Fitting multispecies surplus production models

Kristin Marshall

Background

This code is meant to replicate the approach of Mueter and Megrey (2006) to fit multispecies surplus production models to stock assessment output (check this) and catches.

The general approach is:

1. Select species in "guild"

2. Pull adult Biomass (B) and Catch (C) information from stock assessments

3. Calculate annual surplus production (ASP) for each stock

4. Sum annual ASP and B across species to the guild-level

5. Plot multispecies ASP v B

6. Fit a schaefer model or Pella-Tomlinson model

7. Expand to include environmental covariates

Select species and ages

For the Norwegian example we are looking at the pelagic guild of zooplankton predators: Atlantic Mackerel, Blue Whiting, and Norwegian Spring Spawning Herring.

Pull data from stock assessments. I thought this was going to be easy - pulling assessment output directly from stockassessment.org. It turns out I can only do that for mackerel. The 2022 blue whiting and norwegian spring spawn herring assessments aren't up there. So, I need to pull the numbers out of tables in the WGWIDE 2022 report.

For all species, I am using spawning biomass, following Mueter's "adult" biomass in the paper.

All the raw data I copied from the WGWIDE report is in the excel file NorwegianAssessmentsData.xls. From this, I manually made .csv files for each species.

Data processing

First we need to pull in the data, align the years and make sure all of the units are the same. Note the herring biomass and catch need to be multiplied by 1000 to be equivalent to the others.

```r
whiting.raw <- read.table("data/norway-bluewhiting.csv", header=T, sep=",")

whiting <- whiting.raw |>
  mutate(Year = str_squish(Year)) |>
  mutate(SSB = str_squish(SSB)) |>
  mutate(Catch = str_squish(Catch)) |>
  mutate_if(is.character, as.numeric)

herring.raw <- read.table("data/norway-ssherring.csv", header=T, sep=",")

herring <- herring.raw |>
  mutate(Year = str_squish(Year)) |>
  mutate(SSB = str_squish(SSB)) |>
  mutate(Catch = str_squish(Catch)) |>
  mutate_if(is.character, as.numeric) |>
  mutate(SSB = SSB*1000) |>
  mutate(Catch = Catch*1000)

mackerel.raw <- read.table("data/norway-mackerel.csv", header=T, sep=",")

mackerel <- mackerel.raw |>
  mutate(Year = str_squish(Year)) |>
  mutate(SSB = str_squish(SSB)) |>
  mutate(Catch = str_squish(Catch)) |>
  mutate_if(is.character, as.numeric)

ms.dat <- mackerel |>
  inner_join(whiting, by= "Year", suffix=c(".mack",".whit")) |>
  inner_join(herring, by = "Year") |>
  mutate(SSB.herr = SSB, .keep = "unused") |>
  mutate(Catch.herr = Catch, .keep = "unused")
```

Calculate surplus production

Now we need to calculate annual surplus production for each stock and year. Note I'm ignoring the delta parameter that Mueter mentions as a correction factor on C, at least for now.

$$ASP_{i,j} = B_{i+1,j} - B_{i,j} + C_{i,j}$$

```
ms.dat <- ms.dat |>
  mutate(asp.mack = SSB.mack - lag(SSB.mack) + lag(Catch.mack)) |>
  mutate(asp.whit = SSB.whit - lag(SSB.whit) + lag(Catch.whit)) |>
  mutate(asp.herr = SSB.herr - lag(SSB.herr) + lag(Catch.herr))
```

Then we need to sum these across the stocks to get guild-level ASP and catch.

```
ms.long <- ms.dat |>
  mutate(asp.sum = asp.mack + asp.whit + asp.herr) |>
  mutate(Catch.sum = Catch.mack + Catch.whit + Catch.herr) |>
  mutate(SSB.sum = SSB.mack + SSB.whit + SSB.herr)|>
  pivot_longer(cols=c(!Year)) |>
  separate_wider_delim(cols=name, names=c("type", "species"), delim = ".")  |>
  filter(species == "sum")
```

Now plot multispecies biomass, catch, and asp through time, similar to Mueter's Fig 2. I added simple smoothers on year to visualize any trends. It looks like production, catch, and biomass all increased until around 2003 and have been fairly steady since then.
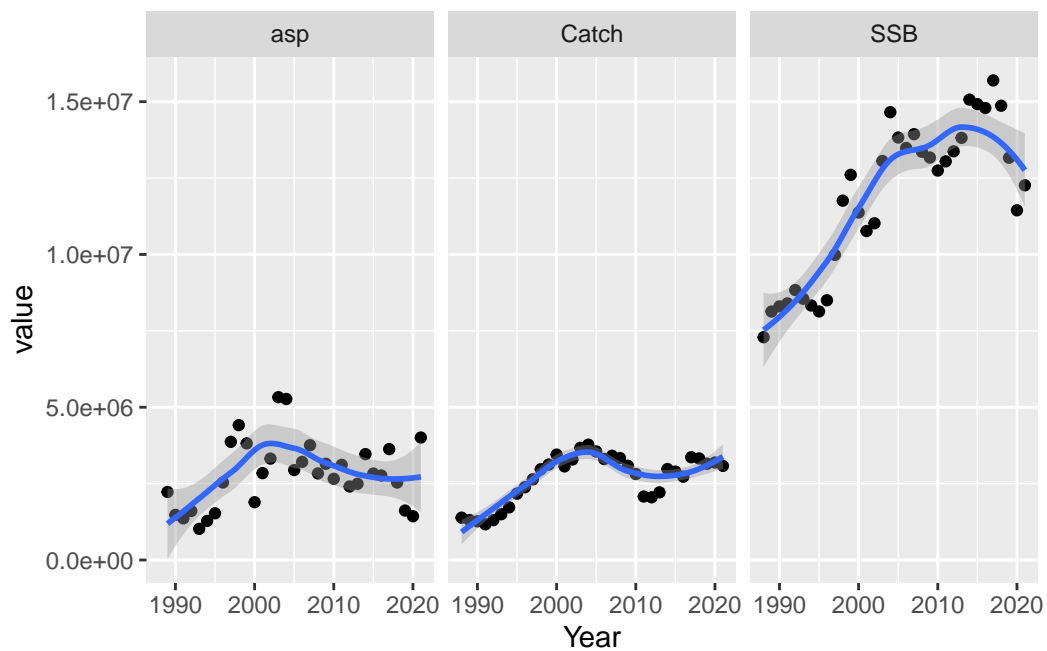
Here is plot of the catch composition through time. Similar to Mueter's Fig 3.
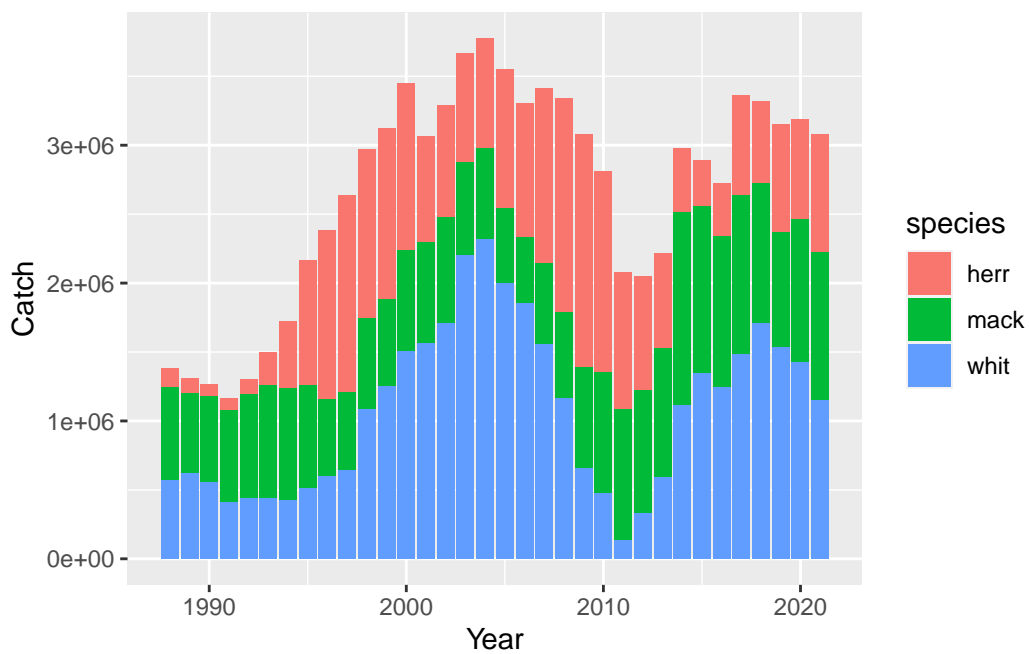
Here is a plot of Biomass by ASP by species, with years labeled. Similar to Mueter's Fig 4, but mine isn't very clean.
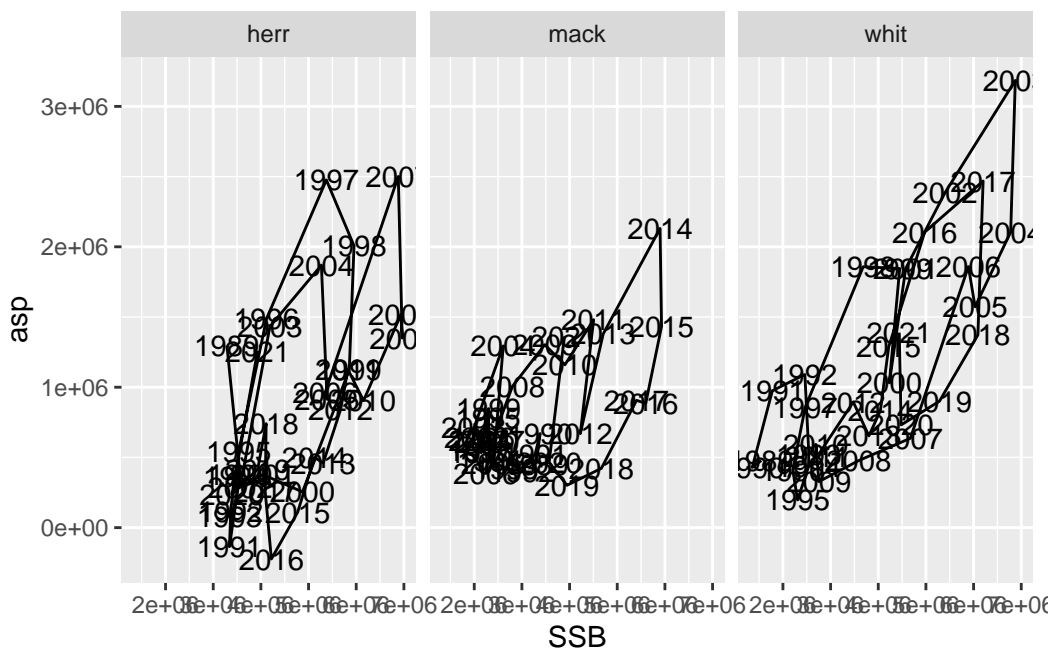
Next, plot summed biomass vs summed asp. This is what we will fit the Schaefer or Pela-Tomlinson curve to, and are the data that go into Mueter's Fig 5. Here I plotted a quadratic glm on top just to see what it looks like. It doesn't look like there will be very strong support for production declining at higher biomass.
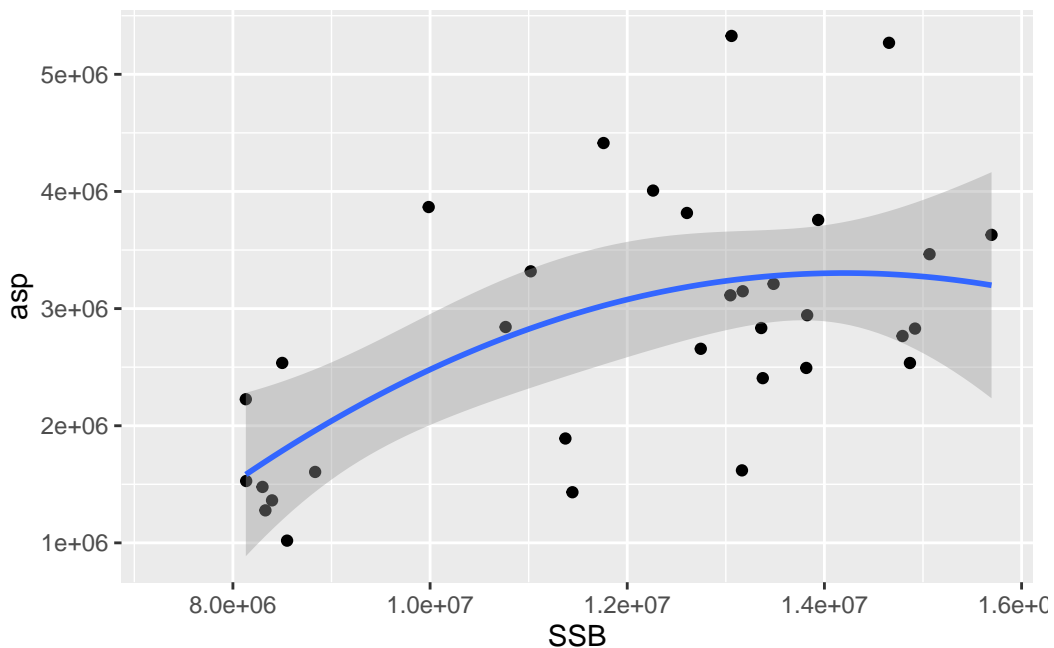
Fit schaefer production model to multispecies asp, biomass, and catch

The Schaefer model is just a multiple regression between ASP, linear and quadratic terms for B. We can fit this as a glm, forcing the intercept to be zero (unlike the above plot where the intercept was estimated.

```
m1 <- glm(asp ~ 0+ SSB + I(SSB^2), data = ms.wide)
summary(m1)
```

```
Call:
glm(formula = asp ~ 0 + SSB + I(SSB^2), data = ms.wide)

Deviance Residuals:
     Min        1Q     Median        3Q        Max
-1473103   -681968   -303233     483697    2260319

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
SSB        2.359e-01  8.366e-02   2.819  0.00832 **
I(SSB^2) -7.277e-11  6.423e-09  -0.011  0.99103
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 8.6432e+11)

    Null deviance: 2.9860e+14  on 33  degrees of freedom
Residual deviance: 2.6794e+13  on 31  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 1004.6

Number of Fisher Scoring iterations: 2
```

If we predict outside the range of the data, we should be able to see where the curve starts to bend down. However, we don't see that because the quadratic term is estimated to be almost zero when I force the intercept to be zero. I think this is telling us that the Schaefer model isn't a good fit to the data.

```
pframe <- data.frame(SSB = with(ms.wide, seq(0,2.5*10^7, length=100)))
pframe$pred <- NA
pframe$pred <- predict(m1, newdata = pframe)
```

```
ggplot(ms.wide, aes(SSB,asp)) +
  geom_point() +
  geom_line(aes(SSB,pred), data=pframe)
```
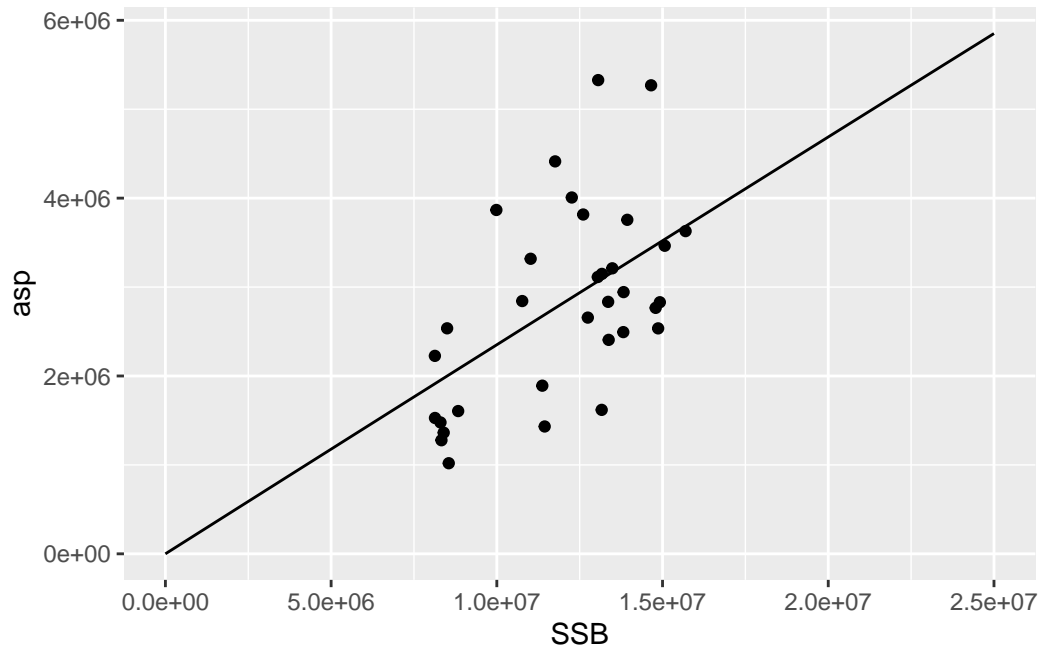

Fit Schaefer and Pella-Tomlinson using optim

Next, let's try to fit the P-T model using optim to see if we can get something that makes sense with a different exponent (other than 2, which is what the schaefer model assumes).

Something isn't working here right now. Optim is converging but the parameters don't make sense, especially the observation error. I'll need to play with initial conditions and debug on Monday...

```
pellatom <- function(pars, asp, bio){

  alpha <-pars[1]

  beta <-pars[2]

  nu <- pars[3]
```

```r
  sigma <- pars[4]

  asp.pred <- alpha*bio + beta*bio^(nu)

  output <- sum(dnorm(log(asp),
                      log(max(0.00001,asp.pred)),
                      sigma,
                      log=TRUE),
                na.rm =T)

  return(output)
}

pars.init <- c(0.02, 0.01, 2, 5)

ms.wide <- filter(ms.wide, !is.na(asp))
asp <- ms.wide$asp
bio <- ms.wide$SSB

res <- optim(pars.init, pellatom, asp=asp, bio=bio, hessian=TRUE)
pars<-res$par

ms.wide$pred <- pars[1]*bio + pars[2]*bio^(pars[3])

ggplot(ms.wide, aes(SSB,asp)) +
  geom_point() +
  geom_line(aes(SSB,pred))
```