ESS 575, Models for Ecological Data

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

January 9, 2019

Lecture in NESB A302, Tues. and Thurs. 9:30-10:50. Lab in NATRS 254, Wed. 11:00-1:00

Rationale: Virtually all progress in science requires using models to gain insight from data. This course is about enhancing 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 understanding, skill, and confidence needed to use mathematical and statistical models in your research.

Learning Outcomes:

1. Understand basic principles of probability and statistical distributions needed to fit 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 of unobserved quantities in Bayesian models. 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, missing data, 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 had a course in calculus, a basic ecology course, and an introduction to statistics, particularly mathematical statistics. None of these is an absolute requirement; 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. The definite integral is the only concept from calculus that we will use frequently. You need to understand this type of integration, but you do not need to be able to perform it.

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.

Discovery: I plan to add several "discovery" exercises this year in keeping with the central goal of the course: to enable you to learn on your own. A discovery exercise will be a challenge that I have not specifically covered in class. You will know the relevant underpinning principles, but you will not have seen the problem solved. You may use any resource (web, books, colleague, etc.) you like to attack the problem. There is no more important skill in science than being able to "figure it out." The quantitatively less adept among us (myself included) may take longer to figure it out than the quantitatively blessed, but our dogged persistence and a principle based approach will allow us to prevail, every time.

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 much the same. Rest assured, however, I will not "lecture from the book." There will be assigned readings from the text. You would be foolish to forgo them. There are few errors in the book[1].Find these described at `https://www.stat.colostate.edu/~hooten//papers/pdf/Hobbs_Hooten_Bayesian_Models_2015_errata.pdf`

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

Working in groups: You will be assigned to a lab group including two or three other colleagues. I feel strongly that your success in science depends on your ability to work effectively with others, so we will practice that critical skill here. 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. Course evaluations from past students were virtually unanimous in their enthusiastic support of group work.

[1] Astonishingly few for an amusing reason I will tell you about.

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 in a handout in the Admin folder of the class repository. I encourage you to think about this problem sooner rather than later. You will probably know most of the material you need to know to get started on your project soon after the spring recess.

Exams: There will be a take-home, open-book "challenge" during the week before spring break, emphasizing some great discovery problems. You may call on any resource you think would be useful in composing your answers – printed material, web searches, conversations with colleagues, divine intervention. I wouldn't think of doing real work without using these; why not practice the way you will work in the future? I call this a challenge rather than an exam because the exercise will ask you to apply what you have learned to novel problems rather than require you to regurgitate what I have told you. The challenge serves two primary purposes. I believe you learn more effectively if you know you are going to be held to account. More importantly, the challenge will allow me to discover areas where my teaching may not have been as successful as I would have liked, motivating targeted review.

Grading: You will get an A in this course if you complete the assigned work with attention and care. 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 and I wouldn't be here if I didn't want to teach.

Sixty-five percent of your grade will be based on 11 lab write-ups. Fifteen percent will come from the mid-term exam. The remaining 20% of your grade will be based on the individual project, described above and in more detail as the course proceeds. 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 including the underpinning math as needed? Was your document attractive and well organized? Did you use proper notation for mathematics and statistics? Was your writing clear, succinct, grammatically correct, and professional?
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. There are files in the `Admin` folder on the class GitHub describing what you need to know to produce stunning lab reports. All equations must be typeset in LaTeX. The R markdown handout contains a section on proper statistical notation I expect you to follow it. There will also be a handout on a few key elements of clear scientific writing. You will suffer much red ink if you fail to pay attention to that handout.

Turning in reports Lab reports are due the before the start of a new lab problem. So, for example, the R lab report (see schedule below) is due no later than 11 AM, Wednesday, Feb. 5. Turn in all assignments the appropriate folder in the course Dropbox, ESS575. You have received an invitation to use that drive in an email. Place an .html and .Rmd file in a folder named using the format `Lab_2_groupname` where `groupname` is your group's name.

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[2]and the R Studio editor, all of which are free, open-source software. A laptop computer will be necessary. 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 and a sharp pencil to class and lab. There are topics that I feel are best presented at the board. I illustrate points on the board 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 be 11:00-1:00 on Tuesdays, 1:00 - 3:00 on Wednesdays, and by appointment. Appointments can be informal – if you stick your head in my office, 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 your 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 for hours. Contact us with a well-framed question.

Teaching assistant You are fortunate that Brian Avila has volunteered to help me teach the laboratories. He has taken ESS 575 as well as Mevin Hooten's Bayesian statistics class. He is a knowledgable R and JAGS programmer. He will be a terrific resource for you. His contact information is `bavila@rams.colostate.edu`. His office is Wagar 203. Office hours are Friday 12:30 - 2:30 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. I occasionally forget to push the new notes to the repository. Please remind me by email if there are no new notes after 9 AM.

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

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

Approximate schedule

Week	Lecture	Reading[1]	Lab
22-Jan	Introduction to class What sets Bayes apart?	Preface, Chapter 1	Learning R
29-Jan	Deterministic models Laws of probability	Chapter 2, 3	Learning R
5-Feb	Distribution theory Moment matching	Chapter 3	Probability, statistical distributions, moment matching
12-Feb	Likelihood Bayes' theorem	Chapters 4, 5.1 - 5.2	Light limitation of trees
19-Feb	Priors and conjugate priors Markov chain Monte Carlo I	Chapters 5.3 - 5.4, 7	The components of Bayesian models
26-Feb	Markov chain Monte Carlo II Simple Bayesian models illustrated with regression	Chapter 7, 8	Coding MCMC in R, inference from a single model
5-Mar	Mid-term challenge		No lab, probably no lecture this week
12-Mar	Spring Break		
19-Mar	Debrief challenge Bayesian multi-level models models and more about priors	Chapter 6	JAGS and rjags
26-Mar	Bayesian analysis of designed experiments Hierarchical models and problem set	Chapter 6	JAGS, rjags, and data simulation
2-Apr	Debrief hierarchical models problem set Model checking and what to do when checks fail	Chapter 8.1 Chapter 9	Nitrous oxide emissions
9-Apr	Model selection Mixture models, zero inflation, occupancy, more about priors	Chapter 10.2	Nitrous oxide emissions
16-Apr	Dynamic models Forecasting	Chapter 8.5	Landscape occupancy of Swiss birds
23-Apr	Modeling spatial dependence Modeling spatial dependence	TBA	Harvest of lynx in Sweden
30-Apr	Communicating a Bayesian analysis Consultation on individual problem	Afterword	Harvest of lynx in Sweden
7-May	Consultation on individual problem		Modeling spatial dependence
14-May	Work on individual problem		No lab

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