Fitting Integrated Population Models to Lower Columbia River Chum Salmon Monitoring Data

Eric Buhle, Kale Bentley, Thomas Buehrens, Mark Scheuerell, Todd Hillson, and …

2020-07-18

Contents

1 Overview

Background on IPMs, outline of **salmonIPM**…

2 Model Description

2.1 Process Model

2.1.1 Egg Deposition

Generic spawner-recruit function

$$E_{jt} = f(S_{jt}|\alpha_{jt}, E_{\max,j})$$

Intrinsic productivity is calculated as a weighted sum of mean age-specific female fecundity, weighted by the spawner age distribution, divided by 2 assuming a 1:1 sex ratio

$$\alpha_{jt} = \frac{1}{2} \sum_{a=3}^{5} q_{jta} \mu_{\text{fec},a}$$

Maxmum egg deposition ("capacity") varies randomly among populations according to the hyper-distribution

$$\log(E_{\max,j}) \sim N(\mu_{E_{\max}}, \sigma_{E_{\max}})$$

Three spawner-recruit functional forms

$$f(S_{jt}|\alpha_{jt}, E_{\max,j}) = \begin{cases} \alpha_{jt} S_{jt} & \text{exponential} \\ \dfrac{\alpha_{jt} S_{jt}}{1 + \alpha_{jt} S_{jt} / E_{\max,j}} & \text{Beverton-Holt} \\ \alpha_{jt} S_{jt} \text{exp} \left(- \dfrac{\alpha_{jt} S_{jt}}{\exp(1) E_{\max,j}} \right) & \text{Ricker} \end{cases}$$

2.1.2 Egg-to-Smolt Survival

$$\begin{aligned} \text{logit}(s_{EM,jt}) &= \text{logit}(\mu_{EM}) + \eta_{EM,j}^{\text{pop}} + \eta_{EM,t}^{\text{year}} + \epsilon_{EM,jt} \\ \eta_{EM,j}^{\text{pop}} &\sim N(0, \sigma_{EM}^{\text{pop}}) \\ \eta_{EM,t}^{\text{year}} &\sim N(\rho_{EM} \eta_{EM,t-1}^{\text{year}}, \sigma_{EM}^{\text{year}}) \\ \epsilon_{EM,jt} &\sim N(0, \sigma_{EM}) \end{aligned}$$

Smolts in year t

$$M_{jt} = s_{EM,t-1} E_{t-1}$$

2.1.3 Smolt-to-Adult Survival

SAR

$$
\begin{aligned}
\text{logit}(s_{MS,jt}) &= \text{logit}(\mu_{MS}) + \eta^{\text{year}}_{MS,t} + \epsilon_{MS,jt} \\
\eta^{\text{year}}_{MS,t} &\sim N(\rho_{MS}\eta^{\text{year}}_{MS,t-1}, \sigma^{\text{year}}_{MS}) \\
\epsilon_{MS,jt} &\sim N(0, \sigma_{MS})
\end{aligned}
$$

2.1.4 Conditional Age-at-Return

Adult age structure is modeled by defining a vector of conditional probabilities, $\mathbf{p}_{jt} = [p_{3jt}, p_{4jt}, p_{5jt}]^\top$, where p_{ajt} is the probability of an outmigrant in year t in population j returning at age a, given that it survives to adulthood. The unconditional probability is given by $s_{MS,jt}pajt$, where both SAR and p_a are functions of underlying annual marine survival and maturation probabilities that are nonidentifiable without some ancillary data. This parameterization resolves the nonidentifiability.

The conditional age probabilities follow a logistic normal process model with hierarchical structure across populations and through time within each population. The additive log ratio,

$$
\text{alr}(\mathbf{p_{jt}}) = [\log(p_{3jt}/p_{5jt}), \log(p_{4jt}/p_{5jt})]^\top
$$

has a bivariate normal distribution:

$$
\begin{aligned}
\text{alr}(\mathbf{p_{jt}}) &= \text{alr}(\boldsymbol{\mu}_\mathbf{p}) + \boldsymbol{\eta}^{\text{pop}}_{\mathbf{p},j} + \boldsymbol{\epsilon}_{\mathbf{p},jt} \\
\boldsymbol{\eta}^{\text{pop}}_{\mathbf{p},j} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}^{\text{pop}}_\mathbf{p}) \\
\boldsymbol{\epsilon}_{\mathbf{p},jt} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{p}).
\end{aligned}
$$

Here the 2×2 covariances matrices $\boldsymbol{\Sigma}^{\text{pop}}_\mathbf{p}$ and $\boldsymbol{\Sigma}_\mathbf{p}$ allow correlated variation among age classes (on the unconstrained scale, not merely due to the mathematical simplex constraint on $\mathbf{p}$) across populations and through time within a population, respectively. For example, some populations or cohorts may skew overall younger or older than average. We parameterize each covariance matrix by a vector of standard deviations and a correlation matrix:

$$
\begin{aligned}
\boldsymbol{\Sigma}^{\text{pop}}_\mathbf{p} &= \boldsymbol{\sigma}^{\text{pop}}_\mathbf{p}\mathbf{R}^{\text{pop}}_\mathbf{p}\boldsymbol{\sigma}^{\text{pop}\top}_\mathbf{p} \\
\boldsymbol{\Sigma}_\mathbf{p} &= \boldsymbol{\sigma}_\mathbf{p}\mathbf{R}_\mathbf{p}\boldsymbol{\sigma}^\top_\mathbf{p}
\end{aligned}
$$

2.1.5 Adult Recruitment

Survival to adults at age, broodstock removal assumed known, harvest assumed to be zero for now

$$
S_{\text{W},jt} = \left(\sum_{a=3}^{5} s_{MS,j,t-a}\, p_{aj,t-a}\, M_{j,t-a}\right) - B_{jt} = \left(\sum_{a=3}^{5} \tilde{S}_{\text{W},ajt}\right) - B_{jt}
$$

Spawner age structure is $\mathbf{q}_{jt} = [q_{3jt}, q_{4jt}, q_{5jt}]$, where $q_{ajt} = \tilde{S}_{\text{W},ajt}/S_{jt}$.

Wild vs. hatchery spawners

$$S_{\mathrm{H},jt} = S_{\mathrm{W},jt} p_{\mathrm{HOS},jt} / (1 - p_{\mathrm{HOS},jt})$$

Total spawner abundance is then $S_{jt} = S_{\mathrm{W},jt} + S_{\mathrm{H},jt}$.

2.2 Observation Model

2.2.1 Fecundity

We modeled observations of fecundity from individual female chum salmon collected at hatcheries. The likelihood for the fecundity of female i of age a is a zero-truncated normal with age-specific mean and SD.

$$E^{\mathrm{obs}}_{a,i} \sim N(\mu_{E,a}, \sigma_{E,a})\, T[0, \infty)$$

2.2.2 Smolt and Spawner Abundance

Informative priors based on Bayesian observation models applied to field data of various kinds

$$\begin{aligned}
\log(M_{jt}) &\sim N(\mu_{M,jt}, \tau_{M,ij}) \\
\log(S_{jt}) &\sim N(\mu_{S,jt}, \tau_{S,ij})
\end{aligned}$$

Some prior observation error SDs are missing or unknown, and so were imputed by fitting a log-normal hyperdistribution to the known SDs

$$\begin{aligned}
\log(\tau_{M,ij}) &\sim N(\mu_{\tau_M}, \sigma_{\tau_M}) \\
\log(\tau_{S,ij}) &\sim N(\mu_{\tau_S}, \sigma_{\tau_S})
\end{aligned}$$

2.2.3 Spawner Age and Origin Composition

Age composition of wild spawners $\mathbf{n}^{\mathrm{obs}}_{ajt} = [n^{\mathrm{obs}}_{3jt}, n^{\mathrm{obs}}_{4jt}, n^{\mathrm{obs}}_{5jt}]^\top$ is assumed to follow a multinomial likelihood with the expected proportions given by the unobserved true state

$$\mathbf{n}^{\mathrm{obs}}_{ajt} \sim \mathrm{Multinomial}\left(\sum_a n^{\mathrm{obs}}_{ajt}, \mathbf{q}_{jt}\right)$$

Hatchery/wild composition of spawners

$$n^{\mathrm{obs}}_{\mathrm{H},jt} \sim \mathrm{Bin}\left(n^{\mathrm{obs}}_{\mathrm{W},jt} + n^{\mathrm{obs}}_{\mathrm{H},jt}, p_{\mathrm{HOS},jt}\right)$$

2.3 Priors

3 Setup and Data

Load the packages we'll need…

```
options(device = ifelse(.Platform$OS.type == "windows", "windows", "quartz"))
options(mc.cores = parallel::detectCores(logical = FALSE) - 1)

library(salmonIPM)
library(rstan)
library(shinystan)
library(matrixStats)
library(Hmisc)
library(tibble)
library(dplyr)
library(tidyr)
library(reshape2)
library(yarrr)
library(magicaxis)
library(viridis)
library(zoo)
library(here)

if(file.exists(here("analysis","results","LCRchumIPM.RData")))
  load(here("analysis","results","LCRchumIPM.RData"))
```

Read in and manipulate the data…

```
# Mapping of location to population
location_pop <- read.csv(here("data","Location.Reach_Population.csv"),
                         header = TRUE, stringsAsFactors = TRUE) %>%
  rename(strata = Strata, location = Location.Reach, pop1 = Population1, pop2 = Population2)

# Mapping of disposition to hatchery vs. wild (i.e., broodstock vs. natural spawner)
disposition_HW <- read.csv(here("data","Disposition_HW.csv"),
                           header = TRUE, stringsAsFactors = TRUE) %>%
  rename(disposition = Disposition) %>% arrange(HW)

# Start dates of hatcheries associated with populations
hatcheries <- read.csv(here("data","Hatchery_Programs.csv"), header = TRUE, stringsAsFactors =

# Spawner abundance data
# Assumptions:
# (1) NAs in hatchery dispositions (incl. Duncan Channel) are really zeros
# (2) NAs in Duncan Creek from 2004-present are really zeros
# (3) All other NAs are real missing observations
```

```r
# (4) When calculating the observation error of log(S_obs), tau_S_obs, assume
#     Abund.Mean and Abund.SD are the mean and SD of a lognormal posterior distribution
#     of spawner abundance based on the sample
spawner_data <- read.csv(here("data","Data_ChumSpawnerAbundance_2019-12-12.csv"),
                         header = TRUE, stringsAsFactors = TRUE) %>%
  rename(year = Return.Yr., strata = Strata, location = Location.Reach,
         disposition = Disposition, method = Method, S_obs = Abund.Mean, SD = Abund.SD) %>%
  mutate(pop = location_pop$pop2[match(location, location_pop$location)],
         disposition_HW = disposition_HW$HW[match(disposition, disposition_HW$disposition)],
         S_obs = replace(S_obs, is.na(S_obs) & disposition_HW == "H", 0),
         S_obs = replace(S_obs, is.na(S_obs) & pop == "Duncan_Creek" & year >= 2004, 0),
         tau_S_obs = sqrt(log((SD/S_obs)^2 + 1))) %>%
  select(year:location, pop, disposition, disposition_HW, method:tau_S_obs) %>%
  arrange(strata, location, year)

names_S_obs <- disposition_HW$disposition
names_B_take_obs <- disposition_HW$disposition[disposition_HW$HW == "H"]

spawner_data_agg <- spawner_data %>% group_by(strata, pop, year, disposition) %>%
  summarize(S_obs = sum(S_obs), tau_S_obs = unique(tau_S_obs)) %>% ungroup() %>%
  pivot_wider(id_cols = c(strata, pop, year), names_from = disposition,
              values_from = c(S_obs, tau_S_obs), values_fill = list(S_obs = 0)) %>%
  add_column(S_obs = rowSums(select(., all_of(paste0("S_obs_", names_S_obs)))),
             tau_S_obs = rowSums(select(., all_of(paste0("tau_S_obs_", names_S_obs))), na.rm =
             B_take_obs = rowSums(select(., all_of(paste0("S_obs_", names_B_take_obs))))) %>%
  mutate(tau_S_obs = replace(tau_S_obs, tau_S_obs == 0, NA)) %>%
  select(-matches(paste(names_S_obs, collapse = "|"))) %>%
  as.data.frame()

# Spawner age-, sex-, and origin-frequency (aka BioData)
bio_data <- read.csv(here("data","Data_ChumSpawnerBioData_2019-12-12.csv"),
                     header = TRUE, stringsAsFactors = TRUE) %>%
  rename(year = Return.Yr., strata = Strata, location = Location.Reach,
         disposition = Disposition, origin = Origin, sex = Sex, age = Age, count = Count) %>%
  mutate(pop = location_pop$pop2[match(location, location_pop$location)],
         origin_HW = ifelse(origin == "Natural_spawner", "W", "H"),
         count = ifelse(is.na(count), 0, count)) %>%
  select(year:location, pop, disposition, origin, origin_HW, sex:count) %>%
  arrange(strata, location, year, origin, age, sex)

# age of wild spawners only
bio_data_age <- bio_data %>% filter(origin_HW == "W") %>%
  dcast(year + strata + pop ~ age, value.var = "count", fun.aggregate = sum)

bio_data_origin <- bio_data %>%
  dcast(year + strata + pop ~ origin_HW, value.var = "count", fun.aggregate = sum)

# Juvenile abundance data
```

```r
# Assumptions:
# (1) Smolts from Duncan Channel represent all naturally produced offspring of spawners
#     in Duncan Creek (hence Duncan_Channel -> Duncan_Creek in location_pop)
# (2) Duncan_North + Duncan_South = Duncan_Channel, so the former two are redundant
#     (not really an assumption, although the equality isn't perfect in all years)
# (3) When calculating the observation error of log(M_obs), tau_M_obs, assume
#     Abund_Mean and Abund_SD are the mean and SD of a lognormal posterior distribution
#     of smolt abundance based on the sample
# (4) If Abund_SD == 0 (when Analysis=="Census": some years in Duncan_Creek and
#     Hamilton_Channel) treat as NA
juv_data <- read.csv(here("data", "Data_ChumJuvenileAbundance_2020-06-09.csv"),
                     header = TRUE, stringsAsFactors = TRUE) %>%
  rename(brood_year = Brood.Year, year = Outmigration.Year, strata = Strata,
         location = Location.Reach, origin = Origin, trap_type = TrapType,
         analysis = Analysis, partial_spawners = Partial.Spawners, raw_catch = RawCatch,
         M_obs = Abund_Median, mean = Abund_Mean, SD = Abund_SD,
         L95 = Abund_L95, U95 = Abund_U95, CV = Abund_CV, comments = Comments) %>%
  mutate(pop = location_pop$pop2[match(location, location_pop$location)],
         tau_M_obs = replace(sqrt(log((SD/mean)^2 + 1)), SD==0, NA)) %>%
  select(strata, location, pop, year, brood_year, origin:CV, tau_M_obs, comments) %>%
  arrange(strata, location, year)

# drop hatchery or redundant pops and cases with leading or trailing NAs in M_obs
head_noNA <- function(x) { cumsum(!is.na(x)) > 0 }
juv_data_incl <- juv_data %>% filter(pop %in% spawner_data$pop) %>%
  mutate(location = factor(location), pop = factor(pop, levels = levels(spawner_data$pop))) %>%
  group_by(pop) %>% filter(head_noNA(M_obs) & rev(head_noNA(rev(M_obs)))) %>% as.data.frame()

# Fish data formatted for salmonIPM
# Drop age-2 and age-6 samples (each is < 0.1% of aged spawners)
# Use A = 1 for now (so Rmax in units of spawners)
fish_data <- full_join(spawner_data_agg, bio_data_age, by = c("strata","pop","year")) %>%
  full_join(bio_data_origin, by = c("strata","pop","year")) %>%
  full_join(juv_data_incl, by = c("strata","pop","year")) %>%
  mutate(B_take_obs = replace(B_take_obs, is.na(B_take_obs), 0)) %>%
  rename_at(vars(contains("Age-")), list(~ paste0(sub("Age-","n_age",.), "_obs"))) %>%
  select(-c(n_age2_obs, n_age6_obs)) %>%
  rename(n_H_obs = H, n_W_obs = W) %>% mutate(A = 1, fit_p_HOS = NA, F_rate = 0) %>%
  mutate_at(vars(contains("n_")), ~ replace(., is.na(.), 0)) %>%
  select(strata, pop, year, A, S_obs, tau_S_obs, M_obs, tau_M_obs, n_age3_obs:n_W_obs,
         fit_p_HOS, B_take_obs, F_rate) %>% arrange(strata, pop, year)

# fill in fit_p_HOS
for(i in 1:nrow(fish_data)) {
  pop_i <- as.character(fish_data$pop[i])
  start_year <- ifelse(pop_i %in% hatcheries$pop,
                       min(hatcheries$start_brood_year[hatcheries$pop == pop_i]) + 1,
                       NA)
```

```r
  fish_data$fit_p_HOS[i] <- ifelse((!is.na(start_year) & fish_data$year[i] >= start_year) |
                                     fish_data$n_H_obs[i] > 0, 1, 0)
}

# # drop cases with initial NAs in S_obs unless bio data is present
# fish_data <- fish_data %>% mutate(n_age = rowSums(select(., n_age2_obs:n_age6_obs))) %>%
#   group_by(pop) %>%  filter(head_noNA(S_obs) | cumsum(n_age) > 0) %>%
#   select(-n_age) %>% as.data.frame()

# subsets for models with specific stage structure
# spawner-spawner: drop cases with initial NAs in S_obs, even if bio data is present
fish_data_SS <- fish_data %>% group_by(pop) %>% filter(head_noNA(S_obs)) %>% as.data.frame()
# spawner-spawner: drop cases with initial NAs in M_obs, even if bio data is present
fish_data_SMS <- fish_data %>% group_by(pop) %>%
  filter(head_noNA(S_obs) | head_noNA(M_obs)) %>% as.data.frame()

# pad data with future years to generate forecasts
# use 5-year (1-generation) time horizon
fish_data_SMS_fore <- fish_data_SMS %>% group_by(pop) %>%
  slice(rep(n(), max(fish_data_SMS$year) + 5 - max(year))) %>%
  mutate(year = (unique(year) + 1):(max(fish_data_SMS$year) + 5),
         S_obs = NA, tau_S_obs = NA, M_obs = NA, tau_M_obs = NA,
         fit_p_HOS = 0, B_take_obs = 0, F_rate = 0) %>%
  mutate_at(vars(starts_with("n_")), ~ 0) %>%
  full_join(fish_data_SMS) %>% arrange(pop, year) %>%
  mutate(forecast = year > max(fish_data_SMS$year)) %>%
  select(strata:year, forecast, A:F_rate) %>% as.data.frame()

# Fecundity data
# Note that L95% and U95% are reversed
fecundity <- read.csv(here("data","Data_ChumFecundity_fromHatcheryPrograms_2017-01-25.csv"),
                      header = TRUE, stringsAsFactors = TRUE) %>%
  rename(stock = Stock, year = BY, ID = Female.., age_E = Age, L95 = U95., U95 = L95.,
         reproductive_effort = Reproductive.Effort, E_obs = Estimated.Fecundity,
         mean_mass = Green.egg.avg.weight, comments = Comments) %>%
  mutate(ID = as.character(ID))

# drop cases with age not in c(3,4,5) or with estimated fecundity missing
# add strata based on stock: Grays -> Coastal, I-205 -> Cascade, Lower Gorge -> Gorge
fecundity_data <- fecundity %>% filter(age_E %in% 3:5 & !is.na(E_obs)) %>%
  mutate(strata = recode(stock, Grays = "Coastal", `I-205` = "Cascade", `Lower Gorge` = "Gorge"
  select(strata, year, ID, age_E, E_obs) %>%
  arrange(strata, year, age_E)
```

Let's look at the first few rows of `fish_data` to see the format **salmonIPM** expects…

```r
head(fish_data_SMS)
```

```
   strata        pop year A S_obs  tau_S_obs M_obs tau_M_obs n_age3_obs n_age4_obs n_age5_obs n
1 Cascade Cascade_MS 2002 1  3160 0.15231109    NA        NA        101        114          7
2 Cascade Cascade_MS 2003 1  2866 0.05572085    NA        NA         19        448         26
3 Cascade Cascade_MS 2004 1  2324 0.08680695    NA        NA         75        203         50
4 Cascade Cascade_MS 2005 1   923 0.03711614    NA        NA          4         38          0
5 Cascade Cascade_MS 2006 1   869 0.02342601    NA        NA          1         41          0
6 Cascade Cascade_MS 2007 1   576 0.05358997    NA        NA         32        115         43
  B_take_obs F_rate
1         15      0
2          0      0
3          0      0
4          0      0
5          0      0
6          0      0
```

4 Retrospective Models

Fit two-stage spawner-smolt-spawner models and explore output…

Density-independent

```
LCRchum_exp <- salmonIPM(fish_data = fish_data_SMS,  fecundity_data = fecundity_data,
                         ages = list(M = 1), stan_model = "IPM_LCRchum_pp", SR_fun = "exp",
                         pars = c("B_rate_all","mu_Emax","sigma_Emax","Emax"),
                         include = FALSE, log_lik = TRUE,
                         chains = 3, iter = 1500, warmup = 500,
                         control = list(adapt_delta = 0.99, max_treedepth = 14))
```

```
print(LCRchum_exp, prob = c(0.025,0.5,0.975),
      pars = c("eta_pop_EM","eta_year_EM","eta_year_MS","eta_pop_p","p",
               "tau_M","tau_S","p_HOS","E","S","M","s_EM","s_MS","q","LL"),
      include = FALSE, use_cache = FALSE)
```

```
Inference for Stan model: IPM_LCRchum_pp.
3 chains, each with iter=1500; warmup=500; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=3000.

                 mean se_mean    sd    2.5%     50%   97.5% n_eff Rhat
mu_E[1]       2332.08    0.63 40.20 2252.79 2332.00 2408.34  4125 1.00
mu_E[2]       2585.64    0.29 22.63 2540.24 2585.87 2629.91  6114 1.00
mu_E[3]       2552.86    1.26 82.05 2398.00 2551.56 2721.12  4225 1.00
sigma_E[1]     607.47    0.39 29.95  551.68  606.34  669.55  5781 1.00
sigma_E[2]     670.51    0.23 16.28  640.30  670.04  703.10  5028 1.00
sigma_E[3]     714.80    1.05 66.11  603.42  707.70  860.59  3995 1.00
mu_EM            0.71    0.00  0.08    0.53    0.71    0.86   907 1.00
sigma_pop_EM     0.85    0.01  0.32    0.36    0.80    1.59  1397 1.00
```

```
rho_EM               0.48    0.01  0.31     -0.34      0.56      0.85   455 1.01
sigma_year_EM        0.89    0.02  0.40      0.22      0.84      1.85   441 1.01
sigma_EM             1.10    0.01  0.23      0.72      1.08      1.61   466 1.01
mu_MS                0.00    0.00  0.00      0.00      0.00      0.00   894 1.00
rho_MS               0.52    0.01  0.22     -0.01      0.55      0.84  1018 1.00
sigma_year_MS        1.03    0.01  0.23      0.67      0.99      1.61  1078 1.00
sigma_MS             0.64    0.00  0.06      0.52      0.64      0.77   823 1.00
mu_p[1]              0.20    0.00  0.02      0.16      0.20      0.24   854 1.00
mu_p[2]              0.75    0.00  0.02      0.71      0.75      0.78  1031 1.00
mu_p[3]              0.05    0.00  0.01      0.04      0.05      0.06   796 1.00
sigma_pop_p[1]       0.20    0.01  0.15      0.01      0.18      0.57   653 1.00
sigma_pop_p[2]       0.14    0.00  0.11      0.01      0.12      0.43   618 1.00
R_pop_p[1,1]         1.00     NaN  0.00      1.00      1.00      1.00   NaN  NaN
R_pop_p[1,2]         0.28    0.02  0.57     -0.89      0.42      0.98  1025 1.00
R_pop_p[2,1]         0.28    0.02  0.57     -0.89      0.42      0.98  1025 1.00
R_pop_p[2,2]         1.00    0.00  0.00      1.00      1.00      1.00  1234 1.00
sigma_p[1]           1.50    0.01  0.13      1.26      1.49      1.77   570 1.00
sigma_p[2]           0.79    0.00  0.08      0.64      0.79      0.97   718 1.00
R_p[1,1]             1.00     NaN  0.00      1.00      1.00      1.00   NaN  NaN
R_p[1,2]             0.72    0.00  0.07      0.57      0.73      0.84   977 1.01
R_p[2,1]             0.72    0.00  0.07      0.57      0.73      0.84   977 1.01
R_p[2,2]             1.00    0.00  0.00      1.00      1.00      1.00    84 1.00
mu_tau_M             0.09    0.00  0.02      0.06      0.08      0.12  2230 1.00
sigma_tau_M          1.31    0.00  0.15      1.05      1.30      1.64  2270 1.00
mu_tau_S             0.12    0.00  0.01      0.10      0.12      0.14  2555 1.00
sigma_tau_S          1.00    0.00  0.06      0.90      1.00      1.14  2568 1.00
lp__            -32316.68    1.59 36.63 -32389.28 -32316.36 -32245.00   532 1.02

Samples were drawn using NUTS(diag_e) at Thu Jul 16 14:50:26 2020.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Beverton-Holt

```
LCRchum_BH <- salmonIPM(fish_data = fish_data_SMS, fecundity_data = fecundity_data,
                        ages = list(M = 1), stan_model = "IPM_LCRchum_pp", SR_fun = "BH",
                        pars = "B_rate_all", include = FALSE, log_lik = TRUE,
                        chains = 3, iter = 1500, warmup = 500,
                        control = list(adapt_delta = 0.99, max_treedepth = 14))
```

```
print(LCRchum_BH, prob = c(0.025,0.5,0.975),
      pars = c("Emax","eta_pop_EM","eta_year_EM","eta_year_MS","eta_pop_p",
               "p","tau_M","tau_S","p_HOS","E","S","M","s_EM","s_MS","q","LL"),
      include = FALSE, use_cache = FALSE)
```

```
Inference for Stan model: IPM_LCRchum_pp.
```

```
3 chains, each with iter=1500; warmup=500; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=3000.

                     mean se_mean    sd      2.5%       50%     97.5% n_eff Rhat
mu_E[1]           2332.90    0.51 40.50   2254.90   2332.64   2410.75  6324 1.00
mu_E[2]           2587.33    0.27 22.15   2544.18   2587.48   2628.76  6537 1.00
mu_E[3]           2554.18    1.02 85.35   2382.57   2552.29   2723.54  7002 1.00
sigma_E[1]         607.64    0.36 30.59    553.00    606.38    669.69  7362 1.00
sigma_E[2]         670.44    0.19 15.69    639.99    670.13    701.69  6791 1.00
sigma_E[3]         714.01    0.85 62.90    605.40    709.50    856.19  5479 1.00
mu_Emax              1.34    0.04  1.06     -0.39      1.21      3.80   754 1.00
sigma_Emax           2.56    0.02  0.93      1.33      2.40      4.97  2195 1.00
mu_EM                0.89    0.00  0.06      0.76      0.90      0.98  1206 1.00
sigma_pop_EM         0.93    0.03  0.72      0.05      0.80      2.68   654 1.00
rho_EM               0.26    0.01  0.40     -0.64      0.33      0.81   859 1.00
sigma_year_EM        1.14    0.03  0.63      0.12      1.08      2.62   616 1.01
sigma_EM             1.39    0.01  0.32      0.84      1.36      2.12   833 1.00
mu_MS                0.00    0.00  0.00      0.00      0.00      0.00  2190 1.00
rho_MS               0.51    0.01  0.23     -0.02      0.55      0.83   975 1.00
sigma_year_MS        1.04    0.01  0.24      0.68      1.00      1.61  2104 1.00
sigma_MS             0.59    0.00  0.06      0.49      0.58      0.71   780 1.00
mu_p[1]              0.20    0.00  0.02      0.17      0.20      0.24  1161 1.00
mu_p[2]              0.75    0.00  0.02      0.71      0.75      0.78  1514 1.00
mu_p[3]              0.05    0.00  0.01      0.04      0.05      0.06  1170 1.00
sigma_pop_p[1]       0.19    0.01  0.15      0.01      0.16      0.58   439 1.00
sigma_pop_p[2]       0.15    0.01  0.12      0.01      0.13      0.44   547 1.00
R_pop_p[1,1]         1.00     NaN  0.00      1.00      1.00      1.00   NaN  NaN
R_pop_p[1,2]         0.28    0.02  0.57     -0.90      0.42      0.98  1057 1.00
R_pop_p[2,1]         0.28    0.02  0.57     -0.90      0.42      0.98  1057 1.00
R_pop_p[2,2]         1.00    0.00  0.00      1.00      1.00      1.00  1360 1.00
sigma_p[1]           1.50    0.00  0.13      1.28      1.49      1.78   745 1.00
sigma_p[2]           0.80    0.00  0.09      0.65      0.80      0.98   832 1.00
R_p[1,1]             1.00     NaN  0.00      1.00      1.00      1.00   NaN  NaN
R_p[1,2]             0.72    0.00  0.07      0.57      0.73      0.84  1279 1.00
R_p[2,1]             0.72    0.00  0.07      0.57      0.73      0.84  1279 1.00
R_p[2,2]             1.00    0.00  0.00      1.00      1.00      1.00   160 1.00
mu_tau_M             0.09    0.00  0.02      0.06      0.08      0.12  3376 1.00
sigma_tau_M          1.31    0.00  0.15      1.06      1.30      1.62  3226 1.00
mu_tau_S             0.12    0.00  0.01      0.10      0.12      0.14  2833 1.00
sigma_tau_S          1.00    0.00  0.06      0.89      1.00      1.13  2377 1.00
lp__            -32316.01    1.51 35.62 -32385.25 -32316.11 -32245.67   558 1.00

Samples were drawn using NUTS(diag_e) at Wed Jul 15 20:22:52 2020.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Ricker

```r
LCRchum_Ricker <- salmonIPM(fish_data = fish_data_SMS, fecundity_data = fecundity_data,
                            ages = list(M = 1), stan_model = "IPM_LCRchum_pp", SR_fun = "Ricke
                            pars = "B_rate_all", include = FALSE, log_lik = TRUE,
                            chains = 3, iter = 1500, warmup = 500,
                            control = list(adapt_delta = 0.99, max_treedepth = 14))

print(LCRchum_Ricker, prob = c(0.025,0.5,0.975),
      pars = c("Emax","eta_pop_EM","eta_year_EM","eta_year_MS","eta_pop_p",
               "p","tau_M","tau_S","p_HOS","E","S","M","s_EM","s_MS","q","LL"),
      include = FALSE, use_cache = FALSE)
```

```
Inference for Stan model: IPM_LCRchum_pp.
3 chains, each with iter=1500; warmup=500; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=3000.

                   mean se_mean    sd    2.5%     50%   97.5% n_eff Rhat
mu_E[1]         2332.94    0.60 41.35 2251.09 2332.48 2415.04  4732 1.00
mu_E[2]         2587.05    0.30 22.65 2541.72 2586.72 2631.02  5776 1.00
mu_E[3]         2551.72    1.33 85.93 2385.53 2551.46 2727.92  4190 1.00
sigma_E[1]       607.40    0.42 29.48  553.07  605.95  668.31  4898 1.00
sigma_E[2]       670.36    0.23 16.84  638.64  670.21  704.25  5419 1.00
sigma_E[3]       713.46    0.90 61.80  604.81  709.86  849.95  4725 1.00
mu_Emax            0.96    0.05  1.18   -0.71    0.78    3.74   554 1.00
sigma_Emax         2.42    0.03  1.02    1.20    2.21    5.06  1116 1.00
mu_EM              0.87    0.00  0.06    0.73    0.88    0.97   667 1.01
sigma_pop_EM       1.03    0.04  0.70    0.09    0.88    2.81   323 1.00
rho_EM             0.22    0.01  0.42   -0.69    0.29    0.81   837 1.00
sigma_year_EM      0.89    0.03  0.56    0.06    0.83    2.15   347 1.01
sigma_EM           1.39    0.01  0.32    0.85    1.36    2.10   568 1.01
mu_MS              0.00    0.00  0.00    0.00    0.00    0.00   975 1.00
rho_MS             0.49    0.01  0.24   -0.08    0.53    0.83   467 1.00
sigma_year_MS      1.04    0.01  0.24    0.69    1.01    1.60  1068 1.00
sigma_MS           0.59    0.00  0.06    0.49    0.58    0.71   674 1.01
mu_p[1]            0.20    0.00  0.02    0.17    0.20    0.24   659 1.00
mu_p[2]            0.75    0.00  0.02    0.71    0.75    0.78   805 1.00
mu_p[3]            0.05    0.00  0.01    0.04    0.05    0.06   821 1.00
sigma_pop_p[1]     0.19    0.01  0.15    0.01    0.16    0.57   426 1.00
sigma_pop_p[2]     0.15    0.01  0.11    0.01    0.12    0.43   401 1.01
R_pop_p[1,1]       1.00     NaN  0.00    1.00    1.00    1.00   NaN  NaN
R_pop_p[1,2]       0.28    0.02  0.57   -0.88    0.41    0.99   704 1.00
R_pop_p[2,1]       0.28    0.02  0.57   -0.88    0.41    0.99   704 1.00
R_pop_p[2,2]       1.00    0.00  0.00    1.00    1.00    1.00   620 1.00
sigma_p[1]         1.49    0.01  0.13    1.25    1.49    1.76   530 1.00
sigma_p[2]         0.80    0.00  0.09    0.64    0.79    0.98   566 1.01
R_p[1,1]           1.00     NaN  0.00    1.00    1.00    1.00   NaN  NaN
R_p[1,2]           0.72    0.00  0.07    0.57    0.73    0.83   870 1.00
R_p[2,1]           0.72    0.00  0.07    0.57    0.73    0.83   870 1.00
```

```
R_p[2,2]           1.00    0.00  0.00      1.00      1.00      1.00    62 1.00
mu_tau_M           0.09    0.00  0.02      0.06      0.08      0.12  2546 1.00
sigma_tau_M        1.31    0.00  0.15      1.05      1.30      1.62  1878 1.00
mu_tau_S           0.12    0.00  0.01      0.10      0.12      0.14  2383 1.00
sigma_tau_S        1.00    0.00  0.06      0.89      1.00      1.14  1985 1.00
lp__          -32317.82    1.58 35.61 -32385.50 -32317.56 -32248.99   509 1.01

Samples were drawn using NUTS(diag_e) at Thu Jul 16 03:13:34 2020.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Model comparison based on LOO. Unhelpful because Pareto ks are too high, but appears to favor Beverton-Holt.

```
LL_LCRchum <- lapply(list(exp = LCRchum_exp, BH = LCRchum_BH, Ricker = LCRchum_Ricker),
                 loo::extract_log_lik, parameter_name = "LL", merge_chains = FALSE)

# Relative ESS of posterior draws of observationwise likelihood
r_eff_LCRchum <- lapply(LL_LCRchum, function(x) relative_eff(exp(x)))

# PSIS-LOO
LOO_LCRchum <- lapply(1:length(LL_LCRchum),
                      function(i) loo(LL_LCRchum[[i]], r_eff = r_eff_LCRchum[[i]]))
names(LOO_LCRchum) <- names(LL_LCRchum)

## Compare all three models
loo_compare(LOO_LCRchum)

       elpd_diff se_diff
exp      0.0       0.0
BH      -3.3       7.3
Ricker -13.5       7.6

## Exponential vs. Ricker
loo_compare(LOO_LCRchum[c("exp","Ricker")])

       elpd_diff se_diff
exp      0.0       0.0
Ricker -13.5       7.6

## Exponential vs. Beverton-Holt
loo_compare(LOO_LCRchum[c("exp","BH")])

    elpd_diff se_diff
exp  0.0       0.0
BH  -3.3       7.3
```

```r
## Beverton-Holt vs. Ricker
loo_compare(LOO_LCRchum[c("BH","Ricker")])
```

```
       elpd_diff se_diff
BH       0.0       0.0
Ricker -10.3       7.1
```

Plot estimated spawner-smolt production curves and parameters for the Beverton-Holt model.

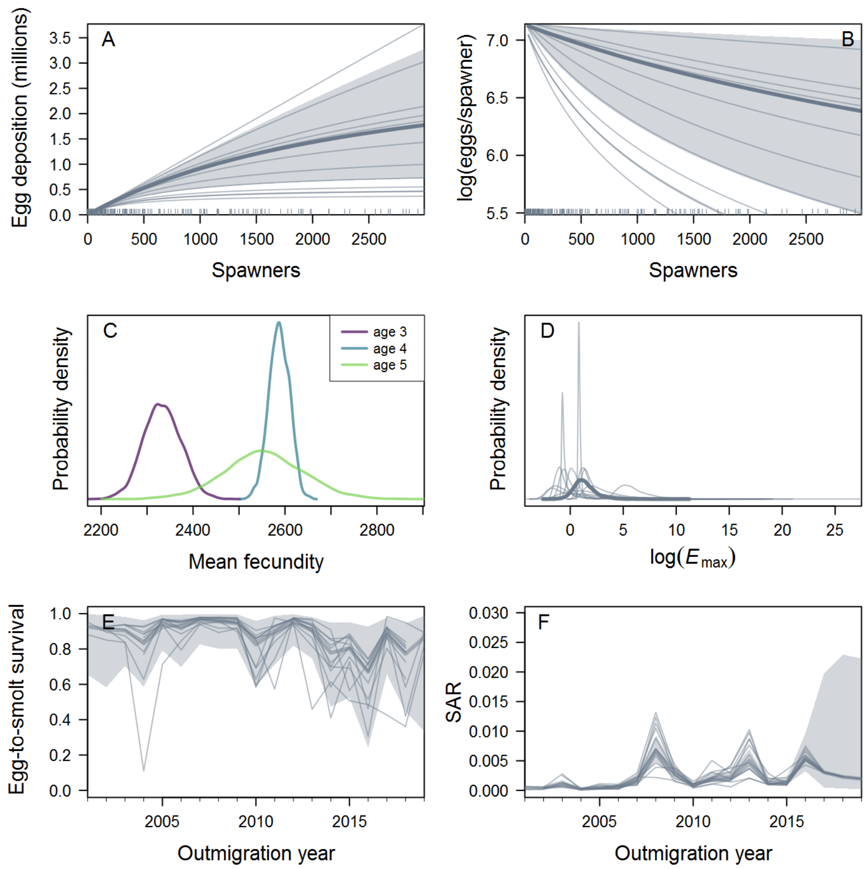

Figure 1: Estimated Beverton-Holt spawner-recruit relationship (A, B) and intrinsic productivity (C) and capacity (D) parameters for the multi-population IPM. Thin lines correspond to each of 12 populations of Lower Columbia chum salmon; thick lines represent hyper-means across populations. In (A, B), each curve is a posterior median and the shaded region represents the 90% credible interval of the hyper-mean curve (uncertainty around the population-specific curves is omitted for clarity).

The Beverton-Holt model is biologically plausible and appears to be supported by LOO, albeit with caveats, so let's tentatively proceed with that model for now. Here are the fits to the spawner data:

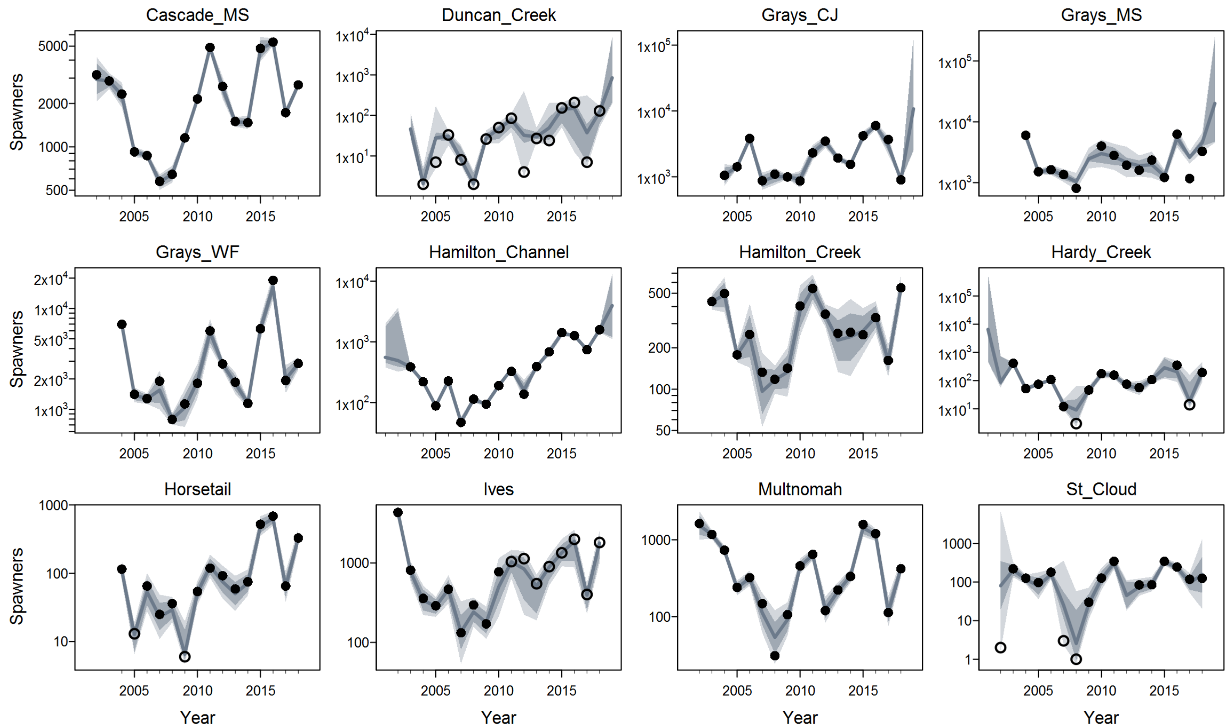

Figure 2: Observed (points) and estimated spawner abundance for Lower Columbia River chum salmon populations. Filled points indicate known observation error SD, while SD for open points is imputed. The posterior median (solid gray line) is from the multi-population IPM. Posterior 90% credible intervals indicate process (dark shading) and observation (light shading) uncertainty.

And here are the fits to the much sparser smolt data:

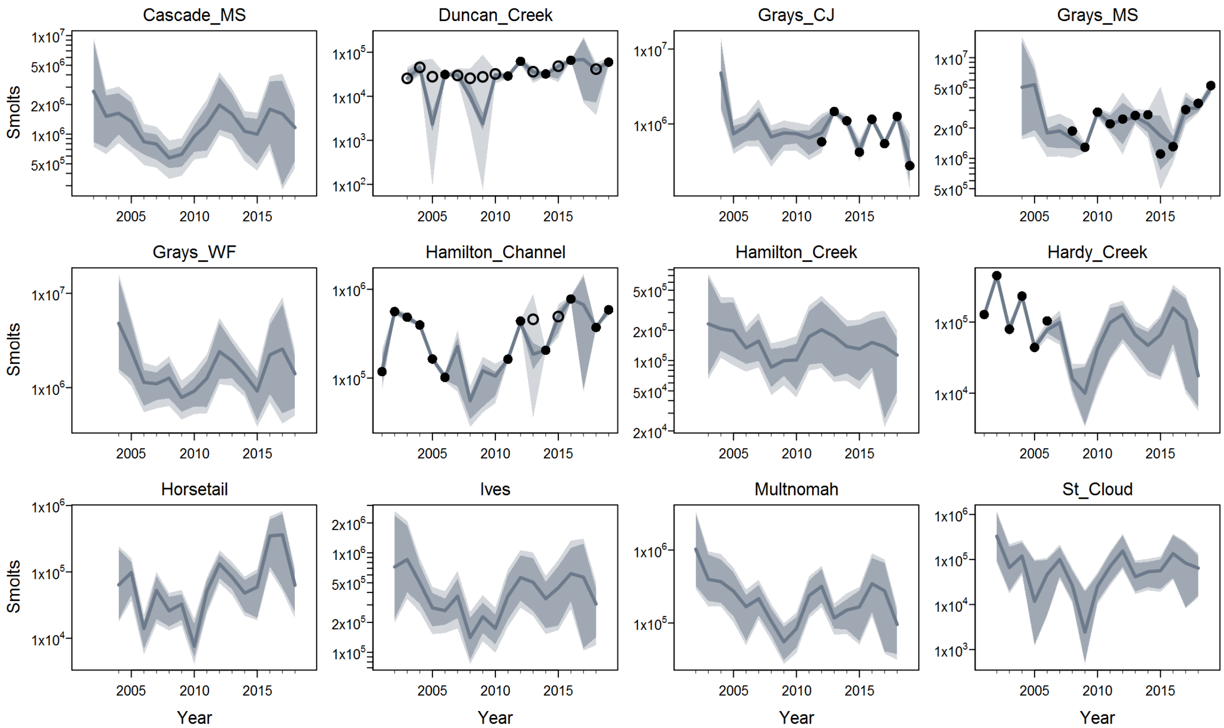

Figure 3: Observed (points) and estimated smolt abundance for Lower Columbia River chum salmon populations. Filled points indicate known observation error SD, while SD for open points is imputed. The posterior median (solid gray line) is from the multi-population IPM. Posterior 90% credible intervals indicate process (dark shading) and observation (light shading) uncertainty.

To understand how the IPM is imputing the observation error SD in cases where it is not reported, let's look at the lognormal hyperdistribution fitted to the known SD values…

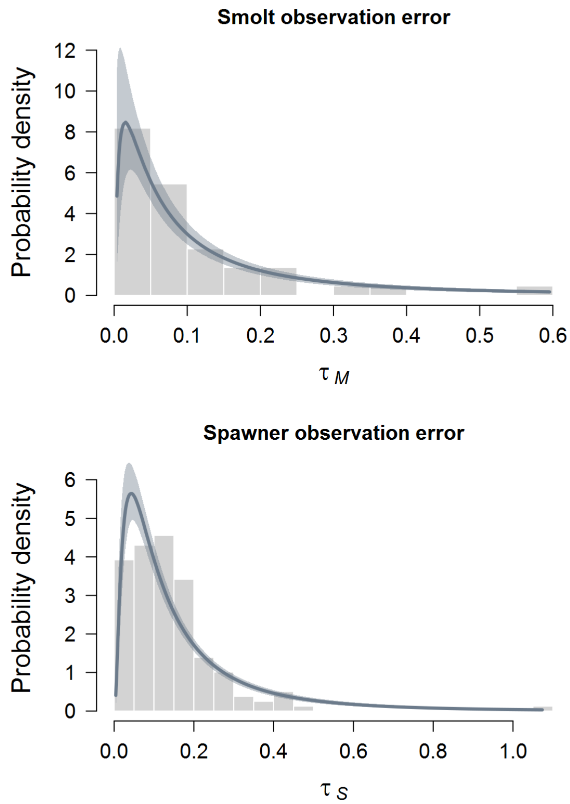

Figure 4: Lognormal hyperdistributions used to impute unknown smolt and spawner observation error SDs in the IPM. The posterior median (line) and 90% credible interval (shading) of the distribution fitted to the known SD values (histogram) are shown for each life stage.

We can also compare the estimated spawner age-frequencies to the sample proportions from the BioData. Age composition varies quite a bit across populations and through time, reflecting fluctuations in cohort strength.

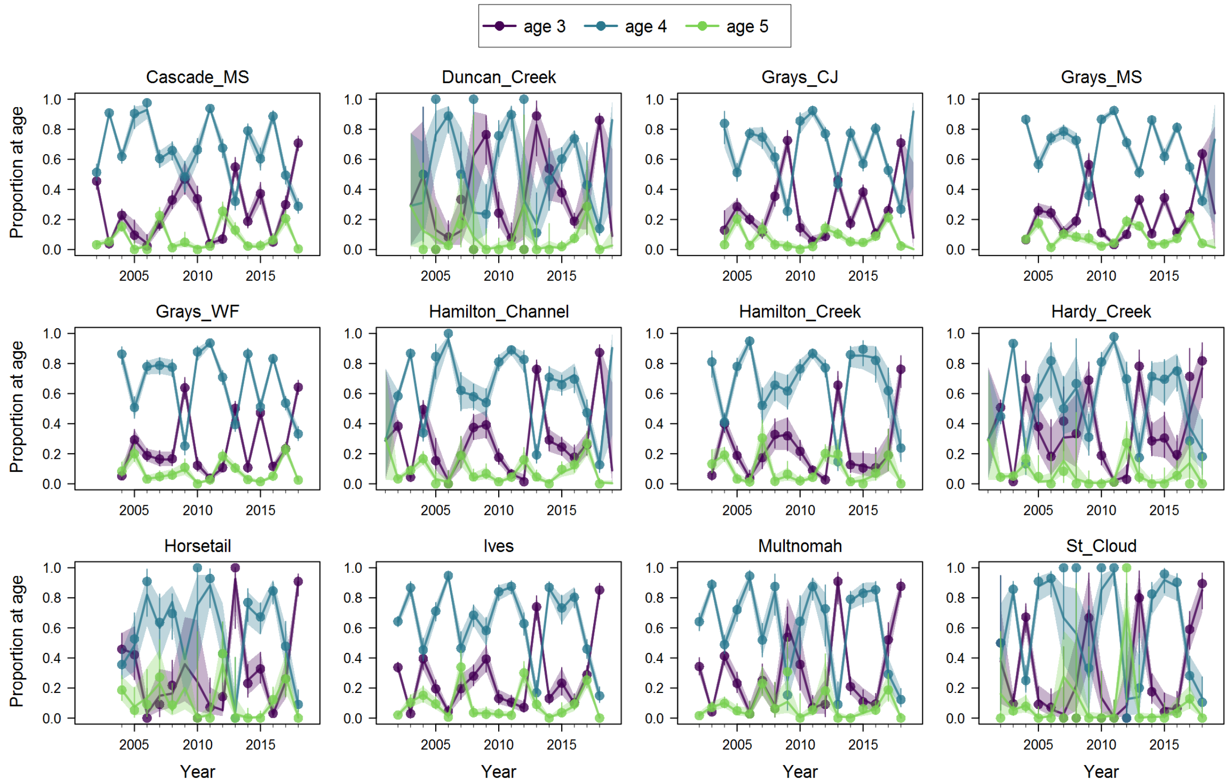

Figure 5: Observed (points) and estimated spawner age composition for Lower Columbia River chum salmon populations. The posterior distribution from the multi-population IPM is summarized by the median (solid line) and 90% credible interval (shading). The error bar around each observed proportion indicates the 90% binomial confidence interval based on sample size.

5 Forecasting

It is straightforward to use the IPM to generate forecasts of population dynamics…

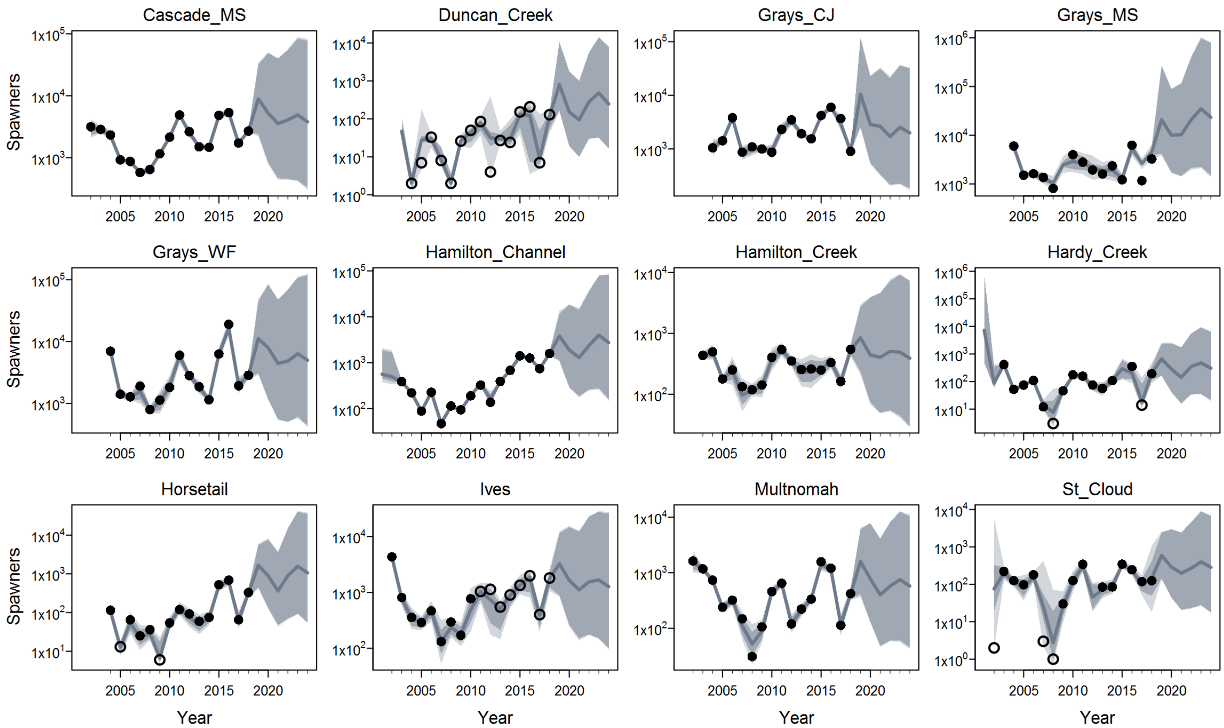

Figure 6: Observed (points) and estimated spawner abundance for Lower Columbia River chum salmon populations, including 5-year forecasts. Filled points indicate known observation error SD, while SD for open points is imputed. The posterior median (solid gray line) is from the multi-population IPM. Posterior 90% credible intervals indicate process (dark shading) and observation (light shading) uncertainty.

Of course we could also look at forecasts of smolts, or any other state variable. Here are the 2020 forecasts of wild spawners for each population…

Population	Estimate
Cascade_MS	5228 (838, 48968)
Duncan_Creek	152 (20, 1779)
Grays_CJ	2851 (548, 22293)
Grays_MS	10003 (1722, 87444)
Grays_WF	7924 (1161, 82872)
Hamilton_Channel	1903 (295, 18195)
Hamilton_Creek	442 (73, 3906)
Hardy_Creek	282 (42, 2410)
Horsetail	888 (132, 7598)
Ives	1645 (267, 14709)
Multnomah	778 (132, 7814)
St_Cloud	291 (45, 2957)
Total	39117 (8820, 280157)