Syllabus

Course Information

- **Course Number**: BEE 4850/5850
- **Course Title**: Simulation Methods for Environmental Data Analysis
- **Credits**: 3

Instructor
- 👤 Vivek Srikrishnan
- ✉ viveks@cornell.edu
- 🏢 318 Riley-Robb

TA
- 👤 TBD
- ✉ TBD
- 🏢 TBD

Meetings
- 📅 MWF
- 🕐 1:25-2:15
- 🏛 TBD

Course Description

Analyzing environmental data often involves iterative Exploratory data analysis plays a critical role in statistical and numerical model development and formulation. However, as environmental data increases in complexity, size, and ubiquity, emphasizing the data-generating process can guide analysis of the dataset and modeling choices. This course will provide an overview of a generative approach to data analysis, which uses simulation and assessments of predictive performance to provide insight into the structure of data and its data-generating process. The goal is to provide students with a framework and an initial toolkit of methods that they can use to formulate and update hypotheses about data which can facilitate principled future uses and inferences. In particular, we will:

- formulate statistical analyses as a method for decision-making under uncertainty;
- interpret graphical and quantitative summaries of counterfactual simulations to test hypotheses and assess the utility of models;
- analyze and account for the structure of model residuals;
- apply Monte Carlo and bootstrap methods to simulate pseudodata from calibrated models;

- assess and select models based on predictive performance; and
- emulate computationally-complex simulations with surrogate models.

Learning Objectives

At the end of this class, students will:

1. Analyze environmental datasets using simulations from calibrated models;
2. Interpret and communicate properties of datasets and models using graphical summaries;
3. Assess model fit and performance by analyzing residual structures and predictive performance;
4. Implement statistical and numerical modeling, calibration, and visualization techniques in modern programming languages.

Prerequisites & Preparation

The following courses/material would be ideal preparation:

- One course in programming (*e.g.* CS 1110, 1112 or ENGRD/CEE 3200)
- One course in probability or statistics (ENGRD 2700, CEE 3040, or equivalent)

In the absence of one or more these prerequisites, you can seek the permission of instructor.

> 💡 What If My Programming or Stats Skills Are Rusty?
>
> If your programming or statistics skills are a little rusty, don't worry! We will review concepts and build skills as needed.

Typical Topics

- Introduction to exploratory data analysis;
- Review of probability and statistics;
- Bayesian decision theory;
- Principles of data visualization;
- Model residuals and discrepancies;
- Censored, truncated, and missing data;
- Monte Carlo and bootstrap methods for calibration (including Markov chain Monte Carlo);
- Predictive model assessment (including cross-validation);
- Model averaging;

- Emulation with surrogate models (including Gaussian processes, polynomial chaos expansions, and machine learning methods)

Course Philosophy and Expectations

The goal of our course is to help you gain competancy and knowledge in the area of data analysis. This involves a dual responsibility on the part of the instructor and the student. As the instructor, my responsibility is to provide you with a structure and opportunity to learn. To this end, I will commit to:

- provide organized and focused lectures, in-class activities, and assignments;
- encourage students to regularly evaluate and provide feedback on the course;
- manage the classroom atmosphere to promote learning;
- schedule sufficient out-of-class contact opportunities, such as office hours;
- allow adequate time for assignment completion;
- make lecture materials, class policies, activities, and assignments accessible to students.

I encourage you to discuss any concerns with me during office hours or through a course communications channel! Please let me know if you do not feel that I am holding up my end of the bargain.

Students can optimize their performance in the course by:

- attending all lectures;
- doing any required preparatory work before class;
- actively participating in online and in-class discussions;
- beginning assignments and other work early;
- and attending office hours as needed.

Community

Diversity and Inclusion

Our goal in this class is to foster an inclusive learning environment and make everyone feel comfortable in the classroom, regardless of social identity, background, and specific learning needs. As engineers, our work touches on many critical aspects of society, and questions of inclusion and social justice cannot be separated from considerations of systems analysis, objective selection, risk analysis, and trade-offs.

In all communications and interactions with each other, members of this class community (students and instructors) are expected to be respectful and inclusive. In this spirit, we ask all participants to:

- share their experiences, values, and beliefs;
- be open to and respectful of the views of others; and
- value each other's opinions and communicate in a respectful manner.

Please let me know if you feel any aspect(s) of class could be made more inclusive. Please also share any preferred name(s) and/or your pronouns with me if you wish: I use he/him/his, and you can refer to me either as Vivek or Prof. Srikrishnan.

> **❗ Please, Be Excellent To Teach Other**
>
> We all make mistakes in our communications with one another, both when speaking and listening. Be mindful of how spoken or written language might be misunderstood, and be aware that, for a variety of reasons, how others perceive your words and actions may not be exactly how you intended them. At the same time, it is also essential that we be respectful and interpret each other's comments and actions in good faith.

Student Accomodations

Let me know if you have any access barriers in this course, whether they relate to course materials, assignments, or communications. If any special accomodations would help you navigate any barriers and improve your chances of success, please exercise your right to those accomodations and reach out to me as early as possible with your Student Disability Services (SDS) accomodation letter. This will ensure that we have enough time to make appropriate arrangements.

> **❗ Important**
>
> If you need more immediate accomodations, but do not yet have a letter, please let me know and then follow up with SDS.

Course Communications

Most course communications will occur via Ed Discussion. Public Ed posts are generally preferred to private posts or emails, as other students can benefit from the discussions. If you would like to discuss something privately, please do reach out through email or a private Ed post (which will only be viewable by you and the course staff).

Announcements will be made on the course website and in Ed. Emergency announcements will also be made on Canvas.

> **❗ Ed Tips**
>
> - If you wait until the day an assignment is due (or even late the previous night) to ask a question on Ed, there is a strong chance that I will not see your post prior to the deadline.
> - But if you see unanswered questions and you have some insight, please answer! This class will work best when we all work together as a community.

Mental Health Resources

We all have to take care of our mental health, just as we would our physical health. As a student, you may experience a range of issues which can negatively impact your mental health. Please do not ignore any of these stressors, or feel like you have to navigate these challenges alone! You are part of a community of students, faculty, and staff, who have a responsibility to look for one another's well-being. If you are struggling with managing your mental health, or if you believe a classmate may be struggling, please reach out to the course instructor, the TA, or, for broader support, please take advantage of Cornell's mental health resources.

> **❗ Mental Health And This Class**
>
> I am not a trained counselor, but I am here to support you in whatever capacity we can. You should never feel that you need to push yourself past your limits to complete any assignment for this class or any other. If we need to make modifications to the course or assignment schedule, you can certainly reach out to me, and all relevant discussions will be kept strictly confidential.

Course Policies

Attendance

Attendance is not *required*, but in general, students who attend class regularly will do better and get more out of the class than students who do not. Your class participation grade will reflect both the quantity and quality of your participation, only some of which can occur asynchronously. I will put as many course materials, such as lecture notes and announcements, as possible online, but viewing materials online is not the same as active participation and engagement. Life happens, of course, and this may lead you to miss class. Let me know if you need any appropriate arrangements ahead of time.

> **!** What If I'm Sick?
>
> Please stay home if you're feeling sick! This is beneficial for both for your own recovery and the health and safety of your classmates. We will also make any necessary arrangements for you to stay on top of the class material and if whatever is going on will negatively impact your grade, for example by causing you to be unable to submit an assignment on time.

Mask Policies

Masks are encouraged but not required in the classroom, per university policy. However, the University *strongly encourages* compliance with requests to mask from students, faculty, and staff who are concerned about the risk of infection. Please be respectful of these concerns and requests if you cannot wear a mask.

> **!** Masks Will Be Required In My Office
>
> I will **require masks to be worn** in my office or during in-person office hours, as we are necessarily interacting in close quarters without great airflow.

Academic Integrity

> **!** Important
>
> **TL;DR**: Don't cheat, copy, or plagiarize!

This class is designed to encourage collaboration, and students are encouraged to discuss their work with other students. However, I expect students to abide by the Cornell University Code of Academic Integrity in all aspects of this class. **All work submitted must represent the students' own work and understanding**, whether individually or as a group (depending on the particulars of the assignment). This includes analyses, code, software runs, and reports. Engineering as a profession relies upon the honesty and integrity of its practitioners (see *e.g.* the American Society for Civil Engineers' Code of Ethics).

External Resources

The collaborative environment in this class **should not be viewed as an invitation for plagiarism**. Plagiarism occurs when a writer intentionally misrepresents another's words or ideas (including code!) as their own without acknowledging the source. **All** external resources

which are consulted while working on an assignment should be referenced, including other students and faculty with whom the assignment is discussed. You will never be penalized for consulting an external source for help and referencing it, but plagiarism will result in a zero for that assignment as well as the potential for your case to be passed on for additional disciplinary action.

AI/ML Resource Policy

As noted, all work submitted for a grade in this course must reflect your own understanding. The use and consulation of AI/ML tools, such as ChatGPT or similar, must be pre-approved and clearly referenced. If approved, you must:

- reference the URL of the service you are using, including the specific date you accessed it;
- provide the exact query or queries used to interact with the tool; and
- report the exact response received.

Failure to attain prior approval or fully reference the interaction, as described above, will be treated as plagiarism and referred to the University accordingly.

Late Work Policy

In general, late work will be subjected to a 10% penalty per day, *which can accumulate to 100% of the total grade.* However, sometimes things come up in life. Please reach out *ahead of time* if you have extenuating circumstances (including University-approved absences or illnesses) which would make it difficult for you to submit your work on time. Work which would be late for appropriate reasons will be given extensions and the late penalty will be waived.

Assessments

Application Exercises: 10%

Most weeks, students will be given a set of exercises (typically involving analyzing a dataset, a model, or a figure, or an overview to a new Julia package) to complete. These will involve a small amount of programming or visual assessment of data or a figure. Application exercises will be provided the previous Friday and are intended to align with the content for the week, and solutions should be submitted by 9:00pm on the Friday at the end of the given week. Application exercise solutions should be submitted as a PDF on Gradescope. The lowest two of these will be dropped. Application exercises which involve programming will be distributed using GitHub Classroom.

Homework Assignments: 50%

Approximately 6 homework assignments will be assigned throughout the semester (roughly one per course module). You will typically have 2 weeks to work on each assignment, though this depends on the module length. Students are encouraged to collaborate and learn from each other on homework assignments, but each student must submit their own solutions reflecting their understanding of the material. Consulting and referencing external resources and your peers is encouraged (engineering is a collaborative discipline!), but plagiarism is a violation of academic integrity.

Some notes on assignment and grading logistics:

- Homework assignments will be distributed using GitHub Classroom. Students should make sure they update their GitHub repositories as they work on the assignments; this helps with answering questions and gives you a backstop in case something goes wrong and you can't submit your assignment on time.
- Homeworks are due by 9:00pm Eastern Time on the designed due date. Your assignment notebook (which include your writeup and codes) should be submitted to Gradescope as a PDF with the answers to each question tagged (a failure to do this will result in deductions).
- Rubrics will be provided for the homeworks as part of the assignments.
- Students in 5850 will be asked to complete additional homework problems which go more deeply into the
- Your lowest homework grade will be dropped. We can discuss arrangements if multiple assignments will be missed for university-approved reasons, preferably ahead of time.
- Regrade requests for specific problems must be made within a week of the grading of that assignment. However, note that regrades can cut both ways: the TA can take away points as well!

Final Term Project: 40%

This course will culminate with a term project. The goal of this project is to apply and extend the tools and approaches we will learn in class. While we encourage drawing on other classes or interests when developing and working on your project, **submitting work from another course or work which was completed prior to the course is not permitted**.

The term project will be completed in small groups (2-3 students) for students enrolled in BEE 4850 and individually for those in BEE 5850. The final deliverable for this project will be a poster summarizing the project and results. Ahead of that, you will submit the following:

- a proposal for feedback on the scope of your project; and
- a mid-semester work plan providing more details on the project design.

Rubrics will be provided for the components of the project.

Tentative Schedule

#	Date	Day	Topic
			Introduction to Exploratory Data Analysis
1	01-22	Mon	Introduction
2	01-24	Wed	Overview of exploratory data analysis
3	01-26	Fri	Github Overview and Work on Homework 1
4	01-29	Mon	Data-generating processes and uncertainty
5	01-31	Wed	What can go wrong with EDA
			Review of Probability and Statistics
6	02-02	Fri	Simulation-based workflows for data analysis
7	02-05	Mon	Probability basics
8	02-07	Wed	Probabilistic models
9	02-09	Fri	Bayes' Theorem
10	02-12	Mon	Loss functions and Bayesian decision theory
11	02-14	Wed	Statistics as decision-making under uncertainty
			Data Visualization and Graphical Checks
12	02-16	Fri	Psychology of memory and vision
13	02-19	Mon	Information channels and Gestalt principles
14	02-21	Wed	Class discussion of figures
15	02-23	Fri	Approaches to visualizing data and uncertainty
	02-26	*Mon*	*February Break*
16	02-28	Wed	Graphical checks for model diagnostics
			Model Calibration
17	03-01	Fri	Understanding model residuals
18	03-04	Mon	Model-data discrepancy
19	03-06	Wed	Missing data
20	03-08	Fri	Censored and truncated data
21	03-11	Mon	Modeling complex residual structures
22	03-13	Wed	Sensitivity analysis
			Simulation Methods
23	03-15	Fri	Introduction to Monte Carlo
24	03-18	Mon	Monte Carlo error analysis
25	03-20	Wed	Monte Carlo simulation: Examples
26	03-22	Fri	Sampling distributions
27	03-25	Mon	The bootstrap
28	03-27	Wed	Using the bootstrap for uncertainty quantification
29	03-29	Fri	Overview of Bayesian computation
	04-01	*Mon*	*Spring Break*
	04-03	*Wed*	*Spring Break*
	04-05	*Fri*	*Spring Break*
30	04-08	Mon	Markov chains

#	Date	Day	Topic
31	04-10	Wed	Markov chain Monte Carlo
32	04-12	Fri	Markov chain Monte Carlo convergence
			Model Selection
33	04-15	Mon	Model assessment via predictive accuracy
34	04-17	Wed	Cross-validation
35	04-19	Fri	Distances between distributions
36	04-22	Mon	Predictive information criteria
37	04-24	Wed	Model averaging and stacking
			Emulation and Surrogate Modeling
38	04-26	Fri	Kriging and variograms
39	04-29	Mon	Gaussian processes
40	05-01	Wed	Polynomial chaos expansion
41	05-03	Fri	Machine learning emulators
42	05-06	Mon	Uncertainty and machine learning models