Models in Ecology

Lab 3

Maria Kuruvilla

Sep 2025

Modelling background and equations

In this lab we use mixed effects models, containing both fixed and random effects, to assess the impacts of sea lice on salmon returns. The approach we will take follows the class reading of Krkošek and Hilborn (2011). However, there are some differences. We will have an updated data set, which include later years with lower sea lice levels. Also, so as to keep things manageable we will only fit a subset of the models given in Krkošek and Hilborn (2011).

We will draw on the theory of Lecture 10 of the BIOL 470/550B class (Mixed Effects Models). Please review these notes. Pay particular attention to the pages on "Which random effects should we include?", "Testing random effects structure" and "Crossed vs. nested random effects". They will all be helpful here.

Please also have the Krkošek and Hilborn (2011) handy as we will be referring to it throughout the lab. Finally, it is easier to read the equations in this lab if you download the file to your computer and read it using Word.

Background: Fitting mixed-effects models in Stan

We will use the rstan package to fit a Bayesian mixed-effects (or hierarchical) model in R. You will need the following: - Recent R ($> 4.2.0$) - Recent Rstudio (or text editor of your choice – but we will be using Rstudio!)

It is recommended that your code for this lab and class is within one folder which is also an Rproject Why Use Projects?

Within R you should install the following packages: - Tidyverse installed in R (we'll mostly use dplyr and ggplot2) - Rstan NOTE: When installing rstan, use the following instructions

Please go through the above instructions for your operating system. Make sure to run the example! Here is a minimal workflow example:

```
install.packages("rstan", repos = c('https://stan-dev.r-universe.dev', getOption("repos")))
example(stan_model, package = "rstan", run.dontrun = TRUE)
```

Here are the list of packages we will be using for this lab. YOu can use the following code to check whether you have already installed these packages and to install them, if not. Then load the packages to use them.

```
list.of.packages <- c("ggplot2", "tidyverse", "here", "tidybayes")

new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
```

```r
if(length(new.packages)) install.packages(new.packages)

#load packages
library(rstan)
rstan_options("auto_write" = TRUE)
options(mc.cores = parallel::detectCores())
library(tidybayes)
library(here)
library(tidyverse)
library(ggplot2)
```

Mixed-effects models for salmon populations

Read Krkosek and Hilborn (2011) (henceforth referred to as KH), focusing on the section Stock recruit model, in which they develop a hierarchical version of the Ricker model that includes random effects to account for spatial and temporal co-variation among populations. The full model they present (equation 7) for the log ratio of recruits to spawners is:

$$ln(R_{river,t}/S_{river,t}) = (\alpha_{river} + \beta_{exposurecategory} + \theta_t + \delta_{t,area}) - b_{river}S_{river,t} + \epsilon_{river,t}$$

where $R_{river,t}$ is the number of recruits from population $river$ in year t, $S_{river,t}$ is the number of spawners, α_{river} is a random intercept for each population, $\beta_{exposurecategory}$ is a fixed effect for the exposure category (exposed/unexposed and time period), θ_t is a random year effect common to all populations in year t, $\delta_{t,area}$ is a random effect for year t and area, b_{river} is a random slope for each population, and $\epsilon_{river,t}$ is the residual error.

- the index river indicates which of the up to 128 pink salmon populations (64 Rivers x 2 options for Odd_Even years) is being considered,
- the index exposure category enumerates the exposure category (in KH there are 6 and in our data set there are 8),
- the index t gives the return year (Return_Yr) and
- the index area describes which of the fisheries management areas (Area) the population is found in.

Note that equation (1) indicates that the population dynamics are analyzed at the level of the individual populations: the index "river" appears in each of the variables. Equation (1) indicates that the analysis will occur at the level of each population. The spawners and recruits are measured for each population: the index "river" appears in the formulas. The density dependence b_river will also be different for each population, as will the noise term $\epsilon_{river,t}$. What we are really interested in from an ecological perspective is how the growth rate r is structured across exposure categories, indexed by exposure category, (rather than across populations, indexed by river). However, to deal with the hierarchy of different factors effecting the growth rate, random effects at the levels of return years (indexed by t), populations (indexed by river) and management areas (indexed by a) are also included in the growth rate analysis. The six exposure classes in KH were given by 3 time periods relative to when infestations occurred (before, during, fallow) x two exposure to sea lice options (exposed and unexposed to sea lice from farmed salmon). Here exposure to the sea lice came from populations being in management area 12 where the fish farms were present. However, the data we will use for this lab come from a newer dataset that includes return years to 2009. Therefore, we will include a fourth time period, "after" infestations (see Peacock et al. (2013) for more on this, where it is shown that changes in farm management greatly reduced the impact of sea lice in later years). Thus, the eight exposure classes we have are unexposed-before, unexposed-fallow, unexposed-during, unexposed-after, exposed-before, exposed-fallow, exposed-during, exposed-after. Of these 8 classes, only one of them (exposed-during) give both the location and time interval where the wild salmon returns were hypothesized

by KH to be influenced by the sea lice on salmon farms that juveniles were exposed to on their out-migration to the ocean.

Download the spawner-recruit "Pink S-R data" as a .csv file. Read those data into R. I have downloaded into a folder called "lab3".

The Return_Yr is the year that the Recruits came back to spawn, meaning that the Spawners column refers to the spawners that started in year Return_Yr-2. Area refers to the Department of Fisheries and Oceans statistical management area, which is the spatial scale at which catch data are reported (see http://www.pac.dfo-mpo.gc.ca/fm-gp/maps-cartes/areas-secteurs/index-eng.html). Note that because odd and even year returns do not mix and are treated as separate populations, a population is given by a combination of River and Odd_Even. Create a column for population (by combining river and Odd_Even). Then calculate survival as log(Recruits/Spawners). Create a column for exposure category (by combining exposure and time_period). Finally, create a column for time series length, start row and end row for each population. We will need this to differentiate the data from different populations in stan.

Now use the unique and the length functions in R to check how many return years, management areas and populations there are in the data frame.

Now you should have all the data you need to fit the models described in KH on page 21: all possible combinations of random effects, each with and without the growth rate structured by exposure category. However, in this lab we focus on four categories of the random effects models in KH, which we call models 1 (no random effects), 3 (random effect for intercept on Return_Yr), 4 (random effect for intercept on Area), and 7 (random effect for intercept on Area nested within Return_Yr). These are given in rows 1, 3, 4, and 7 of KH Table 2. For each of these models we are interested in the effect of the exposure category on the population growth rate. Thus, we fit each of the models with no exposure categories (model a) and with the 8 exposure categories given above (model b). Note that the population growth rate alpha is the intercept term in equation (1).

You can visualize the data using the following code. Since there are many populations, we can group the plots by area.

In summary we will have models - 1a: no random effect, no exposure categories, - 1b: no random effects, exposure categories, - 3a: random effect for intercept on Return_Yr, no exposure categories, - 3b random effect for intercept on Return_Yr, exposure categories, - 4a random effect for intercept on Area, no exposure categories, - 4b: random effect for intercept on Area, exposure categories, - 7a: random effect for intercept on Area nested within Return_Yr, no exposure categories, and - 7b: random effect for intercept on Area nested within Return_Yr, exposure categories.