be for decisions about making population projections (say to derive catch advice) of the different assumptions about recruitment made in these models?
7. ***BONUS*** *(+20 points)* Based on your evaluation of model diagnostics and data exploration, what are some suggested improvements or next steps for modeling decisions? Implement one of these and compare the results of your alternative model to the model fit in (2).
8. ***BONUS*** (*+40 points*) Use the biological information for floundah (weight at age, maturity, etc.). Assume M=0.4. Estimate F_{MSY} for floundah for the full range of steepness (steepness can take values between 0.2 & 1.0). Plot how the F_{MSY} estimates change as a function of steepness. How do the estimates for F_{MSY} compare with $F_{40\%}$? (and also the F_{MSY} obtained from the Schaefer model for these data) What might the implications of assuming SPR-based proxy reference points be for this stock? (say for example using our fixed assumption of $h = 0.7$)
   Note: No TMB programming is required for this part of the assignment, you can do this all by making use of functions contained in `spr_calcs.R` to calculate YPR and SBPR.
   hint: You will (likely) want to write a function that allows you to optimize Yield given input values for F - this function will make use of the YPR & SBPR functions.

Appendix 1: Model Specification

Process model

The population dynamics equations are:
$N_{t+1,1} = R_{t+1}$ for $a = 1$
$N_{t+1,a+1} = N_{t,a}e^{-(s_a^f F_t + M)}$ for $2 <= a < A$
$N_{A,t+1} = N\_A - 1, te^{-(s\_A-1^f F\_t+M)} + N\_A, te^{-(s\_A^f F\_t+M)}$ for $a = A$

where
$N_{t,a}$ is the numbers at age a in year t,
R_t are the recruits in year t,
M is the instantaneous rate of natural mortality,
F_t is the fully selected instantaneous fishing mortality rate in year t,
s_a^f is the fishery selectivity at age a, and A is the maximum modeled age.

Recruitment in year t is assumed to be lognormal deviations from a Beverton-Holt stock-recruit relationship:
$R_t = \frac{4hR_0SSB_t}{SSB_0(1-h)+SSB_t(5h-1)}e^{(\eta_t - 0.5\sigma_R^2)}$

where
R_0 is the magnitude of unfished recruitment,
h is the steepness (fraction of unfished recruitment at 20% unfished spawning biomass),
η_t is the recruitment deviation in year t, distributed as $\eta_t \sim N(0, \sigma_R^2)$,
σ_R is the standard deviation of the annual recruitment deviations (process error variance),
SSB_t is the spawning biomass at time t, calculated as: $SSB_t = \sum_{a=2}^{A} N_{t,a}w_a m_a$,
SSB_0 is the unfished spawning biomass at time t.

Selectivity at age for both the survey and fishery are assumed to be logistic:

$s_a^{s/f} = 1/(1 + e^{-\ln(19)(a - A50_{s/f})/(A95_{s/f})})$

where $A50_{s/f}$ is the age at 50% selectivity, and $A95_{s/f}$ is the difference (in age) between the age at 50% & 95% selectivity

Initial conditions

$N_{1963,1} = \phi R_0$
$N_{1963,a} = N_{1,a-1}e^{-M}$ for $2 < a < A$
$N_{1963,A} = \frac{N_{1,A}e^{-M}}{(1-e^{-M})}$ for $a = A$

where ϕ is the fraction of unfished recruitment in 1963 (year 1).

Objective Function

The objective function is the sum of the negative log-likelihoods for the individual likelihood components (see observation below), as well as the contribution of the process errors (recruitment deviations).

Observation model

Catches As before, we assume the logs of the annual catches are also normally distributed with a standard deviation of 0.05.The likelihood function is then:

$$L_{catch}(\theta|D) = \prod_t \frac{1}{0.05\sqrt{2\pi}} \exp^{-\frac{[\ln C_t - \ln(\hat{C_t})]^2}{2(0.05)^2}}$$

where $\hat{C}_t = \sum_{a=1}^{A} N_{t,a}w_a \frac{s_a^f F_t}{Z_{t,a}}(1 - e^{-Z_{t,a}})$

with $Z_{t,a} = s_a^f F_t + M$

Survey Index As before, we assume that the survey index observations are proportional to the available biomass in the middle of the year, with the logs of the index being normally distributed, i.e.

$$L_{survey}(\theta|D) = \prod_t \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{[\ln I_t - \ln \hat{I}_t]^2}{2\sigma_I^2}}$$

where $\hat{I}_t = q_I \sum_{a=1}^{A} s_a^s N_{t,a} w_a m_a e^{-0.5(s_a^f F_t + M)}$

(note that given our assumption that both fishing and natural mortality take place throughout the year we can compute the mid-year available biomass directly rather than calculating as the average of the biomass at the beginning of both years t and $t+1$, as done for the production models. *hint* maybe use the previous calculation initially while you get other new pieces of the code to work)

Age Composition Data The age composition data (for both the fishery and the survey) are assumed to be multinomial with given input sample size $N_t = 100$.

The contributions to the objective function for the age composition data are:

$$nLL_{age_{f/s}} = -\sum_t N_t \sum_a p_{t,a}^{f/s} \ln(\hat{p}_{t,a}^{f/s})$$

with the predicted values for the fishery age composition being:

$$\hat{C}_{t,a} = \frac{N_{t,a} s_a^f F_t}{Z_{t,a}} \left(1 - e^{-Z_{t,a}}\right)$$

$$\hat{p}_{t,a}^f = \hat{C}_{t,a} / \sum_a \hat{C}_{t,a}$$

Similarly, the predicted values for the survey age composition are:

$$\hat{p}_{t,a}^s = \frac{s_a^s N_{y,a} e^{-0.5 Z_{t,a}}}{\sum_a s_a^s N_{t,a} e^{-0.5 Z_{t,a}}}$$

Note 1: A more robust version of the nll for the proportions at age that includes constants not added in the above is:

$$nLL_{age_{f/s}} = -\sum_t N_t \sum_a p_{t,a}^{f/s} \ln(\hat{p}_{t,a}^{f/s} / p_{t,a}^{f/s})$$

To implement this you will need to add a small constant to the observed proportions at age and renormalize so the sum over ages = 1.

Note 2: Rather than programming the nLL for the composition data directly, you may instead want to take advantage of the built-in multinomial distribution function in TMB (`dmultinom()`). To use this you will need to convert the age data to numbers at age rather than proportions based on the annual sample size (`dmultinom()` expects the vector of observations to be numbers, not proportions, e.g. see the documentation).

Appendix 2 : Tips

- Start from the age-structured production model, **don't** try to write the whole program from scratch.
- Spend some time with data exploration (ie #1) first. Get a feel for the data objects and their structure before beginning the modeling.
- A suggestion would be to add the contribution to the objective function for the composition data first, using the age-structured production model. You can include these data before modifying the recruitment specification, or indeed estimating the selectivity for both fishery and survey (i.e. use the same fixed inputs for these that you had previously).
- During model development, you can create another estimated dummy parameter, set its initial value to 0, and have the objective function be something like: `nll += dummy*dummy;`
  This will force the model to exit after one function call, and allow you to inspect things in the REPORT easily, and check your population equations. You can turn this code off once you need to.
- Check that your numbers at age and spawning biomass calculations are working correctly by setting the fishing mortality rates to be (close to) 0. If the population starts with a recruitment equal to R_0, in the absence of fishing (and recruitment variability), the stock should stay in equilibrium at the unfished state and spawning biomass in each year should equal SSB_0.
- When you are troubleshooting the recruitment deviation specification, estimate these as fixed effects and don't estimate σ_R. Once you have something working you can then change the dev vector to random effects.
- There are lots of bonus points to be gained, even if work on those sections doesn't include the complete analytical solution.