# Screw trap model

mark sorel

8/21/2020

## Purpose

The purpose of this document is to describe the methods I am using to estimate the abundance and timing of juvenile spring Chinook salmon emigrating from the Chiwawa River, Nason Creek, and the White River.

### Data

The emigration of juveniles from natal streams in the Wenatchee River basin has been monitored with downstream-migrant screw traps operated near the mouth of the Chiwawa River since 1997, Nason Creek since 2004, and the White River since 2006 (Johnson et al. 2018). Most juveniles are captured emigrating from their natal stream as subyearling parr in summer or fall, or yearling smolts in spring.

### Catch data

One component of the data are the numbers of fish captured on each date that a trap was operating without major disruption. The catch is divided among subyearlings and yearlings based on the length and date of emigration. For details of the age delineation, see `src/juvenile age explore.Rmd`. In addition, the river discharge on each date is used in the model, as a covariate on trap capture efficiency.

### Eficiency trials

Another component of the data are the numbers of fish marked and released upstream of the trap and the number of those fish that were recaptured. These mark-recapture trails are conducted on several days in each year at each trap to evalute the capture probabilities in the traps.

### Model

Each stream and age of fish is currently modeled seperately (6 total models). I will describe the model for a single stream by age combination below.

### Daily emigrant process model

The number of emigrants $\hat{m}[t, y]$ on day $t$ in year $y$ is modeled on the log scale as a function of a yearly average $\mu[y]$, accross-year average daily errors $\delta[t]$, and day- and year-specific errors $\epsilon[t, y]$:

$$\log(\hat{m}[t, y]) = \mu_m[y] + \delta[t] + \epsilon[t, y]$$

The accross-year average daily errors $\delta[t]$ and the year-specific daily errors $\epsilon[t, y]$ are modeled as stationary 1st order autoregressive processes with autocorrelation coefficients $\phi_d$ and $\phi_e$, and marginal variances $\sigma_d^2$ and $\sigma_e^2$, for example:

$$\delta[0] \sim N(0, \sigma_d^2)$$
$$\delta[1] = \phi_d \; \delta[1] + \sqrt{1-\phi_d} \; \gamma[1], \quad \gamma[1] \sim N(0, \sigma_d^2)$$
$$\delta[t] = \phi_d \; \delta[t-1] + \sqrt{1-\phi_d} \; \gamma[t], \quad \gamma[t] \sim N(0, \sigma_d^2)$$

The year-specific autoregressive processes are assumed to have a common autocorrelation coefficient $\phi_e$ and marginal variance $\sigma_e^2$.

### Trap Processs

The capture probability $p[t, y]$ of a fish emigrating past a trap is modeled on the logit scale as a function of discharge:

$$\text{logit}(p[t, y]) = \nu + \beta * \text{disch}[t, y]$$

where $\nu$ is an intercept and $\beta$ is a coefficient for the effect of z-scored discharge on logit capture probability.

### Likelihood

The number of fish recaptured $rec[t, y]$ out of the number of marked fish released upstream of the trap $rel[t, y]$ is assumed to be binomially distributed:

$$rec[t, y] \sim \text{Bin}(rel[t, y], \; p[t, y])$$

Thus it is assumed that all released fish pass the trap on the same day of their release, given that the majory of recaptured fish are recovered the morning following their release. Recaptures occuring one or more days after a release are included in the recapture total for that day. This likely causes some bias in the estimates of $p$, but the bias is likely small given that the majority of fish meet the assumption.

The expected catch in the trap $\hat{c}[t, y]$ is equal to the product of the number of emigrants and their capture probabilities,

$$\hat{c}[t, y] = \hat{m}[t, y] \; p[t, y]$$

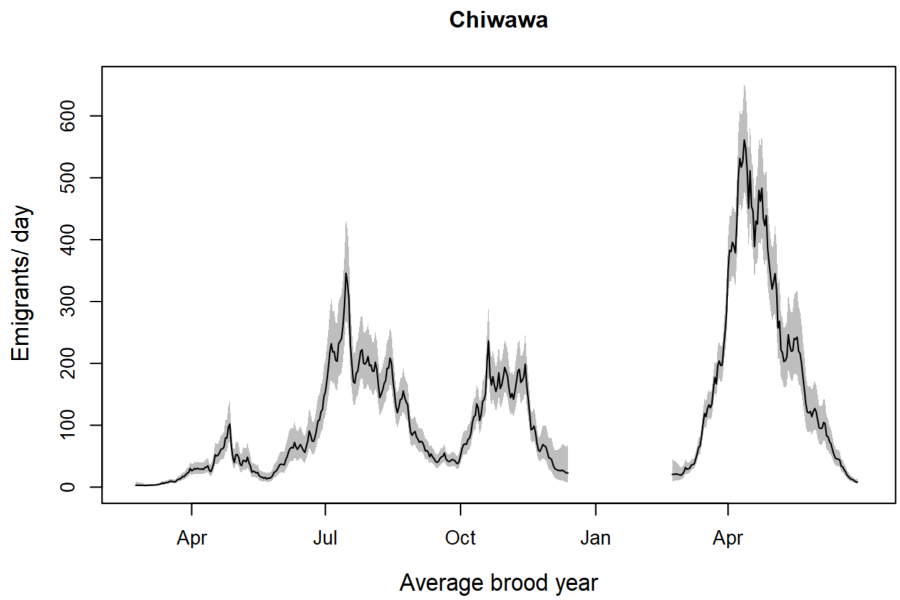
The observed catches $c[t, y]$ are assumed to follow a negative binomial distribution with mean $\hat{c}[t, y]$ and variance $\hat{c}[t, y] + \frac{\hat{c}[t,y]^2}{\phi_{nb}}$. Thus, the additional variance of observations above that of a Poisson with mean $\hat{c}[t, y]$ is controlled by the inverse of the paramater $\phi_{nb}$, scaled by the square of the mean. The overdispersion parameter $\phi_{nb}$ is assumed to be the same for all years.

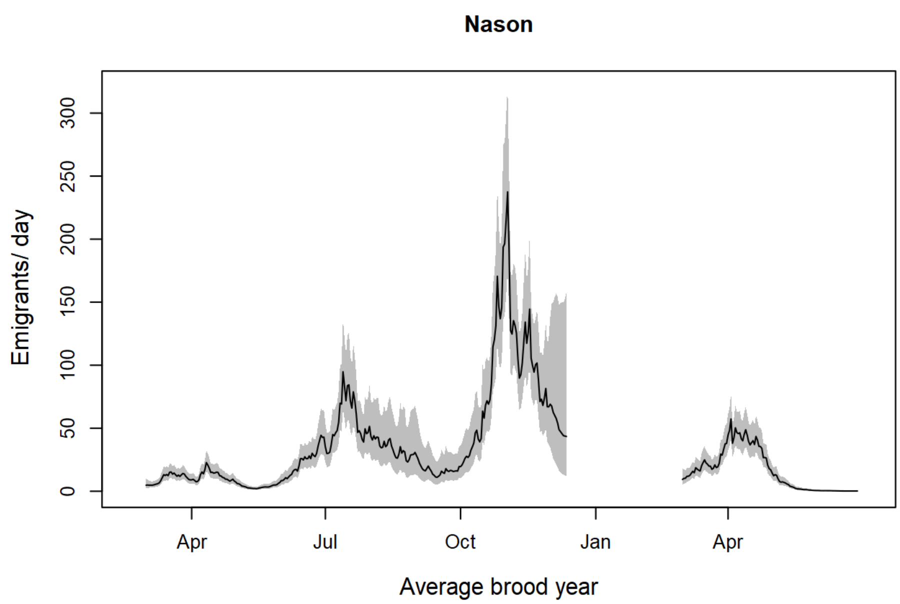
### Parameter estimation

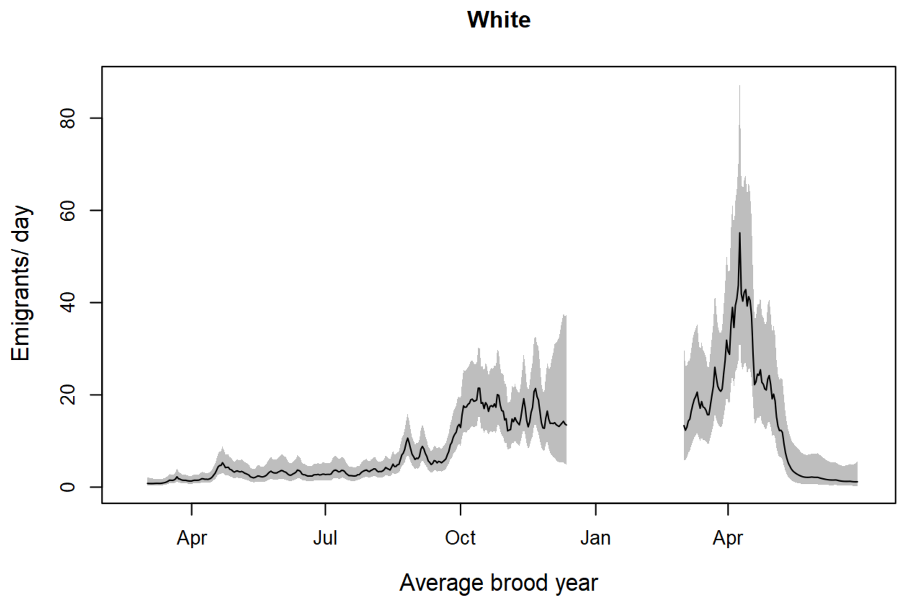
The parameters are estimated within a mixed effects model, treating the autocorrelated errors $\delta[t]$ and $\epsilon[t, y]$ as random effects. Paramaters are estimated by maximizing the marginal likelihood calculated by Template Model Builder within R. Empiracal Bayes estimates of the random effects are used to estimate the unobserved numbers of migrants. Standard deviations are estimated useing the delta method.

### Results

Plots of the geometric mean numbers of daily emigrants across years for each stream.


Chiwawa
Emigrants/ day
0
100
200
300
400
500
600
Apr
Jul
Oct
Jan
Apr
Average brood year


Nason
Emigrants/ day
0
50
100
150
200
250
300
Apr
Jul
Oct
Jan
Apr
Average brood year




For a life cycle model, I need some way of modeling the transition from spawners to emigrants. My current thinking is to aggregate emigrants into discrete groups based on emigration day of year. This could be simpler than modeling the number of *daily* emigrants within the life cycle model, but still accounts for some of the diversity in emigration timing, which I must consider becasue it effects survival between emigration from the natal stream and smolting.

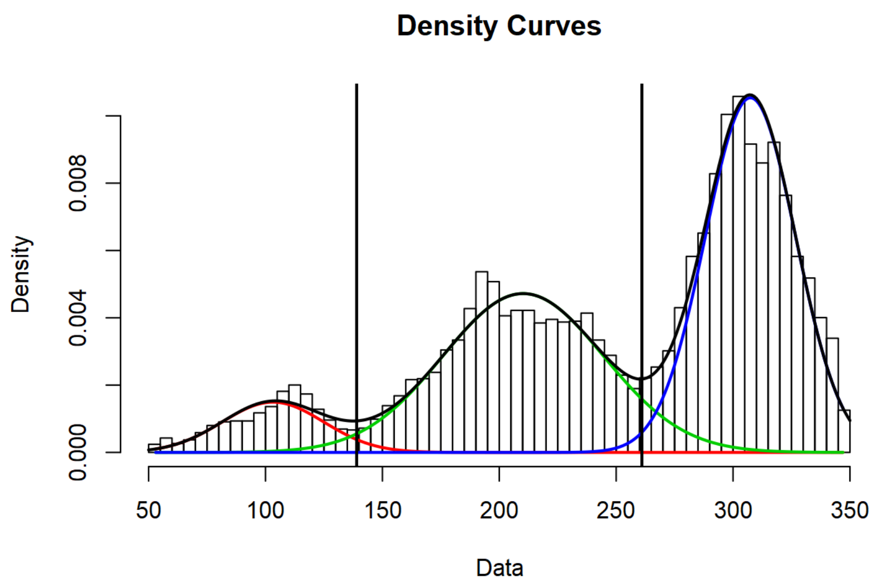
Based on visual examination of the average time series of dialy emigrants, it appears that they could be divided into three groups of subyearling emigrants and one group of yearlings My approach for dividing the subyearlings into groups was to take the geometric mean accross streams of the geometric means across years of daily emigrants for each stream. I then rounded the average number of dialy emigrants to the nearest fish and fit a mixture distribution of three normals to this "average year". Finally, I found the two days corresponding with the lowest densities of the mixtrure distribution between the modes to use as cutoffs for aggregating emigrants.

```
#geometric mean subyearlings across streams
across_stream_geomean<-data.frame(x=1:365,
                                  y=exp(rowMeans(geo_means$all_log_means[,1:3],na.rm=T))) %>%
  filter(!is.na(y)) # drop days with no observations

#expand to daily observation for each (rounded) fish
observation_for_each_average_fish<-rep(across_stream_geomean$x,
                                       times=round(across_stream_geomean$y))

#fit three-normal mixture distribution
mix_geomean<-mixtools::normalmixEM(observation_for_each_average_fish,
                                   lambda=c(.2,.4,.5),
                                   mu=c(100,200,300),)
```

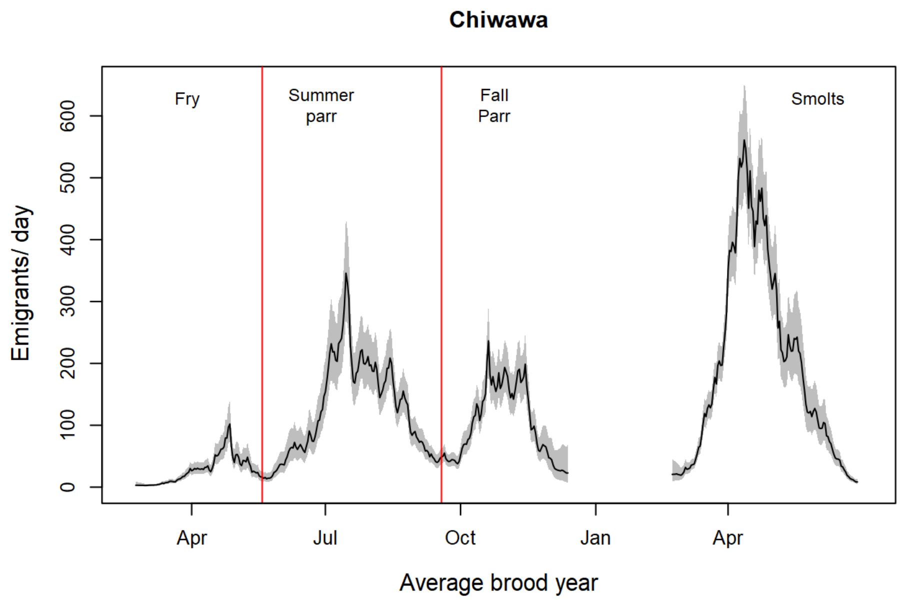
```
number of iterations= 136
```

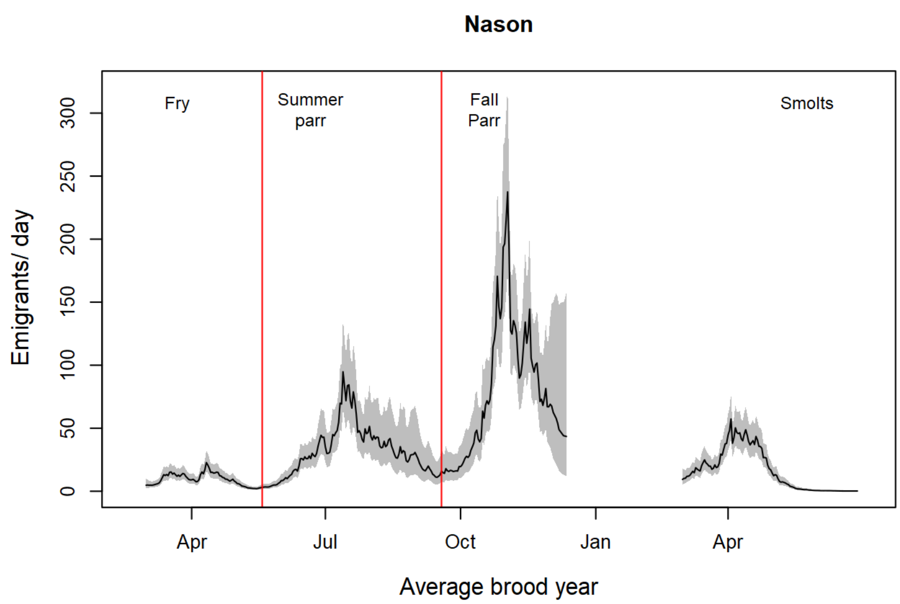
```
#calculate density
mix_geomean_dens<-mix_geomean$lambda[1]*
```

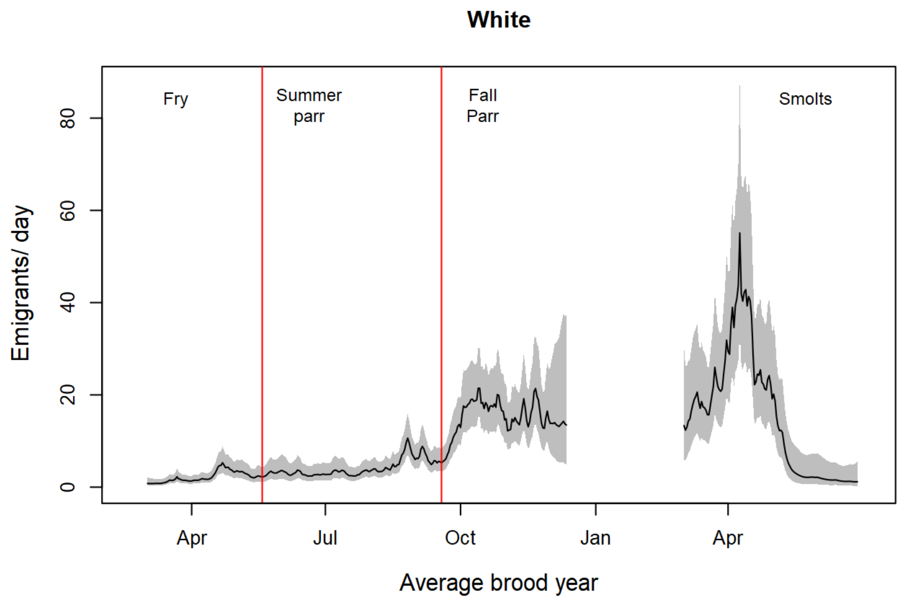
```r
  dnorm(seq(50,350,by=1),mix_geomean$mu[1],mix_geomean$sigma[1])+
  mix_geomean$lambda[2]*
  dnorm(seq(50,350,by=1),mix_geomean$mu[2],mix_geomean$sigma[2])+
  mix_geomean$lambda[3]*
  dnorm(seq(50,350,by=1),mix_geomean$mu[3],mix_geomean$sigma[3])

#plot
plot(mix_geomean,2,breaks=100)
points(seq(50,350,by=1),mix_geomean_dens,type="l",lwd=2)
breaks<-numeric(2)
for ( i in 1:2) breaks[i]<-which.min(
  mix_geomean_dens[((round(mix_geomean$mu[i]):round(mix_geomean$mu[(i+1)]))-49)])+
  round(mix_geomean$mu[i])

abline(v=breaks,lwd=2)
```




```r
x<-plot_all_geomean_daily_emigrants(all_emigrants_estimates,breaks=breaks)
```


Chiwawa
Fry
Summer parr
Fall Parr
Smolts
Emigrants/ day
0
100
200
300
400
500
600
Apr
Jul
Oct
Jan
Apr
Average brood year


Nason
Fry
Summer parr
Fall Parr
Smolts
Emigrants/ day
0
50
100
150
200
250
300
Apr
Jul
Oct
Jan
Apr
Average brood year




### Alternative approach

While the approach outlined above seems straightforward and reasonable, I could explicitly model the mixture within the estimation model for the numbers of daily emigrants. By this I mean I could replace the average-annual errors $\delta[t]$ in the daily emigrant process model described above with a smoothing function that uses 3 normal distributions as the bases. The model would be

$$\hat{m}[t,y] = \mu_m[y] \sum_{j=1}^{3} \left( \lambda[j] \frac{1}{\sigma_n[j]\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu_n[j]}{\sigma_n[j]}\right)^2} \right) e^{\epsilon[t,y]}$$

where $\lambda[j]$ is the weight for mixture component $j$ and the sum across all 3 $\lambda[j]$ is unity, $\mu_n[j]$ is the mean, and $\sigma_n[j]$ is the standard deviation.I would integrate the analyses of all three streams if I were to use this approach. I could then generate the cutoff using the same approach of finding the saddle point between the modes. I think that cutoffs are probably preferable to deviding the emigrants on each day according to the mixture because when it comes time to estimate survival rates of each group, it would be easier to be able to just assign a given fish (capture history) to a group.