Bearded seal CKMR modeling

Paul Conn

6/15/2023

Statistical modeling of kinship relationships

In this document I describe statistical CKMR models for bearded seal kinship data. The ultimate goal is to fit a number of different types of CKMR models, which progress in their complexity. First, I consider a basic model in which half-sib relationships and ages are known with certainty. Second, I look at the case where a cutoff for half-sibs based on PLOD scores is developed. In this case, we need to construct pairwise relationship probabilities that account for the probability a half-sib is missed because of a low PLOD score. Next, we'll look at the case where putative half-sibs actually represent a mixture of half-sibs and grandparent-granchild pairs. Finally, we'll consider models that include aging error. Each of these CKMR models will rely on a common age structured population model representing bearded seal population dynamics; I describe this model next before moving on to the CKMR models.

Bearded seal population dynamics model

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

$$N^F_{t,0} = N^M_{t,0} = 0.5 \sum_a N^F_{t-1,a} \phi_a f_{t-1,a},$$

where $N^F_{t,a}$ 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., $N^F_{t,a} = N^M_{t,a} = N_{t-1,a-1}\phi_{a-1}$ for $a > 0, t > 0$). During the initial year of the population dynamics model ($t = 1$), we set abundance values equal to stable stage proportions from the associated matrix population model (Caswell 2001).

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 it will generally be necessary to provide informative priors on survival and 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

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

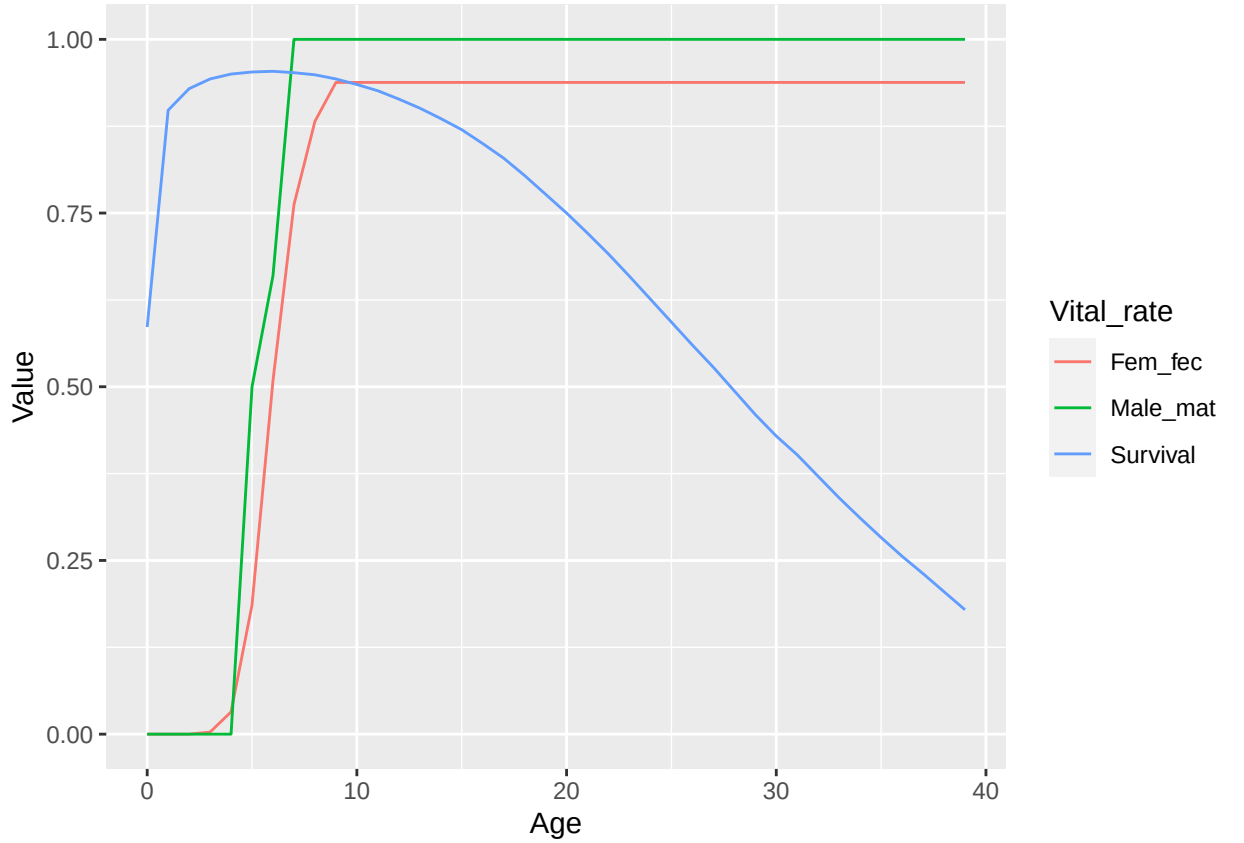
where η_1, η_2, and η_3 are estimated parameters. The values of these parameters from hierarchical analysis (Trukhanova, Conn, and Boveng 2018) were $\eta_1 = 0.055$, $\eta_2 = 2.80$, and $\eta_3 = 0.076$ (Conn et al. 2020).

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("../data/Maturity.csv")
Survival = read.csv("../data/Survival_ests.csv")
Reprod = read.csv("../data/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)
ggplot(Plot_df)+geom_line(aes(x=Age,y=Value,colour=Vital_rate))
```



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 $\lambda \approx 1.0$. 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.

CKMR modeling for certain ages

Our first CKMR model will assume ages are certain. Inference will be 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 will base inference on the joint pseudo-likelihood

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

where L_{pop} is a product Bernoulli likelihood for parent-offspring pairs, L_{hsp} is a product Bernoulli likelihood, $f(\eta)$ 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 $\lambda \neq 1.0$. I now describe each of these components (including data and parameter specifications) in turn.

The likelihood for parent-offspring kin comparisons (L_{pop}) is a product Bernoulli distribution, specified as

$$L_{pop} = \prod_i \prod_j p_{ij1}^{y_{ij1}} (1 - p_{ij1})^{1 - y_{ij1}} I_1(i,j).$$

Here, i and j index two individuals, b_i and b_j denote the years of their births respectively, y_{ij1} is a binary random variable that equals 1.0 if i and j are parent-offspring pairs, and is zero otherwise. and $I_1(i,j)$ is an indicator function. The indicator function is used to prevent double counting, and also to omit certain comparisons that are likely to violate independence assumptions for the Bernoulli model. For instance, we set $I_1(i,j) = 0$ whenever the year of i's birth (b_i) is greater than or equal to the year of j's birth (b_j); we also set $I_1(i,j) = 0$ whenever (1) i is female and (2) i and j are both harvested in the year of j's birth (i.e., $d_i = d_j = b_j$). The latter restriction is to prevent dependency in harvests of mothers and pups, which can positively bias CKMR abundance estimates.

Calculations of p_{ij1} differ based on whether the potential parent is male or female. In addition to reproductive output being based on maturity-at-age for males and fecundity-at-age for females, there is also the issue of timing of reproduction and pupping. Specifically, females need to survive to the spring when their pup is born, while males could still potentially breed the preceding May-June and still sire a pup if they were harvested after that. In general, the p_{ij} are based on the concept of relative reproductive output (Bravington et al. 2016). If s_i denotes the sex of individual i ($s_i = 0$ if i is female, $s_i = 1$ if i is male), we have

$$p_{ij1} = \begin{cases} \frac{m_{b_j - b_i - 1}}{\sum_a N^M_{b_j - 1, a} m_a} & \text{if} \quad d_i \geq b_j - 1 \;\&\; (b_j - b_i) < 40 \;\&\; s_i = 1 \\ \frac{f_{b_j - b_i}}{\sum_a N^F_{b_j, a} f_a} & \text{if} \quad d_i \geq b_i \;\&\; (b_j - b_i) < 40 \;\&\; s_i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that I include the $(b_j - b_i) < 40$ restriction since our population model is restricted to ages 0-39 and parameters are undefined past this range.

For HSPs, the likelihood L_{hsp} is again a product Bernoulli,

$$L_{hsp} = \prod_i \prod_j p_{ij2}^{y_{ij2}} (1 - p_{ij2})^{1 - y_{ij2}} I_2(i,j),$$

where the "2" subscript simply denotes that data, success probabilities, and constraints are particular to HSPs. In this case, we never directly observe the common parent, although we assume that we know what sex it is since maternally related HSPs share mitochondrial DNA. Technically, it is possible that paternally

related HSPs could have the same mitochondrial DNA by random chance, though this is a low probability (e.g, $\approx 1\%$) because mitochondrial haplotype diversity is quite high. For HSPs, we require that $I_2(i,j) = 0$ whenever $b_i > b_j$ to prevent double counting.

Success probabilities p_{ij2} once again reflect relative reproductive output, but we must sum over possible parent ages since we never observe the parent directly. Also, the prospective parent must survive from $b_i \to b_j$ (for females), and from $b_{i-1} \to b_{j-1}$ for males. Once again, we need to have age restrictions to prevent the potential parent from achieving an age ≥ 40 where parameters are undefined. Letting $\delta_{ij} = b_j - b_i$ and doing some algebra, we have

$$p_{ij2} = \begin{cases} \frac{\sum_{a=0}^{40-\delta_{ij}} N^M_{b_i-1,a} m_a m_{a+\delta_{ij}} \prod_{c=a}^{a+\delta_{ij}-1} \phi_c}{\{\sum_c N^M_{b_i-1,c} m_c\}\{\sum_c N^M_{b_j-1,c} m_c\}} & \text{if} \quad s_i = 1 \\ I_{ij} \frac{\sum_{a=0}^{40-\delta_{ij}} N^F_{b_i,a} f_a f_{a+\delta_{ij}} \prod_{c=a}^{a+\delta_{ij}-1} \phi_c}{\{\sum_c N^F_{b_i,c} f_c\}\{\sum_c N^F_{b_j,c} f_c\}} & \text{if} \quad s_i = 0. \end{cases}$$

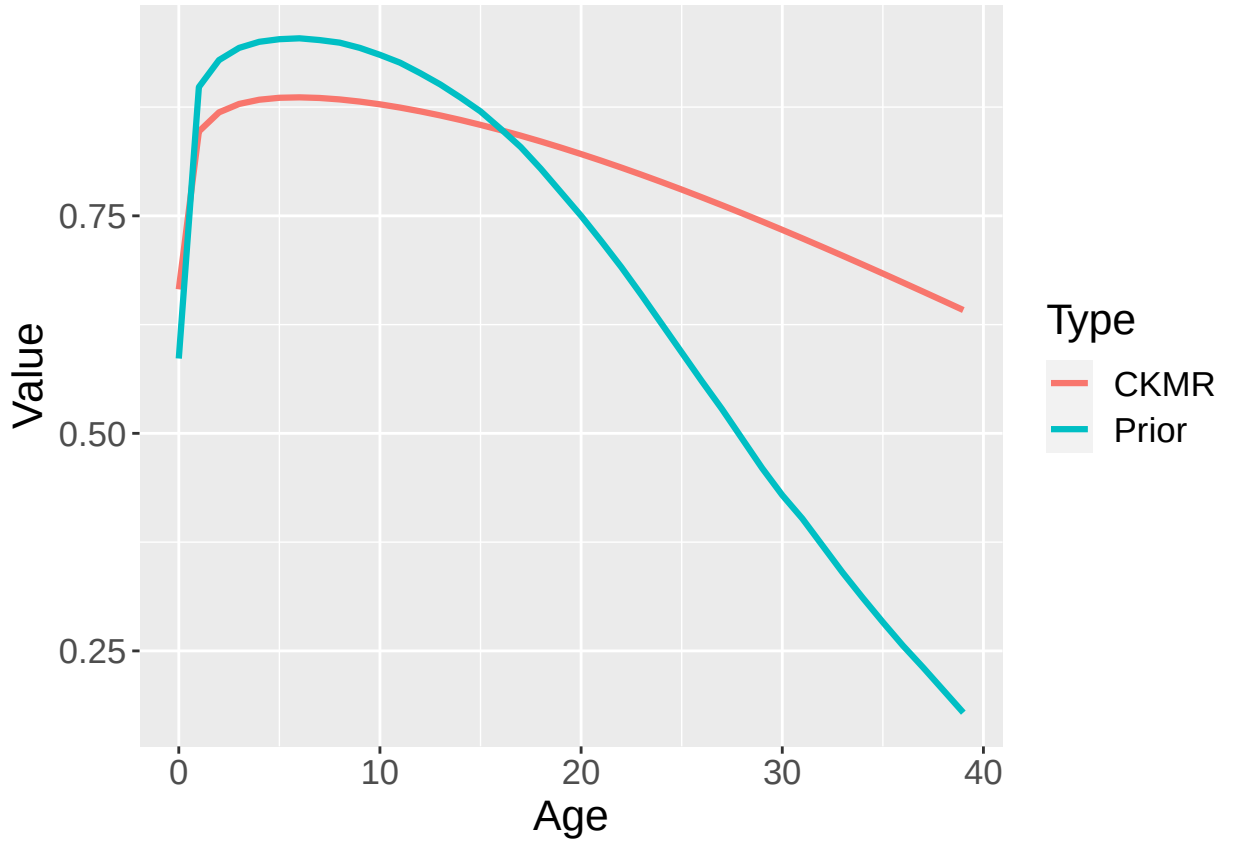
The indicator $I_{ij} = 1$ if $\delta_{ij} > 0$ and is used to disallow maternal MHPs from occurring if birth years of the prospective kin pairs are the same (since females only have one pup per year). For males, whenever $\delta_{ij} = 0$, we set $\prod_{c=a}^{a+\delta_{ij}-1} \phi_c = 1.0$.

The last two terms in the integrated pseudo likelihood are $f(\boldsymbol{\eta})$ and Λ_λ. For $f(\boldsymbol{\eta})$, I specified independent Gaussian prior distributions for RAW parameters, with a mean set to the values estimated from hierarchical meta-analysis (Trukhanova, Conn, and Boveng 2018), and with a standard deviation set so as to achieve a coefficient of variation (CV) of approximately 0.2 on the real scale. For Λ_λ, I set a Gaussian penalty on the realized finite rate of population growth λ, such that $\lambda \sim \text{Normal}(1.0, 10^{-8})$. Log link functions were used on all parameters (abundance, RAW parameters) to constrain real-valued estimates to be > 0.

To fit this model, I programmed the log-likelihood in Template Model Builder (TMB; (Kristensen et al. 2015)). Conditioning on observed kinship observations (i.e., the y_{ijk}), I then minimized it as a function of bearded seal abundance 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 separate comparison computations every time the likelihood is evaluated (though pairwise comparisons may be easier to implement in ageing error models). Here is code to build and fit a "certain age" CKMR model to bearded seal data. We will start our population model one generation (40 years) prior to data collection beginss to allow us to model relative reproductive output of parents of half-siblings that are encountered near the beginning of the study.

The joint negative log pseudo-likelihood was minimized very quickly (in 1.59360480308533 seconds), and gives an estimate of $\hat{N} = 1.64 \times 10^5$. Owing to the heavy constraint on $\lambda = 1.0$, the estimate is fairly precise with SE $= 3.18 \times 10^4$. Recall that the initial combination of survival-at-age and fecundity-at-age conspired to produce a population rate of increase that was roughly 4% per year; since fecundity-at-age was set to be constant during estimation, the minimization procedure was only able to achieve a constant population growth rate by adjusting survival-at-age parameters ($\eta_1, \eta_2,$ and η_3). Presumably there was also some information about adult survival from the differences in half-sibling pair ages. At any rate, estimated survival-at-age peaked at a lower value but had a higher estimated value than the prior at later ages. Let's take a look at the difference between the two

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




```
png("bearded_surv_prior_posterior.png")
ggplot(Plot_df) +
  geom_line(aes(x = Age, y = Value, colour = Type), size = 1.1) +
  theme(text = element_text(size = 16))
dev.off()
```

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

The estimate obtained here is almost certainly negatively biased. There are several of factors contributing to this conclusion:

-I am currently not allowing for the fact that some of the half-siblings are likely grandparent-grandchild pairs, which aren't generally possible to tell apart. This inflates the number of half-sibling pairs, and will tend to decrease the resulting abundance estimate.

-Despite slightly greater probabilities of HSPs for females relative to males, there were only 8 maternal HSPs compared to 17 males.
By contrast, conditional on estimated parameters, the number of expected maternal half-sibling pairs was 4.58×10^{215} and the number of expected paternal HSPs was -1.31×10^{228}. A binomial CDF test with equal probability of PHS and MHS gives a p-value of 0.05, suggesting some evidence for increased paternal HSPs above and beyond what we would expect by random chance. Heterogeneity in reproductive success of sexually mature males would also tend to inflate HSP matches and result in lower abundance estimates.

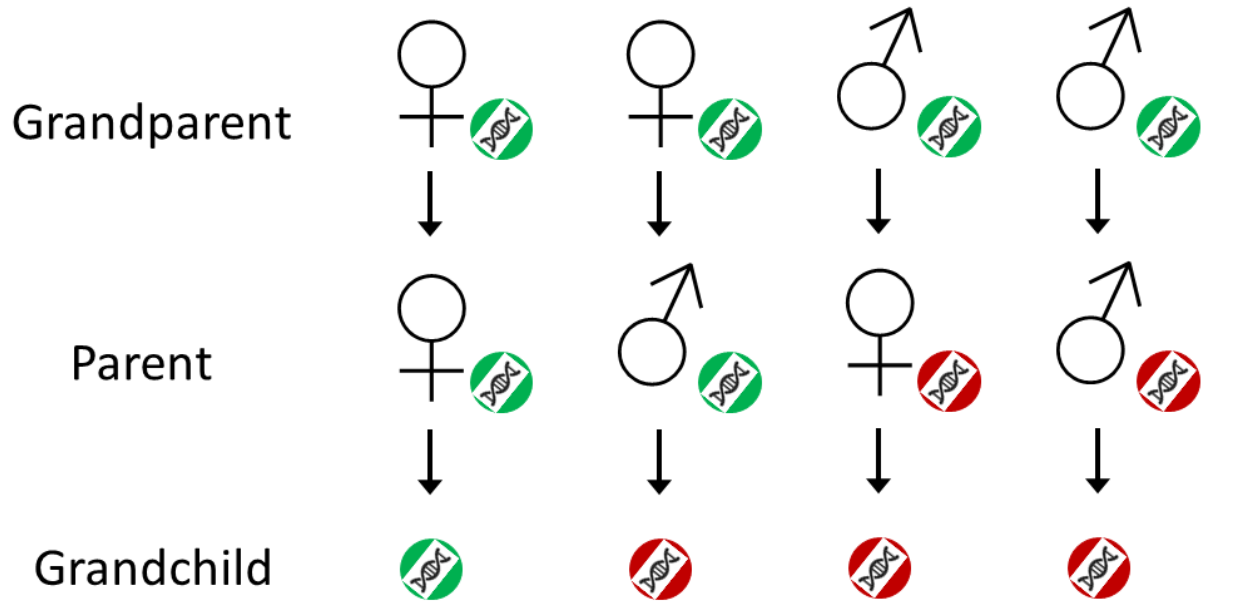
-We haven't yet removed any half sibs with low PLOD scores. There is thus the possibility that the number of half-sib matches is too high because of contamination by lower order kin. This would also negatively bias our abundance estimator.

Our initial estimate is also likely too precise, both because of the above factors and because we haven't

appropriately accommodated aging uncertainty. These shall be the next focuses of our modeling efforts.

Accounting for grandparent-grandchild pairs

Although there were relatively few comparisons with birth gaps long enough to actually be grandparent-grandchild pairs (GGPs), it would probably be worth accounting for this possibility, especially since the absence of grandparent-granchild pairs could also be informative. In order to account for GGPs, we can model apparent HSPs as a mixture of HSPs and GGPs. This works differently for comparisons that share the same mitochondrial haplotype vs. those that don't. In particular, if the older animal is female, there is only a $\approx 50\%$ chance that GGPs will share mitochondrial DNA since the (unobserved) mother must also be female, as shown in this cartoon:



Here, the green DNA bits show inheritance of mtDNA, with green mtDNA being identical to that possessed by the grandparent. Importantly, if the potential grandparent is male, the chance that the GGP shares mtDNA is negligible (this would not be the case in populations with low mtDNA haplotype diversity). For purposes of bearded seals, we'll start by grouping apparent half-sibling comparisons by the sex of the older individual (for same-age comparisons, the sex of an arbirtrary seal is chosen), and also by birth years, and death year of the older animal. Associated matches can either share mtDNA or not (final array dimension).

```
n_comp_HSGGP_sibidibj <- array(0, dim = c(2, n_yrs, n_yrs_data, n_yrs))
n_match_HSGGP_sibidibjmij <- array(0, dim = c(2, n_yrs, n_yrs_data, n_yrs, 2))

for (iseal in 1:(n_seals - 1)) {
  for (jseal in (iseal + 1):n_seals) { # prevent double counting
    if (BY[iseal] <= BY[jseal]) {
      n_comp_HSGGP_sibidibj[(Sex[iseal] == "M") + 1, BY[iseal], DY[iseal], BY[jseal]] <-
        n_comp_HSGGP_sibidibj[(Sex[iseal] == "M") + 1, BY[iseal], DY[iseal], BY[jseal]] + 1
    }
    if (BY[iseal] > BY[jseal]) {
      n_comp_HSGGP_sibidibj[(Sex[iseal] == "M") + 1, BY[jseal], DY[jseal], BY[iseal]] <-
        n_comp_HSGGP_sibidibj[(Sex[iseal] == "M") + 1, BY[jseal], DY[jseal], BY[iseal]] + 1
    }
    # we'll need to make probs 0 for MHSPs since only one pup/yr; restrictions will need to
```

```
    # be made on ages in TMB to prevent GGP comparisons when death of grandparent occurs in inadmissabl
  }
}

# matches
# HSPs
for (imatch in 1:nrow(CKMR_certain_age$matches_PHS)) {
  if (BY[CKMR_certain_age$matches_PHS[imatch, "i"]] <= BY[CKMR_certain_age$matches_PHS[imatch, "j"]]) {
    n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_PHS$i[imatch]] == "M") + 1, BY[CKMR_certain
      n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_PHS$i[imatch]] == "M") + 1, BY[CKMR_certa
  } else {
    n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_PHS$j[imatch]] == "M") + 1, BY[CKMR_certain
      n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_PHS$j[imatch]] == "M") + 1, BY[CKMR_certa
  }
}

for (imatch in 1:nrow(CKMR_certain_age$matches_MHS)) {
  if (BY[CKMR_certain_age$matches_MHS[imatch, "i"]] <= BY[CKMR_certain_age$matches_MHS[imatch, "j"]]) {
    n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_MHS$i[imatch]] == "M") + 1, BY[CKMR_certain
      n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_MHS$i[imatch]] == "M") + 1, BY[CKMR_certa
  } else {
    n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_MHS$j[imatch]] == "M") + 1, BY[CKMR_certain
      n_match_HSGGP_sibidibjmij[(Sex[CKMR_certain_age$matches_MHS$j[imatch]] == "M") + 1, BY[CKMR_certa
  }
}
```

We'll need to reformulate our .cpp TMB file to account for updated half-sib/GGP probabilities. There are four different possibilities, associated with whether (1) the sex of the parent, and (2) whether or not the two individuals compared share mitochondrial DNA. Let's first describe what the probability of a GGP is for these cases, before describing combined probabilities (GGP + HSP). To do this, let s_i denote sex of the older seal (with $s_i = 1$ if i is male), and m_{ij} be a binary random variable that takes on the value 1 if indivduals i and j share mtDNA.

Case 1: $s_i = 0$, $m_{ij} = 1$

We'll denote the probability of a GGP sharing mtDNA when the potential grandparent is female as $Pr(GGP, m_{ij} = 1 | s_i = 0, d_i, b_i, b_j)$. Note that this expression depends on the time of death of the potential grandparent (for instance, if it dies before it was old enough to have potentially reproduced, it is clearly not a grandparent). Note also that we are assuming (by virtue of high mtDNA haplotype diversity) that a grandparent and grandchild can only share mtDNA if the unobserved parent is female . Accordingly,

$$Pr(GGP, m_{ij} = 1 | s_i = 0, d_i, b_i, b_j) = \sum_{t=b_i}^{\min(d_i, b_j)} \frac{f_{t-b_i} N_{t,0}}{\sum_a f_a N_{t,a}} \frac{\{\prod_{k=t}^{b_i - 1} \phi_{k-t}\} 2 f_{b_j - t}}{\sum_a f_a N_{b_j, a}} = \sum_{t=b_i}^{\min(d_i, b_j)} \frac{f_{t-b_i} N_{b_j, b_j - t}}{\sum_a f_a N_{t,a}} \frac{2 f_{b_j - t}}{\sum_a f_a N_{b_j, a}}$$

Here, relative reproductive success is conditional on the unknown age of the parent, so we must sum over the possible years (t) of the mother's birth. The $N_{b_j, b_j - t}$ in the numerator arises because the potential parent can be any of the females born in year t that survive to the year of j's birth. For seals, many of the f_a values are zero for low ages, so practically speaking we must have a sufficient birth gap (and late enough time of death for the potential grandparent) to enable $Pr(GGP) > 0$. As in previous calculations, this formulation requires a number of things to hold like equal male:female sex ratios, equal survival among sexes, etc.

Case 2: $s_i = 0, m_{ij} = 0$

The only way for a grandmother and grandchild not to share mtDNA is if the unobserved parent is male, so our answer will be similar but will involve male maturity-at-age indexed to the year before j's birth:

$$Pr(GGP, m_{ij} = 0 | s_i = 0, d_i, b_i, b_j) = \sum_{t=b_i}^{\min(d_i, b_j - 1)} \frac{f_{t-b_i}}{\sum_a f_a N_{t,a}} \frac{N_{b_j-1, b_j-t-1} 2 m_{b_j-t-1}}{\sum_a m_a N_{b_j-1,a}}$$

Case 3: $s_i = 1, M_{ij} = 1$

Now we're on to the males. This one is easy; by assumption we assume that $Pr(GGP, m_{ij} = 1 | s_i = 1, d_i, b_i, b_j) = 0$ for reasons stated previously.

Case 4: $s_i = 1, M_{ij} = 0$

This case can happen whether the offspring of i is male or female, so we have to account for both. Fortunately, it is very similar to what we have written already, though it involves male maturity in the year previous to the birth of the prospective parent:

$$\begin{aligned} Pr(GGP, m_{ij} = 0 | s_i = 1, d_i, b_i, b_j) = & \sum_{t=b_i+1}^{\min(d_i, b_j)} \frac{m_{t-b_i-1}}{\sum_a m_a N_{t-1,a}} \frac{N_{b_j, b_j-t} 2 f_{b_j-t}}{\sum_a f_a N_{b_j,a}} \\ & + \sum_{t=b_i+1}^{\min(d_i, b_j)} \frac{m_{t-b_i-1}}{\sum_a m_a N_{t-1,a}} \frac{N_{b_j-1, b_j-t-1} 2 m_{b_j-t-1}}{\sum_a m_a N_{b_j-1,a}} \end{aligned}$$

After formalizing these probabilities in a new .cpp file, let's recompile code with updated data structures and see what we get.

The time to fit the combined GGP-HSP model increased quite a bit (up to 33.2966520786285), which isn't surprising since likelihood calculations require an additional type of kin and are disaggregated (now requiring calculations by the sex and year of death of the potential parent). The abundance estimate has also gone up to 2.06×10^5. The expected number of grandparent-grandchild given our data and model fit is 5.42.

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], HSP
    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], HSP
  }
}
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
```

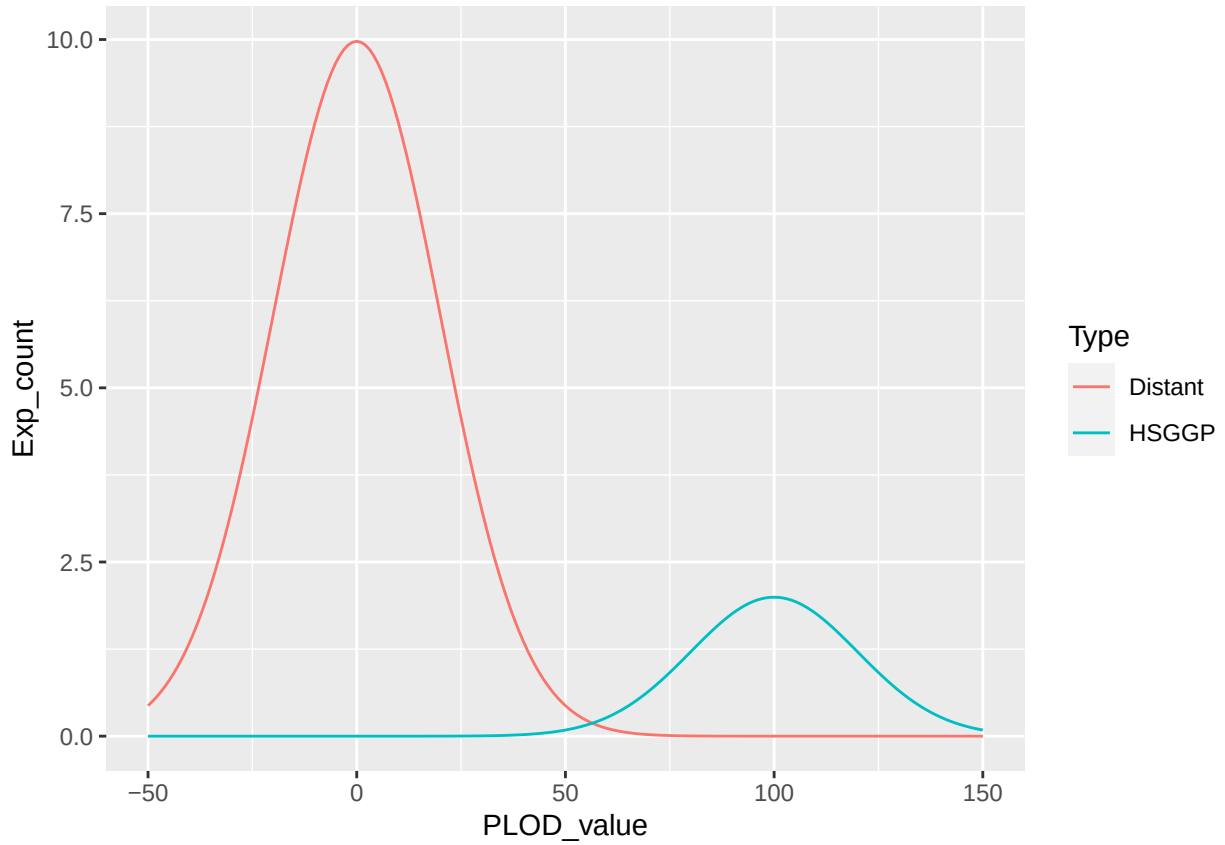
```
## 5          1      40      14      51     1  3.7e-06  4.3e-07         0.90
## 6          1      41       7      52     1  3.7e-06  4.3e-07         0.90
## 7          2      52      13      52     1  2.1e-05  0.0e+00         1.00
## 8          2      48      10      53     1  1.1e-05  0.0e+00         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).

Investigating alternative PLOD thresholds

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 non-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.

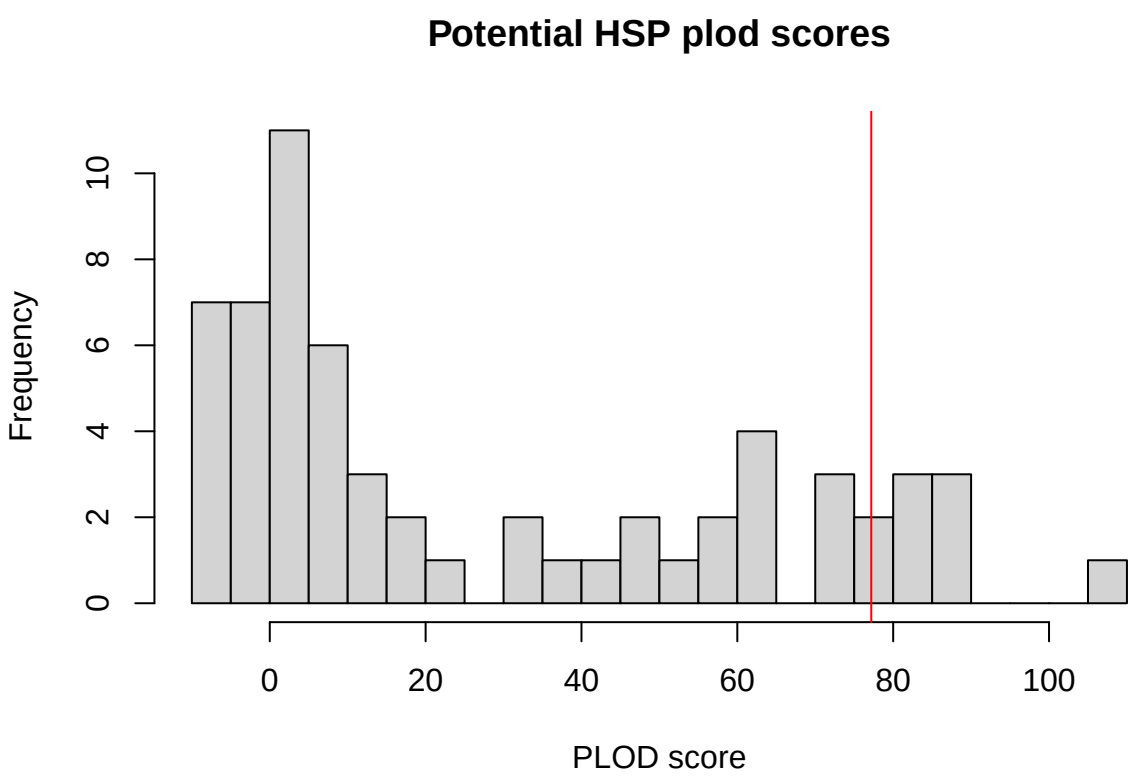
$$d = \int_{x=70}^{\infty} f(x; \mu, \sigma^2) dx$$

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")
```




```
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).

```
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., 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 look at analyses with each of these possible HSP plod cutoffs (30, 40, or 50).

Let's run a series of models accounting for different thresholds and false negative HSP/GGP probabilities (I'm hiding code because I have to recalculate the n_match arrays for HSGGPs.)

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 $\hat{N} = 2.06 \times 10^5$; a 40 cutoff resulting in $\hat{N} = 2.32 \times 10^5$, and a 50 cutoff resulting in $\hat{N} = 2.61 \times 10^5$. 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. However, right now, it is quite real.

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 some alternative trend scenarios, including (1) a constant 2% rate of increase over time, and (2) a constant 2% rate of 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 PLOD=40 scenario...

This exercise illustrates several phenomenon. First, abundance estimates intersect in 2003, 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 three trend (λ) values, with L = -304.96, -303.21, and -304.77 for decreasing, stable, and increasing population models, respectively. Interestingly, there appears to be slight evidence that the population is either increasing or decreasing, rather than being stable, but our experience is that it can actually be quite hard to estimate population trend from close kin data - we would certainly want a lot more kin pairs for reliable trend estimation.

Male heterogeneity

Looking at our data, thre are a total of 17 HSP/GGPs that don't share mtDNA, and 8 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 chance of observing so few maternal HSPs is 0.05. This could have happened by random chance, but it also may have happened because of 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). Bearded seals are known to maintain underwater territories during breeding season, and it may be the case that there may be some competition for mates.

In order to account for this possibility, we conducted a sensitivity scenario where we assumed the number of male breeders is an unknown fraction (π) of the total number of reproductively mature males. Specifically, we set $N_{t,a}^{M} = 0.5 N_{t,a} \pi$ every where 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 $\hat{N} = 4.09 \times 10^5$.

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

```
N_df <- data.frame(matrix(0, 6, 5))
colnames(N_df) <- c("lambda", "PLOD_cutoff", "male_het", "N_hat", "CV")
N_df$lambda <- c(1, 1, 1, 1, 1.02, 0.98)
N_df$PLOD_cutoff <- c(40, 40, 50, 30, 40, 40)
N_df$male_het <- c("no", "yes", "no", "no", "no", "no")
N_df$N_hat <- c(Report_40$N[1], Report_het$N[1], Report_50$N[1], Report_30$N[1], mean(Report_lambda_inc$
N_df$CV <- c(
  SD_N_40[1] / Report_40$N[1], SD_N_het[1] / Report_het$N[1], SD_N_50[1] / Report_50$N[1], SD_N_30[1] /
  mean(SD_N_lambda_inc[32:62] / Report_lambda_inc$N[32:62]), mean(SD_N_lambda_dec[32:62] / Report_lambda
)
print(N_df)
```

```
##   lambda PLOD_cutoff male_het  N_hat   CV
## 1   1.00          40       no 231814 0.21
## 2   1.00          40      yes 408651 0.35
## 3   1.00          50       no 261075 0.22
## 4   1.00          30       no 206333 0.19
## 5   1.02          40       no 243463 0.20
## 6   0.98          40       no 229311 0.17
```

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.