Ecological Models and Data
Block 3: Jul 10, 2023 - Jul 28, 2023

Instructor(s): Dr. Kyle Wilson, Central Coast Indigenous Resource Alliance (e: klwilson.ccira@gmail.com)

Teaching Assistant: Hannah Watkins, Simon Fraser University (e: hwatkins@sfu.ca)

Overview: Ecology is often applied and quantitative in practice. We observe nature, taking data across time and space, and then want to infer whether these observations are consistent with some hypothesized pattern or process. Unfortunately, variation is a pervasive feature of nature and it can confound our ability to make meaningful inferences of the world around us. Enter statistics – which provides important tools that help us to surmount these challenges.

This course provides a bridge between introductory statistics and becoming an ecological practitioner by learning to confront models with real-world data. Students will learn basic and applied skills in this course including: conceptualizing models, translating concepts into code in R or Stan, fitting models to data via probability distributions, generating predictions, and comparing models. This course will go over likelihood and Bayesian approaches throughout the course and compare these two paradigms. Students will learn (or relearn) statistical approaches including linear, generalized linear, nonlinear, hierarchical, and integrated models. In each case, we will try to connect these lessons to local ecosystems and real-world ecological data. This course will allow students to address their own statistical questions through a course project offering opportunities to collect their own field data in local ecosystems, tackle their own pre-existing datasets, or using model simulation to generate data and understanding key concepts in statistics.

Research Skills: Students will improve on their abilities to translate ecological questions into quantitative models and confront models with data – each of which are key to modern ecological practice.

Practical Skills: Ecology has increasingly become a quantitative, data-rich, and applied science. Students will become comfortable handling and manipulating large datasets to answer questions of interest and generate quantitative information through reproducible analyses, which are important skills for careers in academic, government, Indigenous, ENGO, or private organizations (even outside of ecology). This course also develops skills in science communication through presentation of data and models in written, visual, and oral form.

Boat Use: You will be given the opportunity to drive boats if you choose to do so. Boat driving is recommended but not required for this course. Students may wish to drive boats so they can collect data by boat. Students who wish to drive boats at BMSC must hold a PCOC and valid first aid certificate and will participate in an introductory boat check-out on the first day of orientation.

Prerequisites: I recommend that students have some comfort with one of the following: (1) introductory statistics, (2) introductory calculus, and/or (3) introductory ecology – however, students lacking any of these pre-requisites can get permission to take this course but should be prepared to do some additional legwork. All students should be willing to learn some statistical theory and apply it with programming and coding, primarily in R and Stan.

Physical Requirements: All students should be comfortable working long hours at their computer. Additionally, we will have opportunities to be outdoors for short lessons or field work on shoreline or from boats. For those interested in this, students should be comfortable on open boats, walking rugged shorelines in all types of weather. Field work is not mandatory, however, and accommodations will be made for students with any disabilities or other accessibility concerns. Students should note that Bamfield Marine Science Center is a remote and rural area and, as such, access to medical services can be somewhat limited (for physical or mental wellbeing).

Textbook: Statistical Rethinking: A Bayesian Course with Examples in R and Stan, 2nd edition. Richard McElreath (2020). I recommend getting a hard copy of the 2nd edition (should be available from university libraries or available for purchase). Alternatively, however, there are online .pdfs available of the slides and lectures and a less expensive 1st edition (I will have an extra copy of the 1st edition available for students).

Class structure:
There are three weeks of classes. The first two weeks will involve lectures, labs, and student-led group discussions. Starting at the end of the first week, students should be developing their independent projects, and the last week will be entirely devoted to completing these projects and writing a report.

Lecture topics:
Lectures will follow material in the text, with some deviations. A detailed course schedule can be found below, although topics may get rearranged according to needs. There will be approximately 2-3 hours of lecture per day. Topics will loosely follow Statistical Rethinking chapters with the below topics:

1. Course outline & introduction to statistical models & inference
2. Likelihood (Frequentist) v. Bayesian (MCMC)
3. Linear models & causal inference
4. Multiple linear regression & model comparisons
5. Colliders, confounds, and overfitting
6. GLM (Binomial, Poisson)
7. GLM II (Categorical, Negative Binomial, etc.)
8. Hierarchical models (LMM/GLMM)
9. GLMM (and more!)
10. Custom models: nonlinear and mechanistic models
11. Simulation and study design
12. Time-series, state space, forecasting

Labs:
There will be six labs on analysing ecological data using R and Stan that will be graded. The labs will require students to work in the statistical program R (and sometimes Stan) and will involve coding. Some of these topics may be quite challenging for you (especially in a time crunch), and I encourage working, coding, and problem solving together in groups. You must submit your own answers and .R script file. Labs are due midnight the following day after they are assigned. The lab topics are:

1. Introduction to R: data manipulation, functions, loops (install Stan)
2. Optimization, likelihoods, and grid search
3. Linear regression and MCMC sampling
4. LM and GLM in Stan and R

5. Hierarchical models: under the hood
6. GLMM in Stan: model comparisons and generated quantities

Optional: If there is time, I may give an additional lab practical that covers time-series modelling and forecasting (with ***no*** homework assigned)

Discussion groups:

Student groups will have to lead a class discussion based on an assigned reading. Graduate students are expected to be the primary leaders among each group, and groups will submit 4-6 discussion questions based on the reading the day before the discussion is to take place. Student leaders are to give a brief intro to the discussion paper, and then moderate the discussion.

Project:

The course project is designed for students to practice ideas learned in class towards confronting models with data applied to a new problem or case study. Projects are to be done in groups (you can choose your own partners), but each student will hand-in their own report. Students can conduct experiments or collect data for their project during the course, use freely available data, or come with their own datasets. The project proposal is due July 17 and the final project is due July 27 by midnight. Students will present their project in an open symposium at the end of the course. Examples of previous student reports can be obtained from the BMSC library.

Grading:

A percentage will be calculated based on:

- leading class discussion: 15%
- lab assignments: 40%
- project: 35% (of which the proposal is 5% and presentation is 5%)
- other discussion/participation: 10%.

The percentage will then be translated to a course grade by your home university. I will provide a detailed rubric outlining the grading scheme for discussions, project reports, and presentations.

Course website:

Coding and other materials will be posted here: https://github.com/klwilson23/bmsc-emd
Lecture slides will be posted: TBD (BMSC often has a preferred course website)

Academic integrity:

Students are expected to be familiar with University standards regarding academic honesty and to uphold the policies of the University (and relevant policies governing Bamfield Marine Science Center) in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behaviour and avoid any behaviour which could potentially result in suspicions of cheating, plagiarism, misrepresentation of facts and/or participation in offense. Academic dishonesty is a serious offense and can result in suspension or expulsion from the University.

Course schedule:

Date	Topic	Lab	Due dates	Discussions
10-Jul	Course outline & introduction to statistical models & inference	Introduction to R: data manipulation, functions, loops (install Stan)		
11-Jul	Likelihoods and Bayesian probability		Lab 1 due	Discussion 1
12-Jul	Linear models & inference & Multiple linear regression & model comparisons (AIC, LOO, WAIC)	Optimization, likelihood profiles, and grid search: quantifying uncertainty		
13-Jul	**Field lecture:** Colliders, confounds, and overfitting		Lab 2 due	Discussion 2
14-Jul	GLM (Binomial, Poisson) & GLM II (Categorical, Negative Binomial, etc.)	Fitting models to data: linear regression and MCMC sampling		
15-Jul	***off***		Lab 3 due	
16-Jul	***Optional hike***			
17-Jul	LMM/GLMM	LM and GLM in Stan		
18-Jul	GLMM and more		Lab 4 due	Discussion 3
19-Jul	Simulation	Hierarchical models: under the hood		
20-Jul	Nonlinear & mechanistic models		Lab 5 due	Discussion 4
21-Jul	Time-series, state space, forecasting	GLMM in Stan: model comparisons and generated quantities		
22-Jul	***off***		Lab 6 due	
23-Jul	***Optional hike***			
24-Jul	Project			Discussion 5
25-Jul	Project			
26-Jul	Project	Optional: time-series modelling and forecasting		
27-Jul	Project			
28-Jul	Final project due: and presentations			