BEE 4850/5850 (Environmental Data Analysis and Simulation) Syllabus

Spring 2024

This is a 3 credit course offered as an elective.

Instructor

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

Meetings

- 📅 MW
- 🕐 11:40-12:55pm
- 🏛 160 Riley-Robb Hall

Course Description

Simulations from numerical and statistical models are a key tool used to quantify and represent our understanding of many different environmental systems, including but not limited to the climate, sea levels, air pollution, and the electric power system. Data analysis of environmental observations and model simulations is an integral part of developing and evaluating models used to 1) test our assumptions about environmental system dynamics; 2) develop new insights into environmental processes; and 3) project future environmental conditions and outcomes. This course will provide an overview of the use of simulation model development and evaluation in environmental data analysis. 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 and models. Students will actively analyze and use real data from a variety of environmental systems. In particular, over the course of the semester, we will:

- conduct exploratory analyses of environmental datasets;
- discuss best practices for and complexities of data visualization;
- calibrate statistical and process-based numerical models using environmental data;
- use simulations from calibrated models to identify key sources of uncertainty and model error;
- assess model fit and adequacy through predictive ability.

Learning Outcomes

After completing this class, students will be able to:

- create, interpret, and critique data visualizations;
- calibrate environmental models to observations, possibly including censored and missing data;
- simulate alternative datasets from models using statistical methods such as the bootstrap and Monte Carlo;
- assess model adequacy and performance using predictive simulations;
- apply and contextualize model selection criteria;
- evaluate evidence for and against hypotheses about environmental systems using model simulations;
- emulate computationally-complex models with simpler representations.

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;
- Statistical methods for calibration;
- Predictive model assessment;
- Emulation with surrogate models

Course Meetings

This course meets MWF from 11:40–12:55 in 160 Riley-Robb Hall. In addition to the course meetings (a total of 42 lectures, 50 minutes each), the final project will be due during the university finals period. In addition to the work during the semester, students can expect to devote, on average, 6 hours of effort during the exam period.

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 how data are generated, collected, and analyzed.

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.

> **!** 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.

Office Hours

Office hours will be held in 318 Riley-Robb before class on MW, from 10-11AM, in 318 Riley-Robb Hall. Some time will be available after class for brief questions, but this is limited. If these times do not work for you, or you need some additional time outside of office hours, please reach out to Prof. Srikrishnan about scheduling a meeting. Depending on schedules, these requests may not be accepted on short notice (*e.g.* homework is due on Friday and you reach out late on Thursday), but with several days notice we should be able to find a time that will work.

Office hours are intended to help all students who attend. This time is limited, and is best spent on issues that are relevant to as many students as possible. While we will do our best to answer individual questions, students asking us to verify or debug homework solutions will have the lowest priority (but please do ask about how to verify or debug your own solutions!). However, we are happy to discuss conceptual approaches to solving homework problems, which may help to reveal bugs.

Space at office hours can be limited (we may shift to the conference room in 316 Riley-Robb if offices are full and it is available). If the room is crowded and you can find an alternative source of assistance, or if

your question is low priority (*e.g.* debugging) please be kind and make room for others.

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.

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 can be submitted up to 24 hours after the due date at a 50% penalty. 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. Note that *e.g.* job interviews or a busy schedule outside of this course are not valid reasons for extensions. If an extension is granted, any late penalties will be waived up to the extension date. In extreme circumstances, assignments can be forgiven, and your grade will be computed as though those did not occur, giving your other assignments more weight.

Regrade Requests

Regrade requests can be submitted up to one week after the graded work is released on Gradescope.

All regrade requests must include a **brief** justification for the request or they will not be considered. Good justifications include (but are not limited to): - *My answer agrees with the posted solution, but I still lost points. - I lost 4 points for something, but the rubric says it should only be worth 2 points. - You took points off for something, but it's right here. - My answer is correct, even though it does not match the posted solution; here is an explanation. - There is no explanation for my grade. - I got a perfect score, but my solution has a mistake* (you will receive extra credit for this! see below!) - *There is a major error in the posted solution; here is an explanation* (full credit for everyone, but Prof. Srikrishnan will decide what constitutes a "major error"! see below!).

> **❗ We Can Only Grade What You Submitted**
>
> All regrades will be assessed based only on the submitted work. You cannot get a higher grade by explanation what you meant (either in person or online) or by adding information or reasoning to what is submitted after the fact. The goal of the regrade is to draw attention to a potential grading problem, not to supplement the submission.

Once Prof. Srikrishnan issues a final response to a regrade request, further requests for that submission will be ignored.

> **⚠ Regrade Requests Can Be A Gamble!**
>
> While you should submit regrade requests for legitimate errors, using them for fishing expeditions can also result in lost points if Prof. Srikrishnan decide that your initial grade was too lenient or if additional errors are identified.

> **💡 What If I Find A Different Type of Mistake?**
>
> - If you submit a regrade request *correctly* reporting that a problem was graded too leniently — that is, that your score was higher than it should be based on the rubric — your score will be increased by the difference. For example, if your original score on a problem was 8/10 and you successfully argue that your score should have been 3/10, your new score will be 13/10.
> - If a significant error is discovered in a posted homework solution or in the exam solutions, everyone will in the class will receive *full* credit for the (sub)problem. Prof. Srikrishnan will decide what is "significant".

Assessments

Technologies

We will use Canvas as a gradebook, and to distribute PDFs of readings (which also be made available through the website, via the Cornell library). Ed Discussion will be used for course communications. Assignments will be submitted and graded in Gradescope.

Students can use any programming language they like to solve problems, though we will make notebooks and package environments available for Julia (which may help structure your assignments if you use a different language) via GitHub. If students use a language other than Julia, we may limited in the programming assistance we can provide (though we're happy to try to help!).

We recommend students create a GitHub account and use GitHub to version control and share their code throughout the semester.

Grading

Final grades will be computed based on the following assessment weights:

Assessment	Weight
Exercises	10%
Readings	10%
Literature Critique	15%
Homework Assignments	30%
Term Project	35%

The following grading scale will be used to convert the numerical weighted average to letter grades:

Grade	Range
A	94–100
A-	90–94
B+	87–90
B	84–87
B-	80–84
C+	77–80
C	74–77
C-	70–74
D+	67–70
D	64–67
D-	61–64
F	< 61

Exercises

Most weeks, students will be given a set of exercises (typically involving analyzing a dataset, a model, or a figure) to complete. These will involve a small amount of programming, a minor calculation, and/or visual assessment of data or a figure. Exercises will be provided the previous Monday 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. These exercises will be given as quizzes on Gradescope and will be auto-graded. The lowest exercise score will be dropped automatically.

Readings

Readings will be assigned for discussion throughout the semester. Students are expected to summarize their takeaways and thoughts on the reading and respond to others' comments on the Ed forum during the week after the reading is assigned (typically Monday – Monday). Collaborative annotation assignments will be set up in Canvas to facilitate reading ahead of the discussion on the Ed forum. For key readings, a student in BEE 5850 will be assigned to lead a 35 minute in-class discussion on the reading.

Literature Critique

Students will select a peer-reviewed journal article related to an application of data analysis and will write a short discussion paper analyzing the hypotheses and statistical choices. Students enrolled in BEE 5850 will also write a referee report, as if they were a peer reviewer. Students will give an in-class presentation of their paper during one class period. Other students will be asked to submit evaluations of the presentations.

Homework Assignments

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:

- 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).
- A standard rubric is available.
- Students in 5850 will be asked to complete additional homework problems which go more deeply into the underlying concepts or apply more advanced techniques.
- Regrade requests for specific problems must be made within a week of the grading of that assignment.

Term Project

Throughout the semester, students will apply the concepts and methods from class to a data set of their choosing. If a student does not have a data set in mind, we will find one which aligns with their interests.

The term project can be completed individually or in groups of 2. Students will provide updates throughout the semester corresponding to the various tasks discussed in each module and will submit a final report at the end of the semester. The deliverables are:

- A proposal describing the research question and hypotheses, the data set, and the numerical or statistical models the student would like to use to test the hypotheses;
- A simulation study applying simulation methods to the models;
- A final presentation and report.

Tentative Schedule

#	Date	Day	Topic
			Introduction and Class Overview
1	01-22	Mon	Overview, GitHub Intro
2	01-24	Wed	Hypothesis Testing and Data Analysis
3	01-29	Mon	Simulating Data-Generating Processes
4	01-31	Wed	Example Application: Climate Change and Coastal Flooding
			Probability and Statistics Background and Review
5	02-05	Mon	Uncertainty and Probability Review
6	02-07	Wed	Probability Models and Model Residuals
7	02-12	Mon	Probability Models II
8	02-14	Wed	Bayesian Statistics and Decision Analysis
9	02-19	Mon	Hypothesis Testing As Decision-Making
			Data Visualization and Graphical Checks
10	02-21	Wed	Data Visualization
	02-26	*Mon*	*February Break*
11	02-28	Wed	Graphical Checks and Model Diagnostics
12	03-04	Mon	In-Class Figure Discussion
			Simulation Methods
13	03-06	Wed	Uncertainty Propagation and Monte Carlo Simulation
14	03-11	Mon	Monte Carlo Simulation II
15	03-13	Wed	Sampling Distributions and The Bootstrap

#	Date	Day	Topic
16	03-18	Mon	The Bootstrap II
17	03-20	Wed	Markov Chain Monte Carlo
18	03-25	Mon	MCMC II
19	03-27	Wed	Literature Presentations
	04-01	*Mon*	*Spring Break*
	04-03	*Wed*	*Spring Break*
			Model Calibration
20	04-08	Mon	Statistical Model Calibration
21	04-10	Wed	Model-Data Discrepancy
22	04-15	Mon	Missing and Censored Data
			Model Selection
23	04-17	Wed	Model Assessment and Predictive Accuracy
24	04-22	Mon	Predictive Information Criteria
25	04-24	Wed	Model Assessment and Selection: Examples
			Emulation and Surrogate Modeling
26	04-29	Mon	Tradeoffs Between Complexity and Ensemble Size
27	05-01	Wed	Emulating Expensive Models
28	05-06	Mon	Project Presentations