ESS 575, Models for Ecological Data

Tom Hobbs, B227 NESB, 491-5738 (o), 202-415-5374 (mobile), email: tom.hobbs@colostate.edu

January 12, 2017

Lecture in NESB A302, Tues. and Thurs., 9:30-10:50. Lab in NESB B302 Wed., 1:00-3:00.

Rationale: Virtually all progress in science requires using models to gain insight from data. This course is about gaining understanding of ecological systems using mathematics, statistics, and observations. The ultimate goal of the course is to master the fundamental principles needed for analysis of a broad range of problems in ecological research undaunted by their idiosyncrasies. The emphasis on basic principles comes from my strongly held belief that what you know is not as important as what you are capable of learning. A principles-based approach assures you will be able to continue learning new, quantitative approaches to research throughout your career.

Target audience: Graduate students and advanced undergraduates

Objectives:

1. Learn modern methods for gaining insight about ecological processes using deterministic models, probability models, and data.

2. Build a foundation of principles needed to support an intuitive, flexible approach to analysis and to foster life-long self-teaching.

3. Provide conceptual grounding needed for effective collaboration with statisticians.

4. Develop the quantitative confidence needed to use mathematical and statistical models in your research.

Learning Outcomes:

1. Understand basic principles of probability and statistical distributions needed to link deterministic models to data.

2. Explain maximum likelihood methods for estimating parameters in ecological models.

3. Explain key principles of Bayesian statistics. Understand the relationship between inference accomplished by maximum likelihood and by applying Bayes' theorem.

4. Be able to diagram, write, and implement hierarchical models appropriate for diverse problems in ecological research.

5. Explain how Markov chain Monte Carlo (MCMC) methods can be used to approximate marginal posterior distributions. Write MCMC algorithms and computer code in R implementing MCMC methods for simple Bayesian models.

6. Use software for implementing MCMC methods (i.e., JAGS, R packages) to approximate marginal posterior distributions of parameters, latent variables, and derived quantities of interest. Be able to evaluate convergence.

7. Understand procedures for model checking and model selection in the Bayesian framework.

Prerequisites: Ideally, students should have a course in calculus, a basic ecology course, and an introduction statistics, particularly mathematical statistics. None of these are absolute requirements; I will review key background concepts as part of the lectures. However, that said, if you don't have at least two of these background courses, you should be prepared to do some remedial work on your own.

Content and teaching approach: My job as a teacher is to accelerate your mastery of material that you would learn slowly or perhaps not at all without my help. I do this by offering a sequence of lectures and closely linked problem sets. I will usually lecture twice a week, but there may be weeks with one lecture and two labs to provide more time to work on problems. It is imperative that students keep up with problem sets, which form the foundation of the course.

Texts: There is a required text, Hobbs, N. T. and M. B. Hooten, *Bayesian models: A statistical primer for ecologists*, Princeton University Press 2015. The book emerged from the course after a decade of teaching it. Many students have told me it offers a great compliment to the lectures and labs because the sequence of ideas is the same. That said, I will not "lecture from the book."

Readings: I will include a few other readings on the class GitHub. They are not mandatory. Most are included because I found them unusually interesting or amusing.

Working in groups: You will be assigned to a lab group including two other colleagues. I feel strongly that your success in science depends on your ability to work effectively with others. Moreover, a team approach to work in the laboratories allows you to teach each other as well as to learn from the teaching assistant and me. It will lighten the work load by allowing you to share tasks. It is more fun.

An individual project: An individual project will be due at the end of finals week. You will write a Bayesian hierarchical model for a problem of your own choosing, hopefully related to your research. A brief write-up will describe the ecological questions addressed, the data, and the model. I will provide more details about what is expected for this project later in the semester.

Exams: None.

Grading: If you complete the assigned work with attention and care, you will get an A in this course. I am far more interested in your mastery of the material than I am in making academic comparisons among you. The material in this course can appear intimidating at first, but the last thing I want is for you to be anxious about it. Everyone who has taken this course has emerged with a sturdy understanding of the key concepts and methods. It may seem daunting at first. Relax. We will get through it. You wouldn't be here if you didn't want to learn.

Seventy-five percent of your grade will be based on 10 or 11 lab write-ups, each worth 50 points / week required to complete. So, if I allocate 2 weeks to a lab topic it is worth 100 points, etc. The remaining 25% of your grade will be based on the individual project, described above. Lab work will require programming in R and JAGS as well as some work with paper and a sharp pencil. For each assignment, each lab group will turn in a *single* electronic copy of a write-up that includes text, figures, and tables communicating your results. Grading will be based on the following:

1. Quality of approach to problem: Did you use a logical, thoughtful process for solving the problem?
2. Quality of presentation: Did you present your findings in a literate document? Did you clearly communicate how you solved the problem, showing mathematical steps or a computer algorithm? Was your document attractive and well organized? Did you use proper notation for mathematics and statistics?
3. Quality of technique: Did you demonstrate mastery of the appropriate methods? Your lab reports should describe model results and discuss them as appropriate.

Preparing reports All lab write-ups and your the report of your individual project must be prepared in R markdown. All equations must be typeset in LaTeX using proper notation. There are files in the `Admin` folder on the class GitHub describing what you need to know to produce stunning lab reports.

Turning in reports Turn in all assignments the appropriate folder in the course Dropbox, ESS575. You have an invitation to use that drive. Put your last name in the title of the file; for group work, all members of the group.

Things you need: A large amount of computer programming will be necessary to successfully complete the course, so students will need easy access to workstations running R, JAGS[1]and the R Studio editor, all of which are free, open-source software. It will be very helpful if you have your own laptop. You should always bring it to lab and you will occasionally need it in class. It will also be useful to bring an old fashioned notebook or tablet to class. There are topics that I feel are best presented at the board. I do this not to avoid preparing handouts, which I will likely do anyway, but instead to make the lecture more intimate and interactive, to make your learning a bit more active, and to slow me down. You will need a GitHub account to access course materials. If you have an account already, good for you. If you don't, please go to `https://github.com` and sign up. It is free.

Accommodation of individual learning needs: If you have learning needs that may affect your performance (sight, hearing, language, or any other reason), please let me know at the beginning of the course. We will work out ways to make the class work for you. I am deaf as a post in my right ear, so you may need to accommodate me as well.

Interaction outside of class: My office hours will 11:00-12:00 on Tuesdays and Thursdays and by appointment. Appointments can be informal–if you stick your head in my door, I often will be able to help you. The TA and I will be available via email to answer questions on your R programming. When you have R questions, be sure to include the script causing you

[1] We will load JAGS later in the semester. You will need R, R studio, and LaTeX loaded for the first lab. A handout will be provided to help you with installation.

problems in your email including *everything* we need to run your code. Your learning will be meaningfully enhanced if you struggle with a problem before you ask a question, but we don't want you to struggle excessively. If you truly aren't getting anywhere on your own, don't spin your wheels. Contact us with a well-framed question.

Teaching assistant You are really lucky that Megan Vahsen has volunteered to help me teach the laboratories. She has taken ESS 575 as well as Mevin Hooten's Bayesian statistics class. She is a knowledgable R and JAGS programmer. She will be a terrific resource for you. Her contact information is mlvahsen@gmail.com. Her office is Plant Sciences C-033. Office hours are Thursday's 2:00 - 4:00 or by appointment.

Class notes: Notes for each lecture will be available as .pdf files on the class GitHub. The updated, current version of the notes will be available no later that 9:00 the day of lecture. I revise every lecture I give, so you would be wise to print the notes the morning of the lecture if you want a hard copy.

My travel: Although I avoid travel during the spring semester, there may be occasions when I must be away. I will provide an alternative learning activity during the rare days when I miss class. Megan will conduct laboratories when I am traveling.

Approximate schedule

Week	Lecture	Reading[1]	Lab
16-Jan	Introduction to class Deterministic models	Chapters 1,2	Learning R
23-Jan	Stochastic models (and what sets Bayes apart) Laws of probability	Chapter 3	Learning R
30-Jan	Statistical distributions Moment matching	Chapter 3	Probability, statistical distributions, moment matching
6-Feb	Likelihood Bayes theorem	Chapters 4,5	Light limitation of trees
13-Feb	Priors I Introduction to hierarchical models	Chapters 5,6	The components of Bayesian models
20-Feb	Practice writing hierarchical models, no lecture this week	Chapters 10, 11	Practice writing hierarchical models
27-Feb	Markov chain Monte Carlo Markov chain Monte Carlo	Chapter 7	Coding MCMC in R, inference from a single model
6-Mar	Review answers to hierarchical modeling problems Bayesian regression		JAGS and rjags
13-Mar	Spring Break		
20-Mar	Bayesian analysis of designed experiments Multi-level models	Chapter 6.2.2	JAGS and rjags
27-Mar	Priors II Model checking	Chapter 5, Chapter 8.1	Nitrous oxide emissions
3-Apr	Model selection Mixture models, zero inflation, and occupancy	Chapter 9	Nitrous oxide emissions cont.
10-Apr	Priors III Dynamic models	Chapter 8.5	Landscape ecology of Swiss birds
17-Apr	Forecasting Modeling temporal dependence	TBA	Harvest of lynx in Sweden
24-Apr	Modeling spatial dependence Modeling spatial dependence		Harvest of lynx in Sweden cont.
1-May	Buffer		Genetic diversity and plant colonization success
8-May	Work on individual problem		Genetic diversity and plant colonization success cont.

[1]Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, N.J.