Fence expansion prelinary report

Trevor Davies

2022-03-01

Contents

1 Introduction and Project Scope

The Babine river has a salmonine enumeration facility encounters all five Pacific salmon species in addition to steelhead. The primary goal of the facility is to enumerate sockeye, chinook and pink salmon whose runs span from mid-July and are generally over by mid-October. The Coho run frequently continues outside the historical monitoring period (i.e after October 15th of each year) and consequently suffers from truncated and incomplete counts. Previous to the work here, Holtby (2002) estimated missing run days by using data from years that had complete counts to estimate the "missing proportion". This approach is limited as it does not make use of anxillary information that may be useful to account for within-year run timing variability.

Here, I employ two approaches to obtain robust estimates of escapement for the Babine Coho salmon fish passage and compare those results to Holtby (2002; need ref). First, similar to Walsworth and Schindler (2015) I employ a Bayesian hierarchical model that uses environmental co-variates to obtain estimates of total run size for Coho at the Babine fishway from 1950 through 2021. Second, if time allows, I will employ a novel hierarchical Generalized Additive Model (GAM) that also uses environmental data to obtain coho passage estimates. I compare and contrast these three methods.

2 Data Description

The data used here spans from 1946 through 2021.

The earliest day in data used in this projects begins in 07-19 and complete counts are assumed to be 1950, 1952, 1953, 1957, 1976, 1977, 1979, 1985, 1989, 1991, 1994, 1995, 1996, 1997, 1998, 1999, 2021.

3 Modelling Approach

Here we will be evaluating and comparing three methods of total escapement estimation. First, is using historical correction method

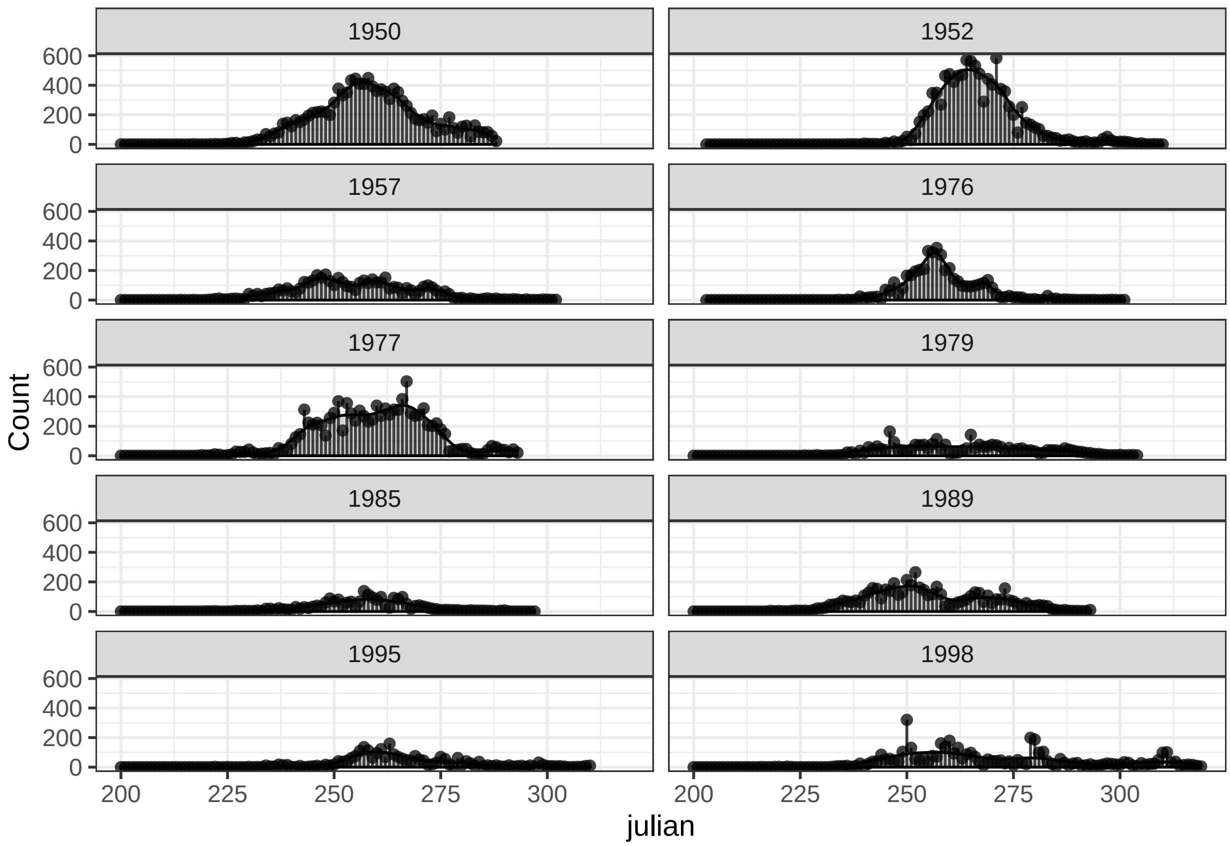

Figure 1: Base years by GAM

See Figure 1.

4 Methods

4.1 Bayesian Hierarchical model

Here, I will be using the migration timing model as described in Walsworth and Schindler (2015). This approach assumes a unimodal distribution of migrations timing and is decribed by the following equations. The following is to fit the model to a single year:

$$E_i = re^{[\frac{-(d_i-p)^2}{\sigma^2}]}/\psi \tag{1}$$

where E_i is the expected daily coho count at the fence on day i; d_i is the numeric day that count was recorded (i); r is the total escapement for that year; p is the day in which peak escapement occured; \$*sigma* is the standard deviation in when the peak run day was observed; and *psi* is a normalizing constant so that:

$$\sum_{i=1}^{n}\{e^{[\frac{-(d_i-p)^2}{\sigma^2}]}\}/\psi = 1 \tag{2}$$

Ultimately, the normarlizing constant ensures that the total escapement parameter r is proportionally allocated to each day of the run.

5 References

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Stan Development Team. 2021. RStan: The R interface to Stan.

Walsworth, T. E., and D. E. Schindler. 2015. Coho salmon escapement and trends in migration timing to a data-poor river: Estimates from a bayesian hierarchical model. Canadian Journal of Fisheries and Aquatic Sciences 72(12):1807–1816.

Appendix A: STAN

Stan is a Bayesian modelling and programming language that can be called from R (R Core Team 2021) via the rstan package (Stan Development Team 2021). Following is the model code

5.1 Appendix A-1: Naive model

1. Naive model: simulated data
2. Naive model: single year
3. Naive model: multiple years
4. Naive model: Hierarchical
5. Informed model: hierarchical.

5.1.1 Naive model: simulated data

First as a proof of concept I generate simulated data and fit the model to ensure I can recapture the parameters. The starting values are as follows:

```
data {
  int N;         // Number of observations (101)
  int y[N];      // Vector of observations
  vector[N] x;  // Vector of DOY
}

// The parameters we are going to estimate in our model
parameters {
  real<lower=0> p;  // day of peak escapement
```

```
  real log_r;       // total escapement
  real<lower=0, upper=15> sigma; //standard deviation in arrival timing
  real<lower=0> reciprocal_phi;  // over dispersion parameter for the negative binomial
}

transformed parameters{
  real r=exp(log_r);
  real phi;
  vector[N] log_phi2;
  vector[N] c_hat;     // expected values of the model
  phi = 1. / reciprocal_phi;
//// These are vectorized
  log_phi2 = -square(x-p) / (2*square(sigma));
  c_hat = log_r + log_phi2 - log_sum_exp(log_phi2);
}

 model {
  //Priors
  reciprocal_phi ~ cauchy(0., 3);
  log_r ~ uniform(1,20);
  sigma ~ cauchy(0., 3);
  p ~ uniform(50, 330);
  // MODEL
  y ~ neg_binomial_2_log(c_hat, phi);
}

generated quantities {
  vector[N] mu;
  //vector[N] log_lik;
  vector[N] y_rep;
  mu = exp(c_hat);
for (i in 1:N) {
  //log_lik[i] = neg_binomial_2_log_lpmf(y[i] | c_hat[i], phi);
  y_rep[i] = neg_binomial_2_log_rng(c_hat[i], phi);
  }
}
```

5.2 Appendix A-2: Final model that includes environmental covariates

Here, I will give a brief example of how stan works in R by demonstrating how to run a bayesisn generalized linear model with normal error structure (equivalent to a least-squares regression). https://mc-stan.org/docs/2_18/stan-users-guide/hierarchical-logistic-regression.html https://mc-stan.org/cmdstanr/articles/r-markdown.html

Appendix B: Comparison with Reference

Appendix C: Generalized Additive Modelling (GAM)

Here I use the

5.3 Appendix C-1: Naive model

First is the Naive single model

5.4 Appendix C-2: Final model that includes environmental covariates

The advantage of a hierarchical structure is that it allows the data from all year to inform each other to ideally obtain better estimates for each years parameters. Hyper priors are quite important particularly on sigma as(see 53 min of video - use half cauchy for sigma)

https://peerj.com/articles/6876/ https://en.wikipedia.org/wiki/Generalized_additive_model