

BEE 4850/5850 (Environmental Data Analysis and Simulation)



Syllabus

Spring 2026

Course Overview

This is a 3 credit course offered as an elective.




Instructor

-  Vivek Srikrishnan
-  vs498@cornell.edu
-  318 Riley-Robb Hall

TA

-  TBD
-  TBD
-  TBD

Meetings

-  MWF
-  11:15-12:05
-  401 Riley-Robb Hall

Course Description

Understanding data is an increasingly integral part of working with environmental systems. Data analysis is a critical part of understanding system dynamics and projecting future conditions and outcomes. This course will provide an overview of a generative approach to environmental data analysis, which uses simulation and assessments of predictive performance to provide insight into the structure of data and its data-generating process. We will discuss exploratory analysis and visualization, model development and fitting, uncertainty quantification, and model assessment. 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:

- conduct exploratory analyses of data, including creating, interpreting, and critiquing data visualizations;
- calibrate environmental models to observations, including missing data;
- quantify and propagate uncertainty using simulation methods such as the bootstrap and Monte Carlo;
- assess model adequacy and performance using predictive simulations;
- evaluate evidence for and against hypotheses about environmental systems using model simulations.

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;
- Principles of data visualization;
- Probability models for data;
- Extreme values;
- Missing data;
- Model fitting;
- Uncertainty quantification with the bootstrap and Monte Carlo;
- Model assessment and comparison.

Course Meetings

This course meets MWF from 11:15-12:05 in 401 Riley-Robb. 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. 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 competency 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.

Textbooks and Course Materials

There is no required text for this class, and all course materials will be made available on the [course website](#) or through the Cornell library. However, the following books might be useful as a supplement to/expansion on the topics covered in class:

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). <http://www.stat.columbia.edu/~gelman/book/BDA3.pdf>
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.). <https://xcelab.net/rm/>
- D'Agostini, G. (2003). *Bayesian Reasoning in Data Analysis: A Critical Introduction*.
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories*. <https://avehtari.github.io/ROS-Examples/index.html>

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 professional and courteous on all course interactions and platforms, and (except in designated off-topic boards or forums) please keep all online discussion relevant to the course. We do not anticipate this as a problem given our experience; almost all students in almost all classes meet these expectations. However, even a single incident can do serious damage to the learning environment and the well-being of your fellow students.

Sexually explicit, harassing, threatening, bullying, trolling, racist, sexist, homophobic, transphobic, or otherwise grossly unprofessional content will be removed. Anyone behaving in these fashions or posting such content will be blocked/banned from the appropriate platform and may be given an F if they are consistently disruptive.

! 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.

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

- **Do not take screenshots of code.** I will not respond. Screenshots can be difficult to read and limit accessibility. Put your code on GitHub, share the link, and point to specific line numbers if relevant, or provide a *simple*, self-contained example of the problem you're running into.
- 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.
- 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

Many policies below (including grading policies) are broken out and discussed further on the course website. Lack of familiarity with any of these policies is not an excuse for violating any of them.

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

Please stay home and rest if you have symptoms of COVID-19 or any other respiratory illness. No masking will be required, but please be respectful of others who may wear masks or take other precautions to avoid illness. This policy may change if there is another outbreak of COVID-19 (or other illness), but will be kept consistent with broader Cornell mask policies.

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

Large language models (LLMs), such as GPT, and other generative AI models are powerful tools for predicting text and code patterns. However, while they can save time (though you'd be surprised at how little time their careful use saves), they often make mistakes that can be hard to detect or fail to communicate key ideas or insights at the expense of more general and banal text. As you are likely to encounter an ever-widening use of these tools when you leave Cornell, it is critical that you learn how to use these tools responsibly and effectively.

However, this class will not focus on teaching you how to use LLMs responsibly; we assume that if you are using these tools, you are doing so in a way which benefits your learning and helps you communicate your already-existing understanding, rather than substituting for it. You are **generally** permitted to use these tools, but you are ultimately responsible for guaranteeing, understanding, and interpreting your results. If you submit work that was LLM generated, and this work receives a poor grade due to the LLM's inability to demonstrate understanding, your grade will reflect that substantive assessment.

General guidelines for AI/ML use:

- **AI tools for code:** You may use LLM tools for code development and debugging, particularly for translating between other languages that you already have a better knowledge of (*e.g.* Python) and Julia. However, LLMs often make bad decisions about how to structure code, can introduce bugs (including suggesting packages that do not exist or outdated syntax), and can mislead you about what your code is doing. You are responsible for understanding and debugging any code involved in solving computational exercises. Notably, if you ask for help debugging LLM-generated code, this will not be possible unless you have already developed your own understanding of what each piece of the code does (or should do) and which parts work and don't work.
- **AI tools for writing:** LLMs **should not be used to generate text that you submit as your own work**. The point of written assessments (including model derivations and interpretations) is to stimulate your own critical thinking and mathematical skills, and you short-cut this process by substituting LLM use for your own process of formulating and articulating ideas. Even using LLMs to do the initial writing and editing that output will result in shallower thinking. As we are not teaching you in this class to critically engage with LLMs, However, you *may* use LLMs or similar tools (*e.g.* Grammarly) to help you edit your writing. As mentioned before, you are ultimately responsible for the content of any work you turn in: if a tool used to edit grammar changed the substance of your writing, you will be responsible for the submission.

To distinguish between these permissible and impermissible uses, you are required to cite your use of the tool (as with any other external reference). In particular, you must disclose how you use the LLM and how you incorporated any output into your final submission. **Failure to fully reference the interaction, as described above, will be treated as plagiarism and referred to the University accordingly.** If you have any questions about whether your planned use of an AI/ML tool complies with the academic integrity policy, please consult a member of the course staff ahead of its use.

Late Work Policy

In general, late work can be submitted up to 24 hours late at a 50% penalty, and will not be accepted after that point. This policy may seem strict, but allows for prompt release of solutions and discussion of assignments. Please reach out *as soon as possible* (ideally before the due date) if legitimate circumstances emerge which prevent you from submitting work within 24 hours of the due date; we will make accommodations for approved reasons, which might include a limited extension or dropping the assignment.

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 explaining 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.

The first regrade request for any submission will be handled by the person who graded that homework problem. The first regrade request for any exam submission will be handled by whoever graded that exam problem. If the submission was graded by the TA, additional regrade requests for the same submission will be handled directly by the instructor. 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 the TA or 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. However, don't fish — your grade might be lowered if the TA finds an independent mistake while regrading.
- 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”.

Office Hours

Office hours with both Prof. Srikrishnan and the TA will be available each week at times specified at the top of this syllabus. Changes to the office hour schedule (cancellations/rescheduling) will be announced in class and on Ed Discussion.

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 or help with syntax 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.

What If I Can't Make Office Hours?

While we will try to select office hours that work for as much of the class as possible, both the course staff and students have busy schedules and no time will work for everyone. **If you need help outside of office hours (*e.g.* office hours do not fit your schedule), please send an email to the TA or Prof. Srikrishnan as soon as possible.** These requests may not be accepted on short notice (*e.g.* if you have a question about a homework due on Thursday and send a request on the immediately prior Wednesday; schedules for course staff may already be full). We recommend starting your homework promptly so you can take advantage of office hours or make an appointment over a longer period.

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 be 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
Labs	10%
Quizzes	10%
Participation (4850)/Literature Critique (5850)	10%
Homework Assignments	15%
Prelims	30%
Term Project	25%

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

Grade	Range
A	93–100
A-	90–93
B+	87–90
B	83–87
B-	80–83
C+	77–80
C	73–77
C-	70–73
D+	67–70
D	63–67
D-	60–63
F	< 60

Participation

Participating fully in the class allows you to gain more from the class and contribute more to the learning of your classmates. Some ways to participate include:

- Asking questions in class or on Ed;
- Answering questions in class or on Ed;
- Actively engaging in in-class activities;
- Coming to office hours.

Note that just passively attending class will not yield full participation points. Participation points are not free: you are likely to lose points if you consistently skip class or do not ask or answer questions online or in person. At the end of the term, you will be asked to evaluate your own participation over the course of the semester, in addition to my documentation of your participation. Participation is a component of the grade for students enrolled in BEE 4850.

Quizzes

Quizzes will be assigned in Gradescope most weeks based on recently discussed content. These will be relatively short and are intended to consolidate material recently discussed in class. The quizzes will be released after Wednesday's class and will be due prior to next Monday's. Quizzes can be retaken as many times as are desired prior to the submission deadline. They may consist of multiple choice questions or questions involving small mathematical, computational, or open-ended problems.

Labs

Several class periods (typically on Fridays) will be dedicated to in-class labs. Students will be given work-sheets and Jupyter notebooks aimed at getting hands-on practice with the prior lecture topic(s). Lab writeups will be due before the class meeting the following Monday after the lab period.

Literature Critique

Students in BEE 5850 will select a peer-reviewed journal article related to an application of data analysis and will write a short discussion paper (2-3 pages) analyzing the hypotheses and statistical choices. Students should feel free to select their own paper or can work with Prof. Srikrishnan to identify one of interest, but in all cases should discuss with Prof. Srikrishnan to ensure that the article is appropriate. The discussion paper will be due towards at the end of the semester.

Homework Assignments

There will be approximately 4 homework assignments assigned. Homework assignments are intended to be more in-depth applications of course material to data analysis problems.

You will generally have two class weeks to work on an assignment. This is intended to provide you enough time to work on the problem and debug and evaluate your code (including troubleshooting any technical problems); these are not reasons for late submission. Each homework assignment will build on material from the prior classes and possibly from the day the homework is assigned.

Students are encouraged to collaborate and learn from each other on homework assignments, but students must submit their own assignments which represent their own 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. While GitHub use is not required for the class aside from accepting and cloning assignments, students are encouraged to update their GitHub repositories as they work on the assignments; this helps with answering questions, keeping solutions synced across groups, 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 (usually a Thursday). Your assignment writeup should be submitted to Gradescope as a PDF with the answers to each question tagged (a failure to do this will result in a non-negotiable 10% point deduction).
- A meta-rubric is provided on the website, under the Homework page. These are not customized for each assignment but the principles will apply generally.
- No homework assignments will be dropped, but you can turn in assignments within 24 hours of the due date with a 50% penalty. If you need a further accommodation for a particular assignment, talk to

Prof. Srikrishnan *before the due date*. Requests for extensions made after the due date will only be considered under extraordinary and unexpected circumstances. Technical challenges submitting assignments are not acceptable reasons for extensions to be granted, and late penalties will apply.

- **Your submitted homework must stand on its own***! We cannot grade you on the basis of information which was not included in the submitted assignment. While regrade requests should include a justification for why your grade is incorrect, we will not consider explanations or additional reasoning outside of the submission.

Prelims

One in-class prelim will be given. The exam is closed-book and closed-note. As a result, the exam will emphasize conceptual material such as model derivations and interpretation of results; any calculations can be done with a pen(cil) and paper. Conflict and extended-time exams will be handled through SDS. Exams will be scanned into Gradescope for grading and feedback.

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. There will be

- Several updates aligning with course models. Students will provide short summaries of their proposed topic and exploratory analysis, proposed probability model(s), simulation studies, and hypothesis tests/model assessments. Each of these should be no more than 2 pages (11 point font, 1 inch margins), not including of figures or references.
- A final presentation and report. The report should be no more than 5 pages (11 point font, 1 inch margins), not including references and figures. The presentation should be no more than 10 minutes and will be delivered in-class; presentations may be spread across multiple class periods if the number of projects requires it. More details and rubrics will be provided later in the semester.

Tentative Schedule

#	Date	Day	Topic	HW	Project
Introduction					
1	01-21	Wed	Course Overview and Introduction	HW1 Released	
2	01-23	Fri	Probability Review		
Exploratory Analysis					
3	01-26	Mon	Exploratory Data Analysis		
4	01-28	Wed	Data Visualization		
5	01-30	Fri	Lab: Data Visualization		
Probability Models					
6	02-02	Mon	Probability Models and Linear Regression	HW1 Due HW2 Released	Exploratory Analysis
7	02-04	Wed	More Linear Regression and Observation Error		
8	02-06	Fri	Maximum Likelihood Estimation and Fisher Information		
9	02-09	Mon	Multiple Regression and Causal Reasoning		
10	02-11	Wed	Direct Acyclic Graphs and Confounds		
11	02-13	Fri	Lab: Reasoning with DAGs	HW2 Due HW3 Released	
	02-16	Mon	February Break		
12	02-18	Wed	Generalized Linear Models		
13	02-20	Fri	More Generalized Linear Models		
14	02-23	Mon	Lab: Selecting a Probability Model		
15	02-25	Wed	Autocorrelation and Time Series		
16	02-27	Fri	Time Series Examples		
17	03-02	Mon	Modeling Extremes: Block Maxima		
18	03-04	Wed	Modeling Extremes: Peak Over Thresholds		
19	03-06	Fri	Lab: Extreme Value Models	HW3 Due	
Simulation Methods					
20	03-09	Mon	Random Variate Simulation	HW4 Released	Candidate Model(s)
21	03-11	Wed	Random Number Generation		
22	03-13	Fri	Prelim		
23	03-16	Mon	Monte Carlo: Theory		
24	03-18	Wed	Monte Carlo: Examples		
25	03-20	Fri	Lab: Monte Carlo	HW4 Due	
26	03-23	Mon	Uncertainty Quantification and the Bootstrap		
27	03-25	Wed	Bootstraps for Structured Data		
28	03-27	Fri	Lab: The Bootstrap		
	03-30	Mon	Spring Break		
	04-01	Wed	Spring Break	HW5 Released	
	04-03	Fri	Spring Break		
29	04-06	Mon	Missing and Censored Data		
30	04-08	Wed	Multiple Imputation		
31	04-10	Fri	Lab: Missing Data		
Hypothesis Testing and Model Assessment					
32	04-13	Mon	Overfitting and Bias-Variance Tradeoff	HW5 Due HW6 Released	Simulation Study
33	04-15	Wed	Scoring Rules and Graphical Checks		
34	04-17	Fri	Cross-Validation		
35	04-20	Mon	Hypothesis Testing		
36	04-22	Wed	Entropy and Information		
37	04-24	Fri	Information Criteria		
Design of Experiments and Value of Information					
38	04-27	Mon	Value of Information	HW6 Due	
39	04-29	Wed	Design of Experiments		
40	05-01	Fri	Prelim 2		
41	05-04	Mon	Class Wrap-Up		