Data weighting in SS

Johnson, K.F., Doering, K.L., Taylor, I.G., and Wetzel, C.R.

Last update: 28 October, 2020

Introduction

Todo: read Punt and Francis papers.

Methods

Overview

Monte Carlo simulations were used to evaluate the performance of three methods of data weighting for stock assessment models. Particularly, Stock Synthesis (citation), an integrated age-structured population modeling framework used to conduct assessments of marine fish populations throughout the world, was used to both generate the truth (operating model; OM) and to estimate the status (estimation method; EM). Combinations of OM, data generation, and EM (hereafter referred to as scenarios) consisted of the following three steps:

- simulate a marine fish population for 100 years with process error,
- estimate quantities of interest by fitting the EM to data sampled with observation error from the OM, and
- compare the estimates to the truth.

```
thedir <- "c:/Users/kelli/Documents/ss3sim_nin"
niter <- 10
fixx <- c(2, 3)
fixn <- seq(25, 200, by = 25)
name.om <- "hakeom"
name.em <- "hakeem"
library(ggplot2)
devtools::load_all("c:/stockAssessment/SS/ss3sim")
```

```
## Loading ss3sim
```

```
vals <- paste0("c(", outer(fixx, fixn, paste, sep = ","), ")")
grid <- expand.grid(fixx, fixn)
scen <- data.frame(
```

```
  "cf.years.1" = "26:100",
  "cf.fvals.1" = 0.747,
  "si.years.2" = "seq(76, 100, by = 1)",
  "si.sds_obs.2" = 0.2,
  "sl.Nsamp.1" = 50,
  "sl.Nsamp.2" = 50,
  "sl.years.1" = "seq(26, 100, by = 1)",
  "sl.years.2" = "seq(76, 100, by = 1)",
  "sl.cpar" = "NULL",
  "sa.Nsamp.1" = 100,
  "sa.Nsamp.2" = 100,
  "sa.years.1" = "seq(26, 100, by = 1)",
  "sa.years.2" = "seq(76, 100, by = 1)",
  "sa.cpar" = "NULL",
  "wc.method" = "DM",
  "wc.fleets" = "1:2",
  "wc.niters_weighting" = 3,
  "co.par_name" = 'c("Age_DblN_peak_Survey(2)")',
  "ce.par_name" = 'c("AgeSel_P_1_Survey")',
  "ce.par_phase" = "c(NA)",
  "om" = file.path(thedir, name.om),
  "em" = file.path(thedir, name.em)
)
scen.all <- data.frame(scen,
  "co.par_int" = grid[, 1],
  "ce.par_int" = grid[, 1],
  "sl.ESS" = grid[, 2])
scen.all <- do.call("rbind", replicate(3, scen.all, simplify = FALSE))
scen.all[, "wc.method"] <- rep(c("DM", "Francis", "MI"), each = NROW(grid))
# Set scenario names if you want
scen.all[, "scenarios"] <- paste0("nin-", seq(1,NROW(scen.all)))
```

Operating model

The simulation was based on a hake-like life history, where the OMs largely used parameters estimated from the 2011 (citation) and 2019 (citation) stock assessments for Pacific hake (*Merluccius productus*). All simulated populations were unfished during the first 25 years to allow processes error to propagate throughout each age. Fishing was implemented in year 26 using a constant level of instantaneous fishing mortality (F) for the remaining 75 years of the simulation.

Selectivity was the largest investigated axis of uncertainty in the OMs.

- First, we decreased the age of inflection for the survey from three to two.
- Second, we implemented time-varying selectivity in the fishery (todo).

Data sampling

Data were sampled from the OMs with observation error to simulate how empirical data are gathered.

Catch data were not subject to observation and were assumed to be known without error.

Sample size for length- and age-composition data were was the largest axis of uncertainty investigated in the data-sampling process.

- Sample sizes increased throughout time for both the fishery and the survey (todo).

Estimation method

The EMs were fit to data generated during the data-sampling process and were largely based on the OMs.

Axes of uncertainty investigated in the EMs included variations of how selectivity was parameterized and how compositional data sources were weighted.

Changes in data weighting were investigated using input sample size and data-weighting algorithms.

- Input sample sizes were either constant throughout time or time varying (todo) and ranged from 25 to 200.
- Compositional data were weighted internally to the EM using Dirichlet-Multinomial parameters or were weighted using the Francis (citation) or McAlister-Ianelli (citation) approach.

```
run_ss3sim(iterations = 1:niter, simdf = scen.all[-(1:4),])
get_results_all()
```

Results

- Input sample sizes smaller than the true value will always lead to underestimation of the true sample size
- As input sample sizes become more positively biased relative to the true sample size, the estimated effective sample size was more negatively biased
- Error in estimates of effective sample size lead to error in estimates of growth parameters
- Estimation of age-composition weightings are independent of successful estimation of length-composition weightings

Tables

Tables suck.

Figures

Figures with captions

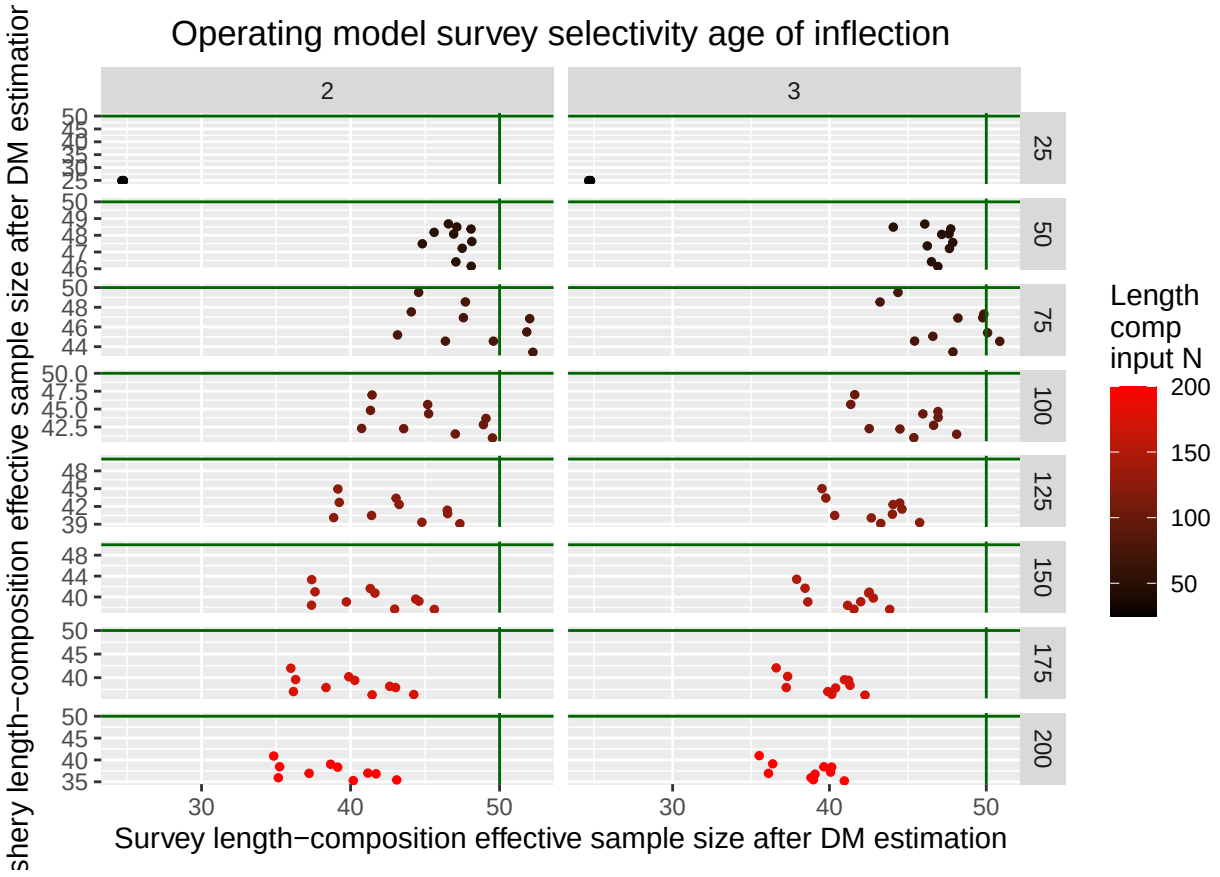


Figure 1: Effective sample sizes for length-composition data after estimating Dirichlet-Multinomial (DM) relative weighting parameters. Colors indicate the input sample size and solid green lines are the true sample size used in the operating model (OM). Columns reflect a change in survey selectivity in the OM. Note the change in the scale for each y axis.

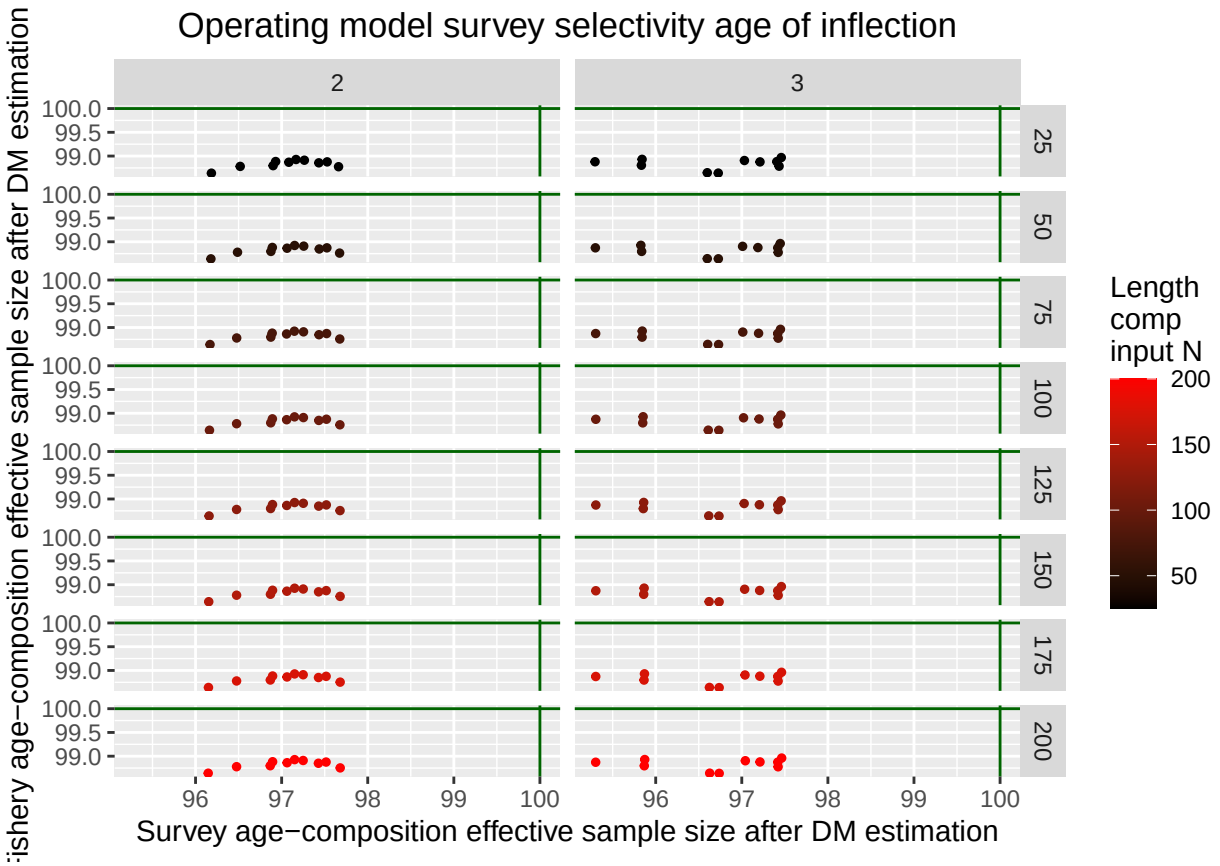
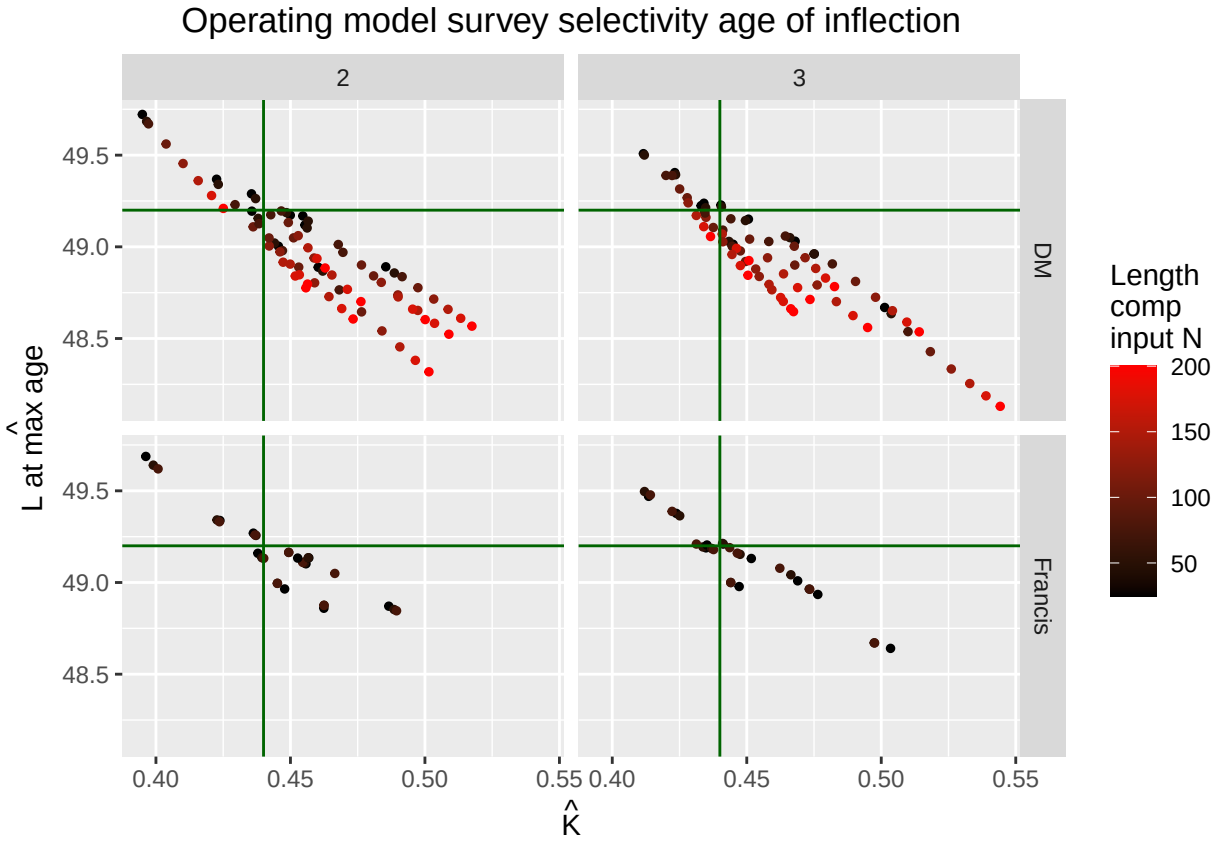
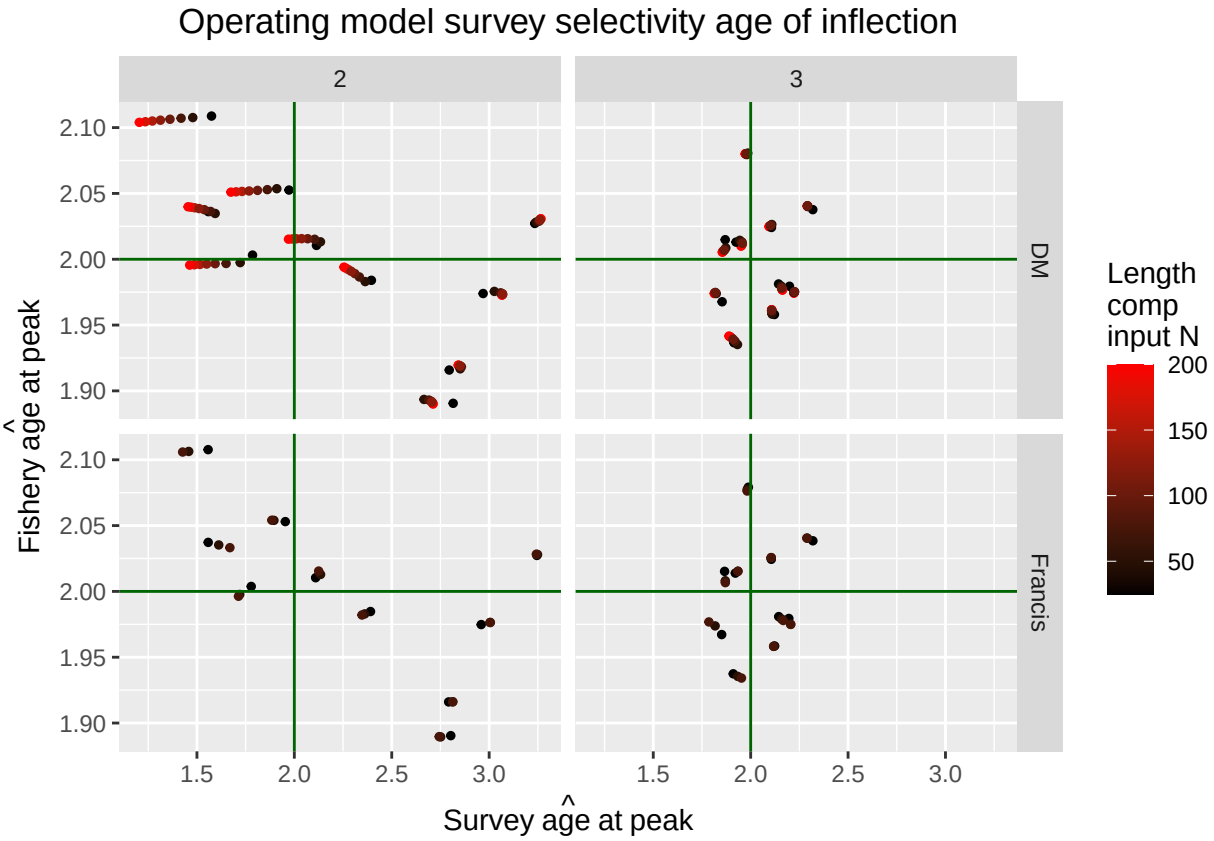


Figure 2: Effective sample sizes for age-composition data after estimating Dirichlet-Multinomial (DM) relative weighting parameters. Colors indicate the input sample size for length-composition data, where the true was 50, and solid green lines are the true sample size used in the operating model (OM). Columns reflect a change in survey selectivity in the OM.

Figures without captions




Operating model survey selectivity age of inflection
2
3
DM
Francis
2.10
2.05
2.00
1.95
1.90
1.5
2.0
2.5
3.0
^Fishery age at peak
^Survey age at peak
Length comp input N
200
150
100
50