ENGG 107: Bayesian Statistical Modeling and Computation

Introduction

(Meeting #1)

Objectives for today:

1. To get to know each other
2. To introduce the motivation, the design, and the logistics of the class

Outline for today

1. Motivation
2. Getting to know each other
3. Overview of the course
4. Logistics

Outline for today

1. Motivation
2. Getting to know each other
3. Overview of the course
4. Logistics

Do you bet on a new coin?

- Someone approaches you in Times Square in NYC.
- They offer you a bet on a coin resting on their hand.
- You see a head on the upwards side of the coin you can see.
- Bet: Pay 10$ upfront. If there is just one tail in 10 repeated tosses of the coin, you get double your payment.
- Do you accept this bet?
- How would you analyze this decision using the statistics you have learned thus far?

Zebras vs horses?

- You stand in Times Square in NYC and hear the sound of hoofs behind you.
- Without turning around, what is your estimate of the probability the this is a zebra or a horse behind you?
- How do you arrive at this estimate?
- Does your estimate change if you were
  - in the zoo in Central Park?
  - in a national park in the Serengeti?
  - at a horse race track?

The German Yodeling Test

- There is genetic trait (i.e., exposure to German yodeling without headache) with a 0.5 % incidence rate in certain populations (let’s assume Germans). You, as a member of the population, test positive, but the test is only 98 % accurate.

- What is the probability that you have the trait after one, two, and three tests?

Please try to solve this warm-up problem
in class for the first yodeling application

More applied:

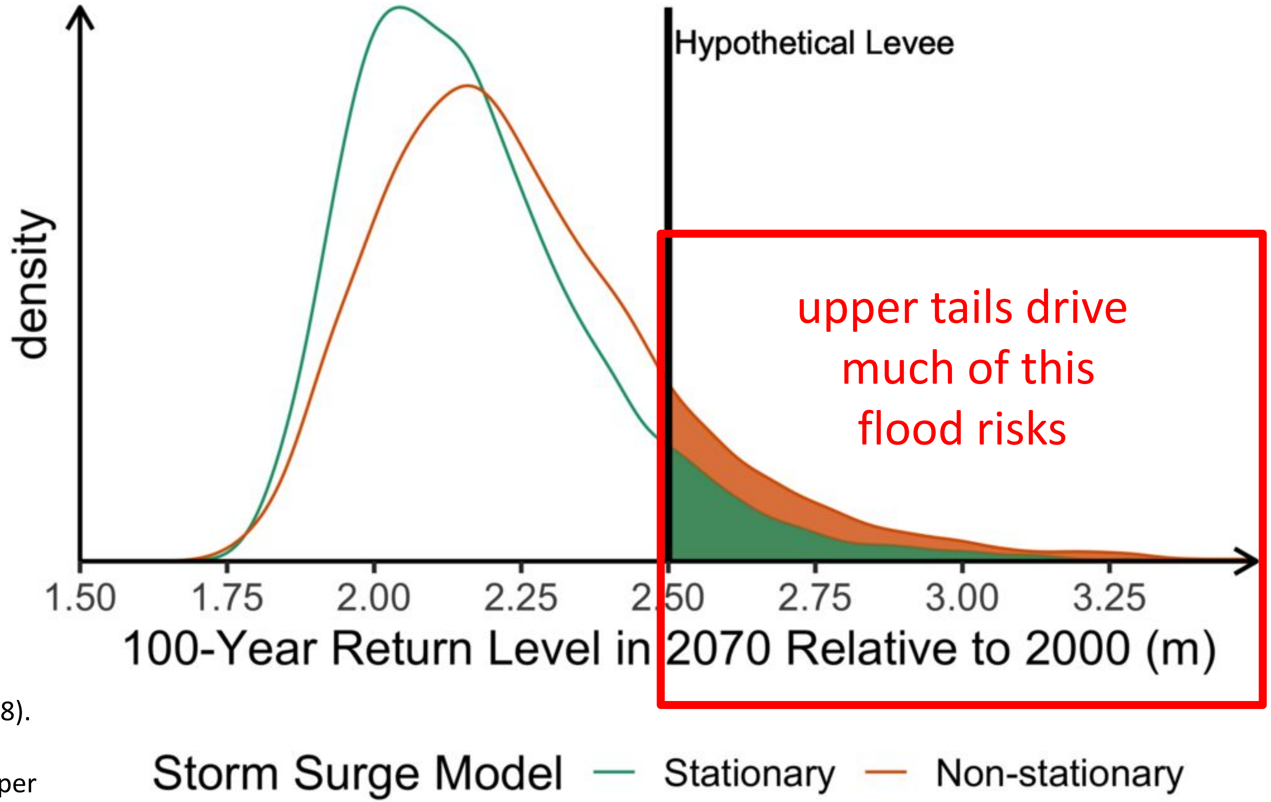
Can you still make it to the next airport three flight hours away?

Picture: Klaus Keller

How high to build a new flood wall?

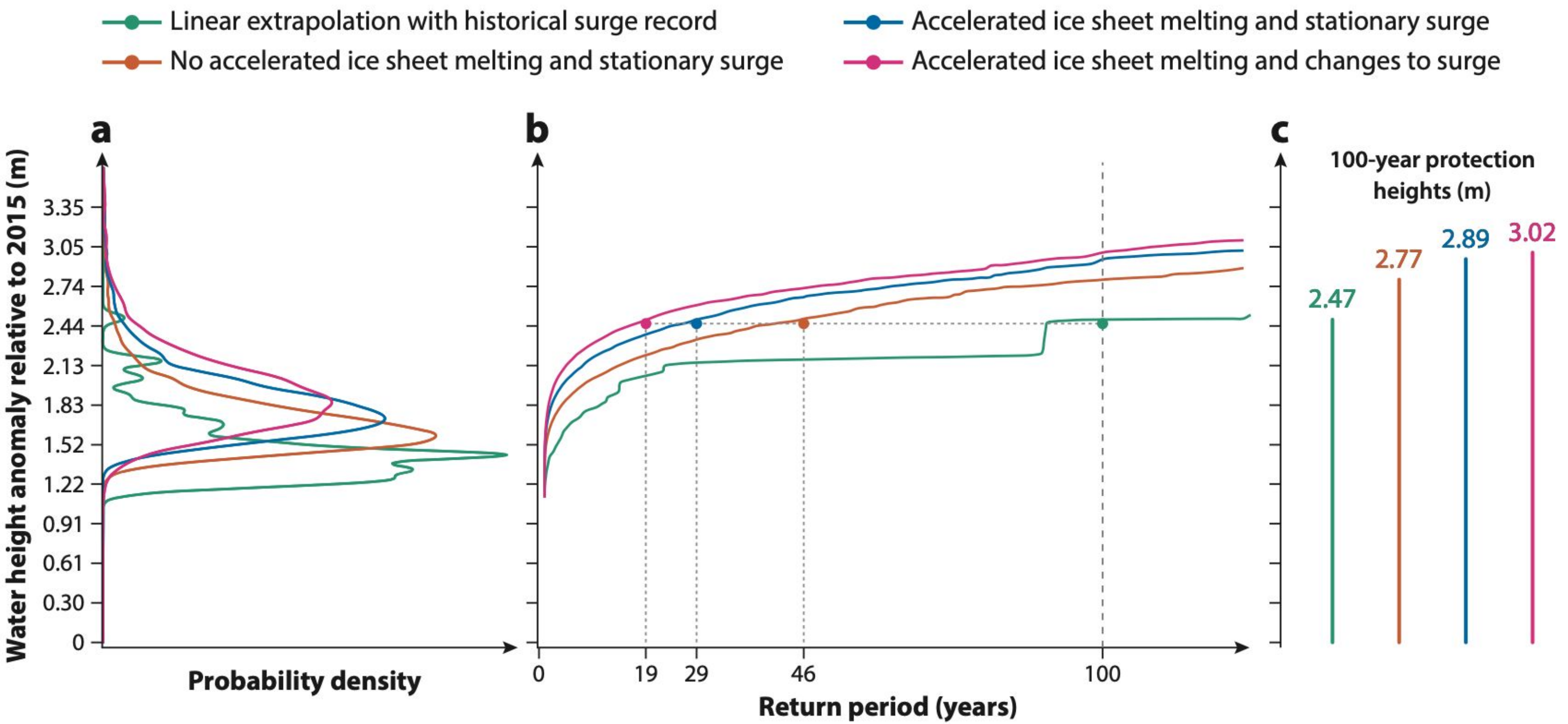
Picture: Klaus Keller,
(Floodwall in Norfolk, Virginia)

1. What is the flooding risk behind a levee?
2. What determines this risk?
3. What aspects of these questions are in the field of engineering?



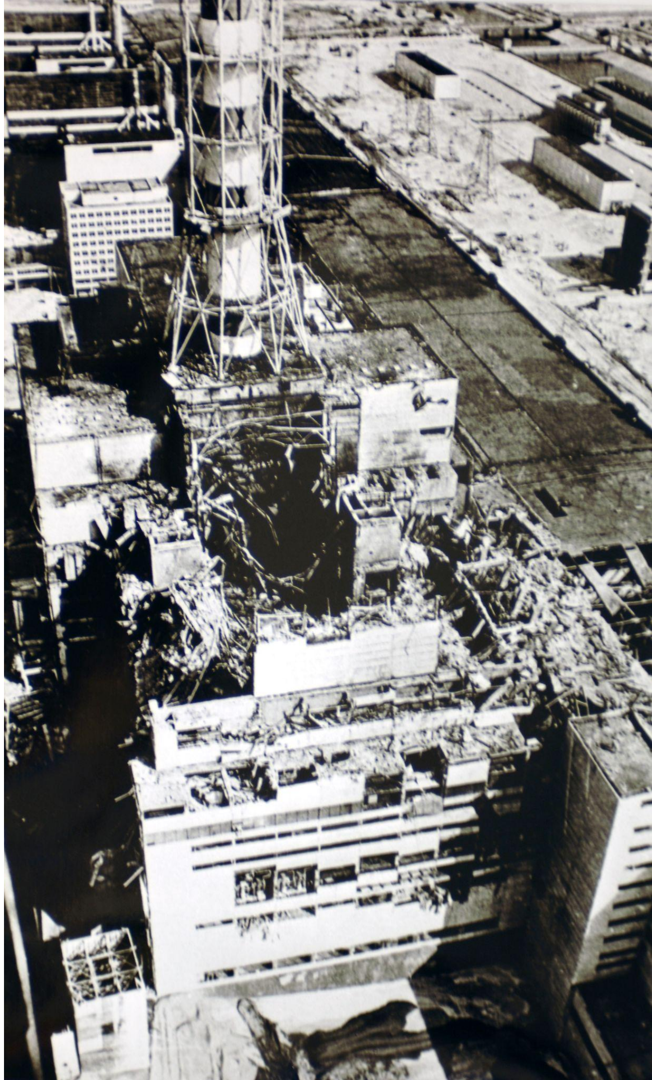
Wong, T. E., Klufas A., Srikrishnan, V., & Keller, K. (2018). Neglecting model structural uncertainty underestimates upper tails of flood hazard. *Environmental Research Letters 13*(7), 074019 (Results redrawn, simplified, and annotated for clarity)

It turns out there may be several answers to parts of the question. How to analyze and communicate these results?



Vivek Srikrishnan, R. Alley, and **K. Keller**: Investing in Science and Using the Results to Improve Climate Risk Management. ***EOS*** (2019) (redrawn for clarity).

How probable is a nuclear accident?



See: Keller, W., & Modarres, M. (2005). A historical overview of probabilistic risk assessment development and its use in the nuclear power industry: a tribute to the late Professor Norman Carl Rasmussen. *Reliability Engineering & System Safety*, *89*(3), 271–285. https://doi.org/10.1016/j.ress.2004.08.022

Motivation

- Statistical inferences in Engineering (and beyond) can be crucial and complex.
  - The number of observations can be small.
  - The predictions can hinge critically on deeply uncertain structural or prior assumptions.
  - The observation and model errors are unknown and not normally distributed.
  - The models can be nonlinear.
  - The observations may not stem from randomized and controlled experiments.
  - There are several possible models to analyze the observations.
  - Information gathered at one scale (*e.g.,* several local observations) has to be used in the analysis at a different scale (e.g., a watershed, an entire plant, ..).
  - How one communicates the results affects outcomes.
  - Poor skills and techniques can have social, economic, ethical, and legal consequences.
- Techniques taught in standard statistics can be of limited utility to address these issues.
- Yet, despite these problems, decisions have to be made

How can we understand and solve (some) of these problems?

Course Objectives

Students will be able to understand, explain, choose, implement, and communicate concepts and techniques of Bayesian analyses **relevant to their own research**.

Disclaimer

- This class will NOT educate to be a card-carrying and Ph.D.-level Bayesian statistician or to be able to perform *complex and publication ready* Bayesian analysis alone.
- You will (hopefully) learn a few important skills.
  - Understand the vocabulary and basic tools.
  - Recognize when simple methods are appropriate (and when not).
  - Know about, how to implement, and how to apply more advanced and more appropriate methods.
  - Know where to get help if you stump onto a problem that requires more research.

Outline for today

1. Motivation
2. Getting to know each other
3. Overview of the course
4. Logistics

Getting to know each other

Please introduce yourself addressing at least the following questions:

1. Why are you in this class?
2. What kind of problems do you typically address in your studies / research (if any)?
3. What kind of statistical inferences do you think you will perform in the future or need to understand?
4. What is your previous training in statistics and programming?
5. If there is one outcome you would like to take away from this class, what would this be?
6. What is your favorite movie (and why)?

Outline for today

1. Motivation
2. Getting to know each other
3. Overview of the course
4. Logistics

Draft Schedule

Week starting	Class 1	Class 2	Class 3
January 2, 2023	No Class	Introduction to the class	Bayesian Approach I
January 9, 2023	Bayesian Approach II	Picking a project	Setting up Computation
January 16, 2023	Analytical Solutions / Kalman	Precalibration /(Bayes) Monte Carlo	MCMC Part 1
January 23, 2023	MCMC Part 2	Bayesian Workflow	Bayesian Workflow
January 30, 2023	Students pitch project ideas	Writing a method section	Catching up / review
February 6, 2023	Convergence diagnostics	Checking Assumptions	Deep Uncertainty
February 13, 2023	Geeting Solid Priors	Links to Decision-Making	Links to Decision-Making
February 20, 2023	Model Choice	Emulation	Sensitivity analysis
February 27, 2023	Communication of results	Student Presentations	Student Presentations
March 6, 2023	Class Debriefing / Resarch links	No Class	No Class

Outline for today

1. Motivation
2. Getting to know each other
3. Overview of the course
4. Logistics

Logistics

- Weekly meetings:
  - Lectures: Classes: MWF 11:30 – 12:35 (009 ESCS)
  - X times on Tu 12:15 – 1:05 PM (**on demand, announced**)
  - Problem sets discussions: Fridays 14:00-15:00
    (meet in my office and outside my office in a work area)
  - Office hours: Fridays 15:00-16:30 (in my office IR 385)
- A mixture of
  - lectures to introduce the example problem, the theoretical background, and the possible approaches to solving the problem and
  - problem centered sessions where we tackle example problems in class exercises and code discussions.
- Does everybody have access to a computer?
- What are your experiences with data analysis programs?
- Materials and assignments via CANVAS

Getting in touch and giving feedback

- Office hours, by appointment, or just drop by.
- Email is, in general, the fastest and easiest way to get in touch with me. *Canvas messaging does not work fast with me.*
- This class is designed to be tuned to students research needs.
- I suggest we play this adaptively.
- Please:
  - Give me feedback on topics, style, speed, workload.
  - Suggest topics that are interesting to you

Expectations from Students

This class will require your sustained attention.

Students are expected to:

- prepare for the meetings (e.g., by carefully reading and synthesizing the reading assignments and being prepared to present their synthesis in class),
- actively contribute to the group discussions, and
- prepare and submit the assignments in time.

The class is designed with the expectation that students spend roughly three times the lecture contact hours outside the class for readings and assignments.

Academic Integrity

- We are all expected to do our own work on exams.
- Plagiarism is the handing in of materials from other sources as your own work without quoting or clear citation.
- Please review the syllabus on CANVAS and return a signed copy on the assigned due date.

Grading

The overall course grade will be based on a weighted average of the grades for these components:

• Problem sets	40%
• Term Project Presentation	20%
• Written Term Project Report	40%

We adopt the Vikrant Vaze Grading Algorithm (1):

"Your final grade will be decided based on your absolute as well as relative performance in the class, whichever is higher. Out of 100 total points, if you score at least min(90, Median + Standard Deviation) points, you are guaranteed an HP grade. If you score at least min(75, Median - Standard Deviation) points, you are guaranteed at least a P grade. Note that this is even more favorable to you than the standard grading on the curve."

(1) Personal communication from Vikrant Vaze

Reading Assignments for Next Class

- Core:
  - Applegate, P. J., & Keller, K. (Eds.). (2016). Risk analysis in the Earth Sciences: A Lab manual. 2nd edition. Leanpub. Open Source and free at: https://leanpub.com/raes **(Chapter 1)**
  - Clyde, M., M. Çetinkaya-Rundel, C. Rundel, D. Banks, C. Chai, L. Huang. (2022). An Introduction to Bayesian Thinking. Retrieved from https://statswithr.github.io/book/ accessed 02/07/2022 **(Chapter 1)**

- Supplementary (if you are interested in background)
  - Freedman, D. (1997). Some Issues in the Foundation of Statistics. In B. C. van Fraassen (Ed.), Topics in the Foundation of Statistics (pp. 19–39). Springer Netherlands. https://doi.org/10.1007/978-94-015-8816-4_4
  - Jefferys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. American Scientist, 80(1), 64–72. https://www.jstor.org/stable/pdf/29774559.pdf
  - Clark, J. S. (2004). Why environmental scientists are becoming Bayesians: Modelling with Bayes. Ecology Letters, 8(1), 2–14. https://doi.org/10.1111/j.1461-0248.2004.00702.x

Review

Do you know

1. Who is in this course?
2. How this course is designed?
3. What you can learn in this course?
4. What is expected from you in this course?