Bearded seal CKMR modeling

Paul Conn

12/2023

Introduction

In this document, we describe CKMR models fitted to bearded seal kinship data.
The theory and equations behind these models is provided elsewhere, so the primary focus here is to ensure repeatability of our analysis. However, we do provide some information on population dynamics equations since this is key knowledge for how initial values were set.

Bearded seal population dynamics model

Underpinning all of our CKMR analyses is an age-structured population dynamics model composed of annual survival probabilities and fecundity values. We will assume a postbreeding census, in which case the number of new recruits each year is given by

$$N_{t,0}^F = N_{t,0}^M = 0.5 \sum_a N_{t-1,a}^F \phi_a f_{t-1,a},$$

where $N_{t,a}^F$ gives the number of age a females (males use the superscritp 'M') alive at time t, ϕ_a is annual survival probability for age a seals, and f_a is female fecundity-at-age (# of pups produced). Note that we assume a 50/50 sex ratio of pups at birth, which is a reasonable assumption given data collected on sex ratios of pups in the field (Fedoseev 2000). Later age classes are propagated forward as a function of age specific survival; i.e., $N_{t,a}^F = N_{t,a}^M = N_{t-1,a-1}\phi_{a-1}$ for $a > 0, t > 0$). In the following CKMR models, we only estimate one abundance paramameter: $N_{0,0}^F$ (where $N_{0,0}^M$ is set equal to $N_{0,0}^F$). The remaining cohort abundances in year 0 are set equal to relevant stable stage proportions from the associated matrix population model (Caswell 2001), and subsequent years abundance are then completely determined.

Priors on life history parameters

Close-kin mark-recapture models only provide limited information on life history parameters. For instance, half-siblings provide information on adult survival (provided that aging has reasonable precision), while parent-offspring pairs provide information on fecundity-at-age. However, bearded seal sample sizes are quite small, so we provided informative priors on survival and fixed fecundity-at-age.

For survival-at-age, we based informative priors on a hierarchical meta-analysis of phocid seal mortality (Trukhanova, Conn, and Boveng 2018). This meta-analysis used a reduced additive Weibull distrubtion (RAW) (Choquet et al. 2011) to model mortality as a function of age for different phocid seal species and populations. The RAW model is characterized by a "bathtub" shape for mortality (i.e, high mortality at young ages, low mortality for young adults, and increasing mortality for the oldest individuals). According to this framework, age-specific annual survival at age a (S_a) is given by

$$S_a = \exp(-(\eta_1 a)^{\eta_2} - (\eta_1 a)^{1/\eta_2} - \eta_3 a),$$

where η_1, η_2, and η_3 are estimated parameters. The values of these parameters from hierarchical analysis (Trukhanova, Conn, and Boveng 2018) were $\eta_1 = 0.055$, $\eta_2 = 2.80$, and $\eta_3 = 0.076$ (Conn et al. 2020).

We set fecundity-at-age values equal to schedules reported by (Conn and Trukhanova 2022), who fitted generalized additive models to data from specimens collected in the Bering and Chukchi Seas. These data represented the proportion of age a females who had given birth or were pregnant in the spring. Sample collections were derived from Native Alaskan subsistence harvests which are monitored by the Alaska Department of Fish and Game, as well as records reported from Russia in the 1980s (Fedoseev 2000).

Although not needed explicitly for population modeling, CKMR paternal kinship probabilities (including half-siblings that are paternally related) rely on relative paternal reproductive output as a function of age. We based these calculations in part on male maturity schedules reported by (Conn and Trukhanova 2022), which were derived from collections from the Bering and Okhotsk seas (Tikhomirov 1966). Survival-, fecundity, and maturity-at-age (m_a) are plotted below.

```
Maturity = read.csv("c:/users/paul.conn/git/ckmr/Bearded_ADFG/Maturity.csv")
Survival = read.csv("c:/users/paul.conn/git/ckmr/Bearded_ADFG/Survival_ests.csv")
Reprod = read.csv("c:/users/paul.conn/git/ckmr/Bearded_ADFG/Reproduction_table.csv")

Male_mat <- rep(1,40)
Fem_fec <- rep(0.938,40)
Male_mat[1:10]=c(0,Maturity$Bearded.male)
Fem_fec[1:10]=c(0,Reprod$bearded)
Plot_df = data.frame("Vital_rate"=rep(c("Fem_fec","Male_mat","Survival"),each=40),
                     "Value"=c(Fem_fec,Male_mat,Survival$bearded),
                     "Age"= rep(c(0:39),3))
library(ggplot2)
vital_plot = ggplot(Plot_df)+geom_line(linewidth=0.8,aes(x=Age,y=Value,colour=Vital_rate,linetype=Vita
  scale_color_manual(values=c("purple","blue","darkorange"),labels = c("Female fecundity", "Male matu
vital_plot
```
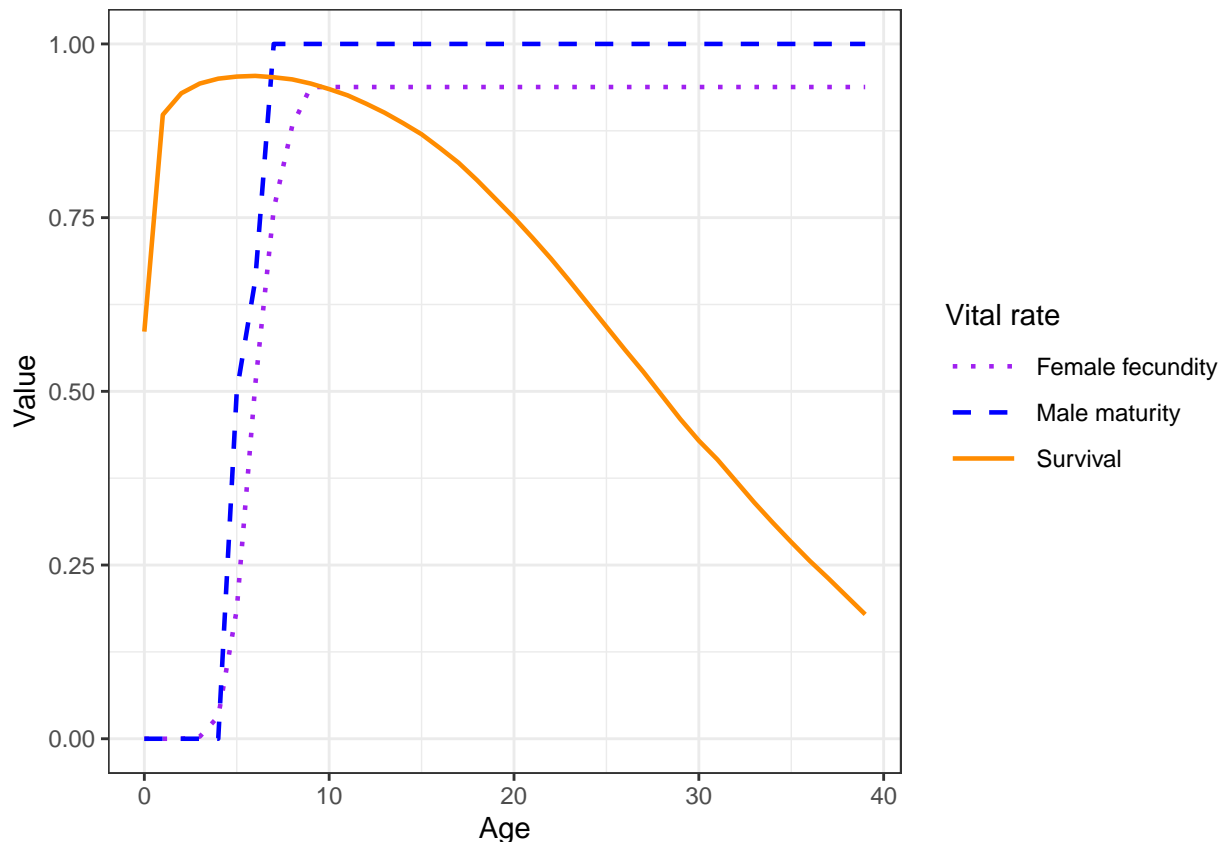
```
png('vital_rate_prior_plot.png')
  vital_plot
dev.off()
```

```
## pdf
##   2
```

One interesting thing to note about using fixed values of fecundity- and survival-at-age is that the corresponding Leslie matrix implies a very specific population trend, and owing to measurement error in estimation of both sets of vital rates it is possible for the implied finite rate of poulation increase (λ) to indicate increasing of decreasing populations. Let's see what λ value these vital rates would imply, should an equilibrium age structure be reached:

```
# set up leslie matrices - via an array (4 matrices, one for each species)
A = matrix(0,40,40)
for(iage in 1:39){
  A[iage+1,iage]=Survival[iage,"bearded"]   #assume post-breeding census
}

#reproduction; nb: adults have to survive to next spring to reproduce
# nb: Leslie matrices are "female only" and assume a 50/50 sex ratio at birth
A[1,]=0.5*Fem_fec*Survival$bearded

eigen(A)$values[1]
```

```
## [1] 1+0i
```

So, it would appear that this combination of fecundity-at-age and survival-at-age is expected to result in about a 4% annual increase in abundance. This is clearly undesirable, because we do not want to presuppose such an increase before we start analyzing CKMR data. There are potentially several fixes to this. First, since survival-at-age is presumably much more uncertain than fecundity-at-age (the former having been produced from a meta-analysis rather than an actual field study), we might consider manipulating survival-at-age values until $\lambda = 1.0$. This was the approach taken by (Conn et al. 2020) when analyzing simulated data that were patterned after bearded seal life history parameters. Alternatively, we could let a CKMR model attempt to estimate updated RAW parameter values, subject to a constraint enforcing a specific value of λ (e.g., 1.0). Given that our kinship data seem like they're too sparse to permit robust inference about population trend, this is the approach we will start off with in our first CKMR analyses with bearded seal data.

Imposing a PLOD threshold

So far, we have been assuming that we have been able to fully discriminate HSP/GGP pairs from more distant kin pairs (e.g., half-aunt-niece, etc.). In truth, it is difficult to discriminate between the two at lower PLOD scores, and it is often worth imposing a lower threshold for PLOD scores to eliminate possible lower order kin. In this case, we can try to account for the HSP/GGPs that are under our assigned lower threshold by doing some creative modeling. The following is an attempt to show our conundrum graphically; here, the red line depicts a hypothetical expected frequency of PLOD scores among unrelated pairs (with a bump centered at zero), and the blue line depicts the same for HSGGPs (here centered at 100). The issue is at scores of e.g. 50-70. These matches could conceivably be of either type.

```
X = c(-50:150)
Y_no = 500*dnorm(X,0,20)
Y_sib = 100*dnorm(X,100,20)
Plot_df = data.frame("PLOD_value"=rep(X,2),"Exp_count"=c(Y_no,Y_sib),Type=c(rep("Distant",length(X)),
library(ggplot2)
ggplot(Plot_df)+geom_line(aes(x=PLOD_value,y=Exp_count,group=Type,color=Type))
```
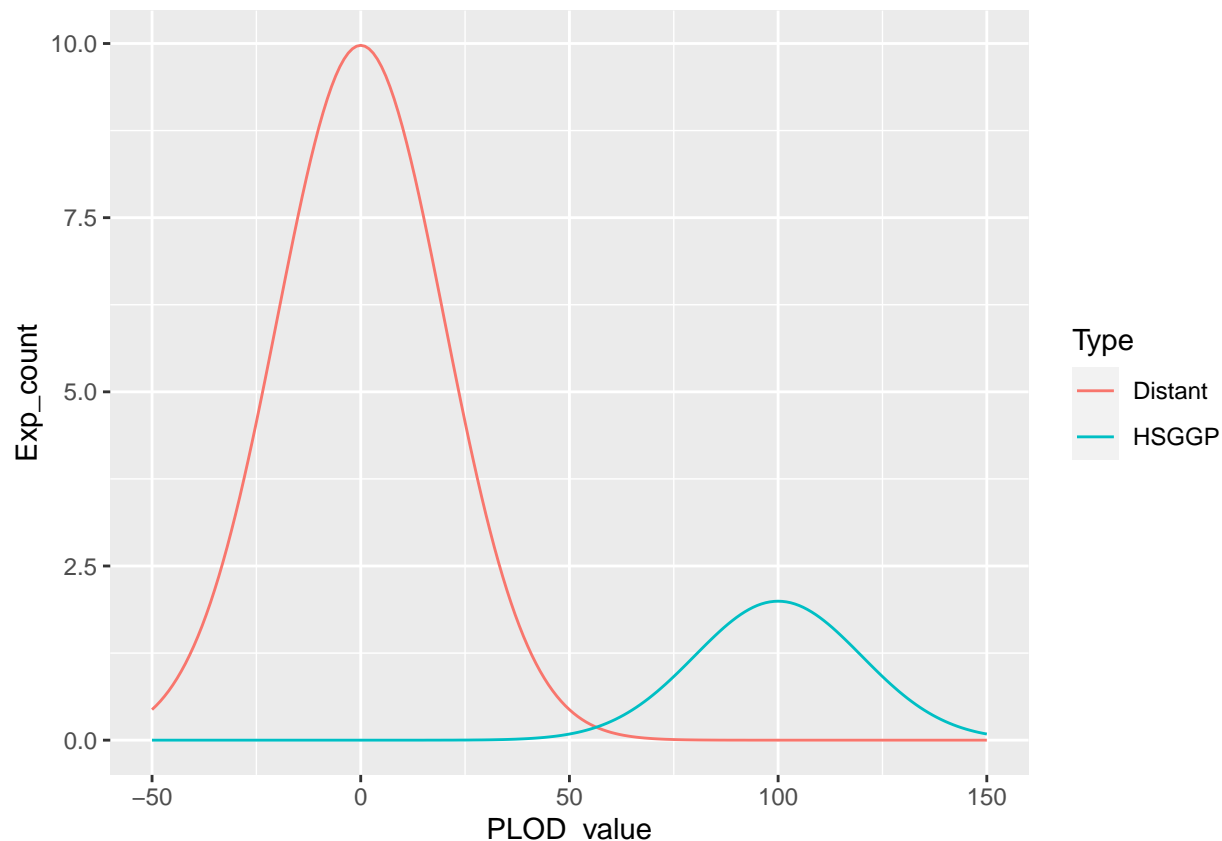
One thing we might do then is to impose a threshold (let's say PLOD_value=70) that essentially makes the probability of a false positive HSGGP negligible. If we knew the parameters of the blue curve and are willing to assume normality, we could then calculate the probability of detecting and including an HSGGP in our modeling procedure as e.g.

$$d = \int_{x=70}^{\infty} f(x; \mu, \sigma^2) dx$$

where $f(x; \mu, \sigma^2)$ is a Gaussian probability density function with mean μ and variance σ^2. To include this quantity in estimation, we can simply replace p_{ij2} with $p_{ij2}d$ everywhere it occurs in the CKMR pseudo-likelihood.

Let's take a look at some possible PLOD thresholds with our bearded seal dataset. We'll start by plotting a histogram of potential HSP scores

```
load('c:/users/paul.conn/git/ckmr/bearded_adfg/HSP_plod_info.RData')

hist(HSP_PLOD_info$HSP_PLODs,breaks=20,xlab="PLOD score",main="Potential HSP plod scores")
abline(v=HSP_PLOD_info$HSPmean,col="red")
```

Potential HSP plod scores
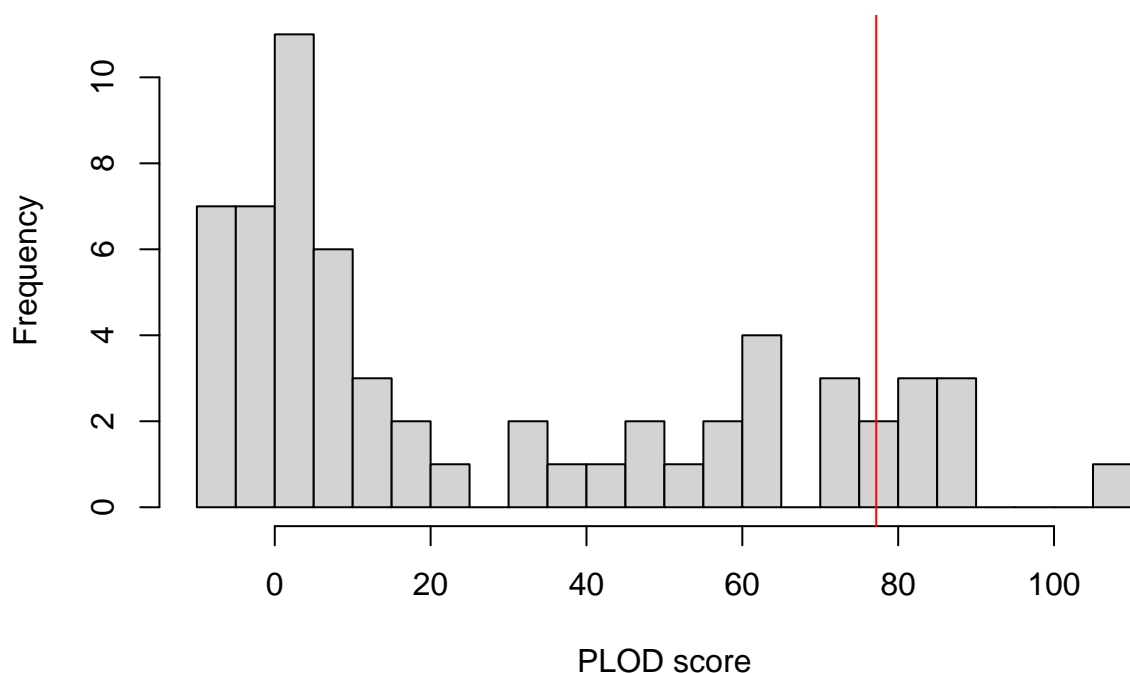

```r
png("HSP_plod_bearded.png")
  hist(HSP_PLOD_info$HSP_PLODs,breaks=20,xlab="PLOD score",main="Potential HSP plod scores")
  abline(v=HSP_PLOD_info$HSPmean,col="red")
dev.off()
```

```
## pdf
##   2
```

Here, the red line shows the theoretical place (based on Hardy-Weinberg) where the peak of PLOD scores should be for HSPs (and GGPs!); the peak closer to zero is for related kin, but lesser so than HSPs (e.g., half-aunt/niece, etc.). In order to set PLOD thresholds we need to know something about the variance of the HSP peak. Because the left hand side is potentially contaminated by weak kin pairs, the only "safe" thing to do is to use the right hand side of the distribution to estimate variance. Let's fit a half-normal distribution to these scores using maximum likelihood. We'll use this fitted model to predict the probability of a HSP occurring with PLOD scores below certain thresholds (specifically, 30, 40, and 50).

```r
Upper_sample = HSP_PLOD_info$HSP_PLODs[which(HSP_PLOD_info$HSP_PLODs>HSP_PLOD_info$HSPmean)]
HalfN_sample = Upper_sample-HSP_PLOD_info$HSPmean

ML_fun <- function(sd_log,Data){
  sd = exp(sd_log)
  -sum(2*dnorm(Data,0,sd,log=T))
}

Estim = nlminb(log(10),ML_fun,Data=HalfN_sample)
sd_est = exp(Estim$par)
```

```
pnorm(30,HSP_PLOD_info$HSPmean,sd_est)
```

```
## [1] 0.00019
```
```
pnorm(40,HSP_PLOD_info$HSPmean,sd_est)
```

```
## [1] 0.0026
```
```
pnorm(50,HSP_PLOD_info$HSPmean,sd_est)
```

```
## [1] 0.02
```

This little experiment is something that should be conducted in all CKMR experiments using half-siblings. However, it also demonstrates the very real complications that can happen with small datasets! In particular, there is a very subjective "gap" around 25 that is tempting to take as a cutoff; however, our analysis suggests that the PLOD cutoff could potentially be much higher (e.g., 40-50). There are some other details that we can look at to justify a decision. In particular, half-aunts/nieces, etc. will typically have larger age differences with a mode near one generation time. In the bearded seal case, potential HSPs with PLOD scores above 25 tend to have lower age differences than matches with lower PLOD scores; however, this isn't definitive. Our view is that we were probably unlucky (in the sense that we had low sample size), and the observed variance to the right hand side of the red line is lower than we would normally observe in a sample of this size. However, it's hard to say for sure, and we'll be somewhat conservative and use a cutoff of 40 as a "base case" in subsequent model runs. Let's now limit our HSP matches to those that have a PLOD threshold of 40.

```
load("CKMR_sample_data.RData")
p_plod <- 1 - pnorm(40, HSP_PLOD_info$HSPmean, sd_est) # to account for false negative HSP/GGPs in CKMR
matches_PHS <- CKMR_certain_age$matches_PHS[which(CKMR_certain_age$matches_PHS$PLOD > 40), ]
matches_MHS <- CKMR_certain_age$matches_MHS[which(CKMR_certain_age$matches_MHS$PLOD > 40), ]
```

This reduces the number of HSPs to 22 from the original 25.

CKMR modeling

Our CKMR models all assume ages are certain and inference is based on maximum marginal pseudo-likelihood inference, with an observation model based on a product Bernoulli likelihood reflecting a large number of pairwise kinship comparisons (Bravington et al. 2016). Specifically, we base inference on the joint pseudo-likelihood

$$L = L_{pop}L_{hsp/ggp}f(\boldsymbol{\eta})\Lambda_\lambda,$$

where L_{pop} is a product Bernoulli likelihood for parent-offspring pairs, $L_{hsp/ggp}$ is a product Bernoulli likelihood for half siblings and grandparent-grandchild pairs (modeled as a mixture), $f(\eta)$ are penalties on RAW survival parameters if they deviate from their prior mean, and Λ_λ is a penalty for population trend that is > 0 when $\lambda \neq \lambda_0$ (where λ_0 is pre-specified by the analyst). These likelihoods are described elsewhere, so we will now turn to model fitting.

All CKMR models were programmed in Template Model Builder (TMB; (Kristensen et al. 2015)). Conditioning on observed kinship observations, the joint negative log likelihood (i.e., $-\log(L)$) was minimized as a function of $N_{0,0}^F$ and $\boldsymbol{\eta}$ parameters using the "nlminb" function in R (R Development Core Team 2017). Some computational efficiency is gained by noting that many of the pairwise kinship probabilities are the same (e.g., for individuals of the same sex and with the same birth years and year of death) and grouping pairs this way. Modeling these sufficient statistics prevents us from having to do $n^2 - n$ separate comparison computations every time the likelihood is evaluated. The next block of code is concerned with computing the number of comparisons and matches for POPs and HS-GGPs. Note that we go back one generation time to start the model to allow us to model relative reproductive output of parents of half-siblings that are encountered near the beginning of the study.

Now, let's actually use these sufficient statistics to fit our CKMR model:

The time to fit the combined GGP-HSP model was (up to 31.0184528827667). The abundance estimate is 2.32×10^5. The expected number of grandparent-grandchild pairs given our data and model fit is 4.81.

Let's take a look at the relative probabilities of individual seals being HSPs vs GGPs.

```
HSPs = which(Data$n_match_HSGGP_sibidibjmij==1,arr.ind=TRUE)
HSPs = data.frame(HSPs)
colnames(HSPs) = c("Older_sex","birth_i","death_i","birth_j","mito")
HSPs$Rel_prob_HSP = HSPs$prob_GGP = HSPs$prob_HSP = 0
for(i in 1:nrow(HSPs)){
  GGP_prob = Report$GGP_table[HSPs[i,1],HSPs[i,2],HSPs[i,3],HSPs[i,4],HSPs[i,5]]
  HSPs$prob_GGP[i] = GGP_prob
  if(HSPs$mito[i]==1){   #use PHSP table
    HSPs$Rel_prob_HSP[i] = Report$PHS_table[HSPs[i,2],HSPs[i,4]]/(Report$PHS_table[HSPs[i,2],HSPs[i,4]
    HSPs$prob_HSP[i] = Report$PHS_table[HSPs[i,2],HSPs[i,4]]
  }
  else{
    HSPs$prob_HSP[i] = Report$MHS_table[HSPs[i,2],HSPs[i,4]]
    HSPs$Rel_prob_HSP[i] = Report$MHS_table[HSPs[i,2],HSPs[i,4]]/(Report$MHS_table[HSPs[i,2],HSPs[i,4]
  }
}
print(HSPs)
```

```
##     Older_sex birth_i death_i birth_j mito prob_HSP prob_GGP Rel_prob_HSP
## 1           2      35       8      47    1  3.0e-06  2.3e-06         0.57
## 2           1      41      13      47    1  9.1e-06  0.0e+00         1.00
## 3           1      46      10      50    1  1.3e-05  0.0e+00         1.00
## 4           1      46       6      51    1  1.1e-05  0.0e+00         1.00
## 5           1      40      14      51    1  3.7e-06  4.3e-07         0.90
## 6           1      41       7      52    1  3.7e-06  4.3e-07         0.90
## 7           2      52      13      52    1  2.1e-05  0.0e+00         1.00
## 8           2      48      10      53    1  1.1e-05  0.0e+00         1.00
## 9           2      44       4      54    1  4.5e-06  0.0e+00         1.00
## 10          1      48       8      54    1  9.1e-06  0.0e+00         1.00
## 11          1      52      12      54    1  1.7e-05  0.0e+00         1.00
## 12          1      50      14      56    1  9.1e-06  0.0e+00         1.00
## 13          1      55      15      56    1  1.9e-05  0.0e+00         1.00
## 14          2      57      17      57    1  2.1e-05  0.0e+00         1.00
## 15          1      51      11      59    1  6.5e-06  0.0e+00         1.00
## 16          2      54      15      60    1  9.1e-06  0.0e+00         1.00
## 17          2      49       9      62    1  2.5e-06  0.0e+00         1.00
## 18          1      48      15      49    2  2.0e-05  0.0e+00         1.00
## 19          1      39       6      50    2  3.9e-06  1.0e-06         0.80
## 20          2      46       6      51    2  1.1e-05  0.0e+00         1.00
## 21          2      50      10      53    2  1.6e-05  0.0e+00         1.00
## 22          2      51      12      53    2  1.8e-05  0.0e+00         1.00
## 23          2      54      14      56    2  1.8e-05  0.0e+00         1.00
## 24          2      54      15      57    2  1.6e-05  0.0e+00         1.00
## 25          1      57      17      59    2  1.8e-05  0.0e+00         1.00
```

It looks like there are 4 seals that have potential to be GGPs, though the relative probabilities still favor these being HSPs (the first row has the highest probability of being a GGP, at 0.43).

Out of curiosity, we might examine what would happen if we fit a CKMR model which didn't allow for the possibility of GGPs. Let's do this.

```r
library(TMB)
TmbFile_noGGP <- "c:/users/paul.conn/git/CKMR/bearded_ADFG/bearded_CKMR/src/certain_ages_plod.cpp"
compile(TmbFile_noGGP)
```

using C++ compiler: 'G__~1.EXE (GCC) 12.2.0'

[1] 0

```r
TmbExec_noGGP <- "c:/users/paul.conn/git/CKMR/bearded_ADFG/bearded_CKMR/src/certain_ages_plod"
dyn.load(dynlib(TmbExec_noGGP))

Data_noGGP <- list(
  "n_yrs" = n_yrs, "n_yrs_data" = n_yrs_data, "n_seals" = n_seals, "n_ages" = n_ages,
  "Male_mat" = Male_mat, "Fem_fec" = Fem_fec, "A" = A,
  "n_match_MPO_bidibj" = n_match_MPO_bidibj, "n_comp_MPO_bidibj" = n_comp_MPO_bidibj,
  "n_match_PPO_bidibj" = n_match_PPO_bidibj, "n_comp_PPO_bidibj" = n_comp_PPO_bidibj,
  mu_log_eta1 = log(0.055), mu_log_eta2 = log(2.8), mu_log_eta3 = log(0.076),
  sd_log_eta1 = 0.07 * abs(log(0.055)), sd_log_eta2 = 0.2 * abs(log(2.8)), sd_log_eta3 = abs(0.08 * log
  lambda_expect = 1.0, min_repro_fem = 3, min_repro_male = 5, p_plod = p_plod
) # SD log multipliers set to

Data_noGGP$n_match_MHS_bibj <- apply(Data$n_match_HSGGP_sibidibjmij[, , , , 2], c(2, 4), "sum")
Data_noGGP$n_match_PHS_bibj <- apply(Data$n_match_HSGGP_sibidibjmij[, , , , 1], c(2, 4), "sum")
Data_noGGP$n_comp_HS_bibj <- apply(Data$n_comp_HSGGP, c(2, 4), "sum")

Obj <- MakeADFun(data = Data_noGGP, parameters = Params, random = Random, map = Map, hessian = FALSE, DI

Obj$fn(Obj$par)
```

[1] 2175170

```r
Report <- Obj$report()


# Minimize negative log likelihood and time it
Start_time <- Sys.time()
Opt <- nlminb(start = Params, objective = Obj$fn, gradient = Obj$gr)
```

```
## outer mgc:  2.1e+07
## outer mgc:  8502685
## outer mgc:  1709685
## outer mgc:  1293724
## outer mgc:  846512
## outer mgc:  38547
## outer mgc:  17560
## outer mgc:  218
## outer mgc:  264
## outer mgc:  1315
## outer mgc:  355
## outer mgc:  164
## outer mgc:  549
## outer mgc:  560
## outer mgc:  505
## outer mgc:  203
## outer mgc:  635
## outer mgc:  276
```

```
## outer mgc:   363
## outer mgc:   979
## outer mgc:   10
## outer mgc:   79
## outer mgc:   170
## outer mgc:   456
## outer mgc:   552
## outer mgc:   282
## outer mgc:   803
## outer mgc:   6.5
## outer mgc:   153
## outer mgc:   179
## outer mgc:   210
## outer mgc:   570
## outer mgc:   124
## outer mgc:   725
## outer mgc:   129
## outer mgc:   110
## outer mgc:   117
## outer mgc:   191
## outer mgc:   211
## outer mgc:   359
## outer mgc:   113
## outer mgc:   100
## outer mgc:   59
## outer mgc:   548
## outer mgc:   78
## outer mgc:   195
## outer mgc:   298
## outer mgc:   284
## outer mgc:   68
## outer mgc:   386
## outer mgc:   120
## outer mgc:   126
## outer mgc:   208
## outer mgc:   225
## outer mgc:   76
## outer mgc:   349
## outer mgc:   65
## outer mgc:   80
## outer mgc:   111
## outer mgc:   123
## outer mgc:   175
## outer mgc:   70
## outer mgc:   237
## outer mgc:   51
## outer mgc:   85
## outer mgc:   103
## outer mgc:   134
## outer mgc:   30
## outer mgc:   101
## outer mgc:   101
## outer mgc:   20
## outer mgc:   34
```

```
## outer mgc:   87
## outer mgc:   55
## outer mgc:   45
## outer mgc:   25
## outer mgc:   39
## outer mgc:   29
## outer mgc:   28
## outer mgc:   2.3
## outer mgc:   4
## outer mgc:   10
## outer mgc:   5.5
## outer mgc:   1.6
## outer mgc:   0.39
## outer mgc:   0.038
```

```r
End_time <- Sys.time()

Report <- Report_NoGGP <- Obj$report()
SD_report_NoGGP <- sdreport(Obj)
```

```
## outer mgc:   0.038
## outer mgc:   0.022
## outer mgc:   0.094
## outer mgc:   50364
## outer mgc:   50120
## outer mgc:   180
## outer mgc:   182
## outer mgc:   48230
## outer mgc:   48192
## outer mgc:   884831
```

```r
N_est_NoGGP <- SD_report_NoGGP$value[which(names(SD_report_NoGGP$value) == "N")]
```

In this case, abundance is estimated as 1.85×10^5 with a CV of 0.21

Male heterogeneity

Looking at the data for the PLOD cutoff of 40, there are a total of 15 HSP/GGPs that don't share mtDNA, and 7 that do share mtDNA. If we look at these records, it looks like there are 4 that have a chance of being GGPs, but in each case they are more likely to be HSPs. If we assume that they are all HSPs, and also assume that we are equally likely to detect maternal and paternal HSPs, the probability of observing so few paternal HSPs is 0.07. This is could have happened by random chance, or it could be a problem with our model. Conditional on the observed data and parameter estimates of our CKMR model, the expected number of MHSPs and PHSPs are 0 and 0, respectively. A χ^2 test with expected and observed numbers of HSPs by sex generates a p-value of 0.17 ($\chi^2 = 1.9, df = 1$).

So why do male HSPs outnumber female HSPs by such a large margin? One explanation is heterogeneity in male reproductive success (e.g., if older or higher quality males are able to breed with more females than younger or lower quality males). For instance, male bearded seals are known to maintain underwater territories during the breeding season, and it may be the case that there may be some competition for mates. Also, male sexual maturity was based on physical characteristics (e.g., presence of semen in testicles of dead specimens), and it may be the case that there is a delay before physically mature males become successful breeders.

In order to account for this possibility, we fit an alternative model where we assumed the number of male breeders is an unknown fraction (π) of the total number of reproductively mature males. Specifically, we set $N_{t,a}^M = 0.5 N_{t,a} \pi$ everywhere that male abundance appears in previous calculations.

After running this model, we have $\hat{\pi} = 0.34$, suggesting that only a relatively small fraction of reproductively mature males are successfully producing offspring each year. As expected, this leads to an increase in estimated abundance, which is now at $\hat{N} = 4.09 \times 10^5$. This model is supported by AIC, with $\delta\text{AIC} = $ -4.23 over the model without heterogeneity. Predictions of the number of maternal and paternal HSPs are now 5.15 and 15.14, much closer to the observed values of 7 and 15.

Although we will keep this heterogeneity model as our "preferred" model, we wanted to see how much of a shift in the male maturity curve would be needed to produce a similar result. After experimenting with a variety the male maturity curve by a number could accomplish the same thing, we found that male sexual maturity would need to be delayed by 8 years to result in $\hat{\pi} = 1.0$:

Alternative trend scenarios

Another possible source of structural uncertainty is with our assumption that abundance is constant over time. Let's look and see what happens when we investigate four alternative trend scenarios, including a constant 2% annual rate of increase/increase and a 4% rate of increase/decrease over time. In practice, it would be impossible to sustain a constant increase or decrease over a long period, but it will be instructive to see what alternative trends do to the overall scale of our estimate. Let's do this with the heterogeneity model.

This exercise illustrates several phenomenon. First, abundance estimates intersect around 2004, but are quite different by the end of the time series. This is one phenomenon with CKMR estimation: precision and accuracy of estimates tend to be better towards to beginning of time series (in the "meat" of observed birth dates). Second, the log pseudo-likelihood values are fairly similar for the five trend (λ) values, with $L = $ -302.07, -300.36, -299.98, -301.45, and -305.65for $\lambda = 0.96, 0.98, 1.0, 1.02,$ and1.04, respectively. There is slightly more evidence for stable, slightly decreasing, or slightly increasing populations than there is for the large decline or large increase scenarios - but we would certainly want a lot more kin pairs for reliable trend estimation.

Alternative PLOD thresholds

Let's take a look at results if we were to use different PLOD thresholds for HSP/GGPs. In these cases we'll still account for false negative probabilities, which will increase if we lower the PLOD threshold, or decrease if we increase the PLOD threshold. However, there is some senstivity to this choice, and it's something to be aware of for CKMR studies with a low number of HSPs (we would expect this not to be much of an issue with higher sample sizes). In particular, we'll look at results for HSP thresholds of 30 and 50 (our base model uses 40).

Looking at estimates from different sensitivity runs, we see that abundance estimates are fairly sensitive to the PLOD cutoff value, with a 30 cutoff resulting in $\hat{N} = 3.65 \times 10^5$; a 40 cutoff resulting in $\hat{N} = 4.09 \times 10^5$, and a 50 cutoff resulting in $\hat{N} = 5.35 \times 10^5$. As more harvest data are collected, presumably resulting in an increased number of kin pairs, we hope that this source of structural uncertainty will diminish.

Let's take a look at a summary of estimated abundance from the combination of sensitivity runs we have attempted so far. For the increasing and decreasing population scenarios, we'll use average abundance from 1990-2020.

```
N_df <- data.frame(matrix(0, 8, 5))
colnames(N_df) <- c("lambda", "PLOD_cutoff", "male_het", "N_hat", "CV")
N_df$lambda <- c(1, 1, 0.96, 0.98, 1.02, 1.04, 1, 1)
N_df$PLOD_cutoff <- c(40, 40, 40, 40, 40, 40, 30, 50)
N_df$male_het <- c("no", "yes", "yes", "yes", "yes", "yes", "yes", "yes")
N_df$N_hat <- c(N_est_TMB[1], Report_het$N[1], mean(Report_lambda_96$N[32:62]), mean(Report_lambda_98$N
N_df$CV <- c(SD_report$sd[1] / N_est_TMB[1], SD_N_het[1] / Report_het$N[1], mean(SD_N_lambda_96[32:62] /
print(N_df)

##   lambda PLOD_cutoff male_het  N_hat   CV
## 1   1.00          40       no 231814 0.21
```

```
## 2   1.00          40        yes 408652 0.35
## 3   0.96          40        yes 437824 0.35
## 4   0.98          40        yes 412120 0.35
## 5   1.02          40        yes 427607 0.35
## 6   1.04          40        yes 469770 0.36
## 7   1.00          30        yes 365233 0.33
## 8   1.00          50        yes 535075 0.41
```

Survival

Let's take a look at the updated survival curve (prior vs. "posterior") from our base model.

```
Plot_df <- data.frame(
  "Type" = rep(c("Prior", "CKMR"), each = 40),
  "Value" = c(Survival$bearded, Report_het$S_a),
  "Age" = rep(c(0:39), 2)
)
library(ggplot2)
ggplot(Plot_df) +
  geom_line(aes(x = Age, y = Value, colour = Type, linetype = Type), size = 1.1) +
  theme(text = element_text(size = 16))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```
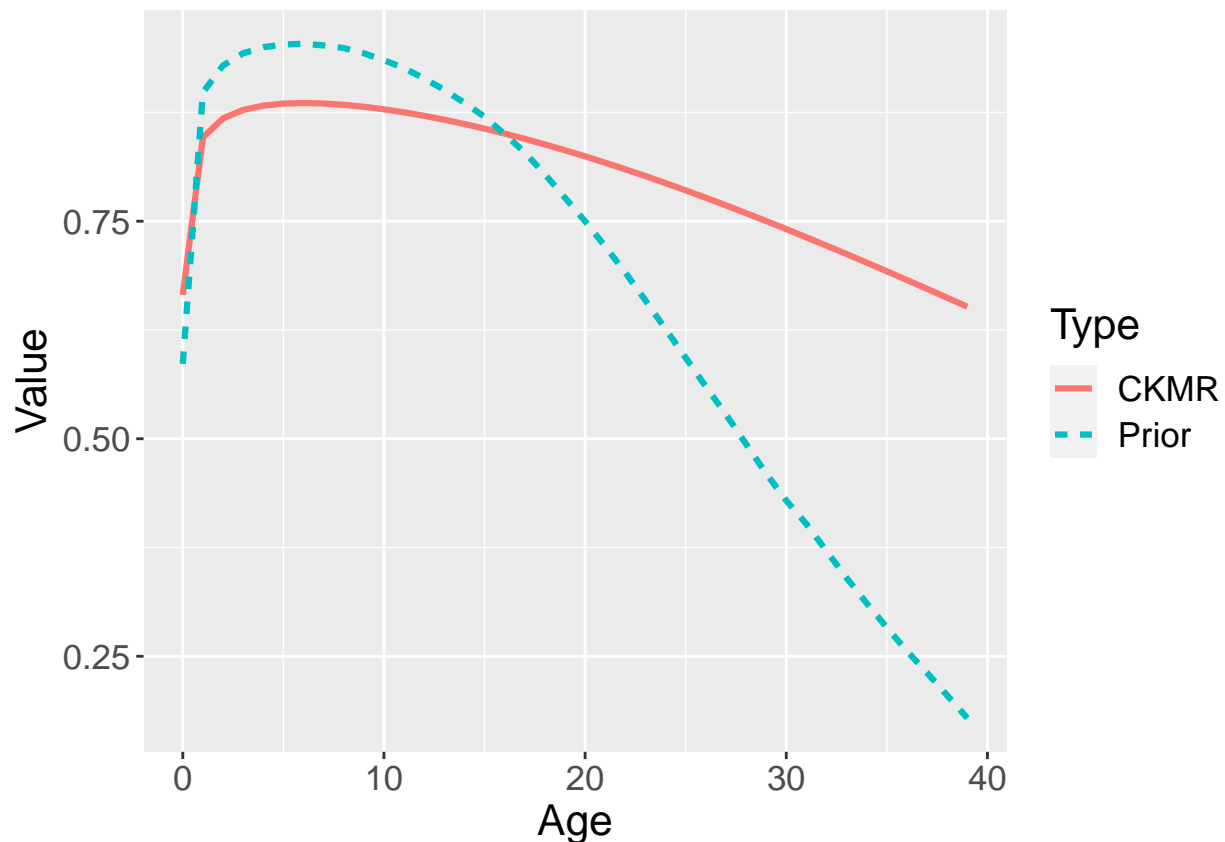
```
png("bearded_surv_prior_posterior.png")
ggplot(Plot_df) +
  geom_line(aes(x = Age, y = Value, colour = Type, linetype = Type), size = 1.1) +
  theme(text = element_text(size = 16))
dev.off()
```

```
## pdf
##   2
```

Aging error

We did not elect tackle aging error, instead electing to assume that ages were known with certainty. Including uncertainty in ages is certainly possible in CKMR estimation (Bravington et al. 2016), and would serve to increase uncertainty in resulting estimates. However, it is difficult to summarize uncertainty in ages, partly because of the way teeth were analyzed. The ages we used were primarily from tooth cementum annuli, and most of these were read by a single reader who assigned a "most likely" age, as well as a range of ages that were plausible. However, these ranges were not always accurate, as we had several kin pairs that indicated aging error magnitudes greater than assigned by the reader. Ideally, aging error could be estimated using a separate experiment where multiple tag readers assess the same tooth (Richards et al. 1992), and then incorporated directly into CKMR estimation.

Comparison with aerial survey estimates

NOAA's Alaska Fisheries Science Center, together with Russian partners, conducted spring aerial surveys over the Bering Sea in 2012 and 2013, and over the Chukchi Sea in 2016. Data from these surveys have been analyzed using spatio-temporal statistical models, which produced abundance estimates. Although as-yet unpublished, bearded seal estimates were 147,000 for the Chukchi Sea; 185,000 for the Russian Bering in 2012; 144,000 for the Russian Bering in 2013; 271,000 for the U.S. Bering in 2012, and 251,000 for the U.S. Bering in 2013. The Chukchi Sea surveys were conducted into late May, so it may not be quite as simple as adding the Chukchi and Bering estimates together (i.e., Chukchi Sea estimates likely includes seals that wintered in the Bering Sea and migrated northward while surveys were being conducted); however, a combined aerial survey estimate around 500,000 seems reasonable. This is considerably higher than we estimated with CKMR, although the model with male heterogeneity in reproductive success comes close. However, we are in some sense only estimating the population of seals that are exposed to Alaska Native subsistence hunters (Conn et al. 2020), so there is good reason to suspect that the population we are estimating with CKMR is somewhat smaller than the entire Beringia DPS. However, it may be the most relevant population estimate for population management purposes.

References

Bravington, Mark V, Hans J Skaug, Eric C Anderson, et al. 2016. "Close-Kin Mark-Recapture." *Statistical Science* 31 (2): 259–74.

Caswell, H. 2001. *Matrix Population Models, 2nd Edition.* Sunderland, MA: Sinauer.

Choquet, Rémi, Anne Viallefont, Lauriane Rouan, Kamel Gaanoun, and Jean-Michel Gaillard. 2011. "A Semi-Markov Model to Assess Reliably Survival Patterns from Birth to Death in Free-Ranging Populations." *Methods in Ecology and Evolution* 2 (4): 383–89.

Conn, P. B., M. V. Bravington, S Baylis, and J. M. Ver Hoef. 2020. "Robustness of Close-Kin Mark-Recapture Estimators to Dispersal Limitation and Spatially Varying Sampling Probabilities." *Ecology and Evolution* 10: 5558–69.

Conn, P. B., and I. S. Trukhanova. 2022. "Modeling Vital Rates and Age-Sex Structure of Pacific Arctic Phocids: Influenc on Aerial Survey Correction Factors." *bioRxiv.* https://doi.org/https://doi.org/10.1101/2022.04.12.487942.

Fedoseev, G. A. 2000. *Population Biology of Ice-associated Forms of Seals and Their Role in the Northern Pacific Ecosystems.* Moscow, Russia: Center for Russian Environmental Policy, Russian Marine Mammal Council.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. 2015. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70: doi: 10.18637/jss.v070.i05.

R Development Core Team. 2017. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Richards, L. J., J. T. Schnute, A. R. Kronlund, and R. J. Beamish. 1992. "Statistical Models for the Analysis of Ageing Error." *Canadian Journal of Fisheries and Aquatic Sciences* 49: 1801–15.

Tikhomirov, E. A. 1966. "Reproduction of Seals of the Family Phocidae in the North Pacific." *Zoologicheskii Zhurnal* 45: 275-281.

Trukhanova, Irina S, Paul B Conn, and Peter L Boveng. 2018. "Taxonomy-Based Hierarchical Analysis of Natural Mortality: Polar and Subpolar Phocid Seals." *Ecology and Evolution* 8 (21): 10530–41.