Statistics and the Future of the Antarctic Ice Sheet

Murali Haran

Department of Statistics, Pennsylvania State University

Frontiers of Science Lectures
Penn State University
January 2020.

Statistics in the 21st Century

Statistics is the dream job of the next decade. – Hal Varian, Chief Economist at Google

- #1 business job and #2 in ranking in best jobs – US News and World Report (2019)
- More statistician/data analyst jobs than qualified candidates

Statistics and Science

The #1 business job is also central to science!

- Discovery: finding information/patterns in data
  Finding a gene (out of thousands) related to a disease

Statistics and Science

The #1 business job is also central to science!

- Discovery: finding information/patterns in data
  Finding a gene (out of thousands) related to a disease
- Separating signal from noise
  Did the medical treatment truly help patients or did they get better by chance/other reasons?

Statistics and Science

The #1 business job is also central to science!

- Discovery: finding information/patterns in data
  Finding a gene (out of thousands) related to a disease
- Separating signal from noise
  Did the medical treatment truly help patients or did they get better by chance/other reasons?
- Synthesis: combining different data with science
  Understand ice sheets from satellite images of ice sheets + ice core data + the physics of ice sheets

Statistics and Science

The #1 business job is also central to science!

- Discovery: finding information/patterns in data
  Finding a gene (out of thousands) related to a disease
- Separating signal from noise
  Did the medical treatment truly help patients or did they get better by chance/other reasons?
- Synthesis: combining different data with science
  Understand ice sheets from satellite images of ice sheets + ice core data + the physics of ice sheets
- Decision-making: you need probability to assess risk
  How high to build levees in New Orleans?

Statistics and Playing in Other People's Backyards

- Philosophy: how should you weigh evidence? What is the nature of truth? Randomness?

Statistics and Playing in Other People's Backyards

- Philosophy: how should you weigh evidence? What is the nature of truth? Randomness?
- Computer Science: clever statistical methods and complicated (and often huge) data sets require sophisticated computing

Statistics and Playing in Other People's Backyards

- Philosophy: how should you weigh evidence? What is the nature of truth? Randomness?
- Computer Science: clever statistical methods and complicated (and often huge) data sets require sophisticated computing
- Mathematics: understanding properties of statistical methods

Statistics and Playing in Other People's Backyards

- Philosophy: how should you weigh evidence? What is the nature of truth? Randomness?
- Computer Science: clever statistical methods and complicated (and often huge) data sets require sophisticated computing
- Mathematics: understanding properties of statistical methods
- Multidisciplinarity: Building statistical models for research in an enormous number of disciplines

Statistics and Playing in Other People's Backyards

- Philosophy: how should you weigh evidence? What is the nature of truth? Randomness?
- Computer Science: clever statistical methods and complicated (and often huge) data sets require sophisticated computing
- Mathematics: understanding properties of statistical methods
- Multidisciplinarity: Building statistical models for research in an enormous number of disciplines

  "The best thing about being a statistician is that you get to play in everyone's backyard." – John Tukey

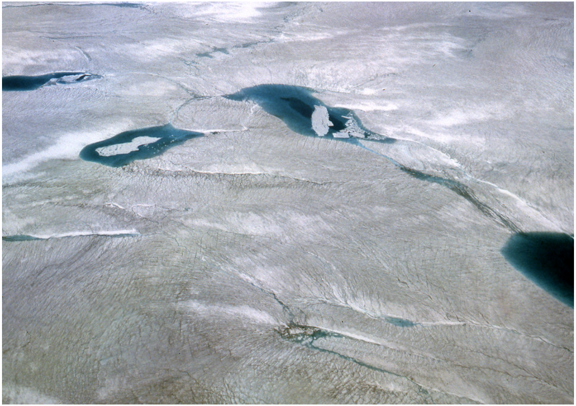
Eberly College's Frontiers of Science, 2020 Edition

- Studying ice sheets (today!)
- January 25: Predicting plant diversity – Pamela Soltis, U of Florida
- February 1: Understanding animal movement and the spread of diseases – Ephraim Hanks, Statistics and Center for Infectious Disease Dynamics, Penn State
- February 8: Characterizing potentially habitable planets – Eric Ford, Astronomy/Astrostatistics, Penn State
- February 15: Disease outbreak control – Katriona Shea, Biology, Penn State
- February 22: Genetic mutations and childhood obesity – Kateryna Makova, Biology, Penn State

What are Ice Sheets?

- Enormous mass of glacial land ice
- Combine for over 99% of the freshwater ice in the world

Greenland



Over 1.7 million km^2 area

Antarctica

Over 14 million km^2 area

Statistics and Ice Sheets

1. Understanding the science of ice sheets: How is it moving? How is it interacting with the ocean waters?
2. Projecting the future: how much of the ice sheet will melt in the next century?
3. Decision-making: should we/how do we plan for ice sheet melting impacts?

Penn State Connections

- Penn State has world class researchers who work on understanding ice sheets
  Penn State Ice and Climate (*PSICE*)
- Many PSU researchers work on climate risk analysis
  Center for Climate Risk Management (*CLIMA*)

Ice Sheets and Sea Level Rise

(*courtesy NASA*)

Ice sheets contain an immense amount of water

- Antarctic ice sheet: equivalent of 58 m. sea level
- Greenland ice sheet: equivalent of 6 m. sea level
- Glaciers and ice caps: equivalent of 0.4 m. sea level

The Impact of Sea Level Rise

- Even a modest contribution to sea level rise can have a major impact on humans, e.g. on low-lying areas and areas prone to storm surges

http://www.miamiherald.com/

Fort Lauderdale, High Tide

Where should we build?

Hurricane Gustav, New Orleans

How high to build the wall?

Industrial Canal levee wall from lower 9th Ward at high water driven in by Hurricane Gustav.

It's tough to make predictions, especially about the future. – Yogi Berra

Wikimedia

A Question

- If you can't get the weather right in 5 days how could you possibly make a prediction about the ice sheet in 50 years?

Time to Talk Baseball Predictions...

- Can you guess what a batter will score the next game?

Time to Talk Baseball Predictions...

- Can you guess what a batter will score the next game?
- Difficult (Are you a gambler?!)

Time to Talk Baseball Predictions...

- Can you guess what a batter will score the next game?
- Difficult (Are you a gambler?!)
- But not a terrible idea to guess his batting average for next season

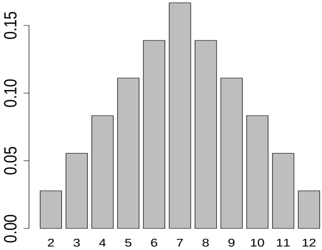
Time to Talk Baseball Predictions...

- Can you guess what a batter will score the next game?
- Difficult (Are you a gambler?!)
- But not a terrible idea to guess his batting average for next season
- The analogy:
  - Next game individual performance is like weather
  - Average for next season is like climate
- One more thing to help with climate projections: physics!

Climate is what you expect. Weather is what you get.

In technical terms

- Weather is a *realization*
- Climate is a *distribution*



A *Realization* of Two Dice Rolls

- Result of a particular roll of dice: a **realization**
- Weather is like seeing what happens when you roll the dice
  Example: The high temperature on January 17, 2020.
- Climate describes what tends to happen in the long run
  - Distribution of temperatures on January 17 over 1910-1940 or 2030-2060

How Do You Predict Weather on January 19?

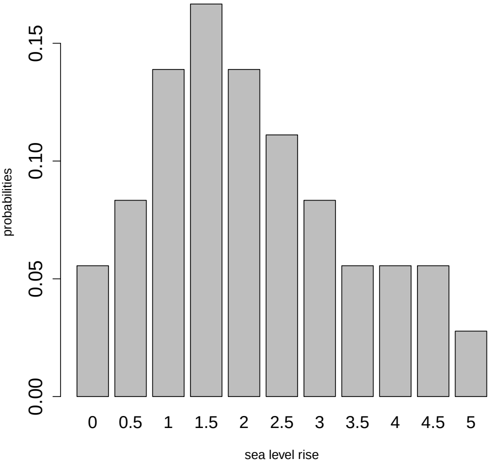
- **Numerical models**: computer simulation of the atmosphere for the next few days
- **Data**: temperature, humidity, wind etc. over past several days
- **Data assimilation**: combine data with model to make forecasts
- Cannot make accurate forecasts beyond a few (5-6) days because errors in the models grow very fast each day

Climate Projection

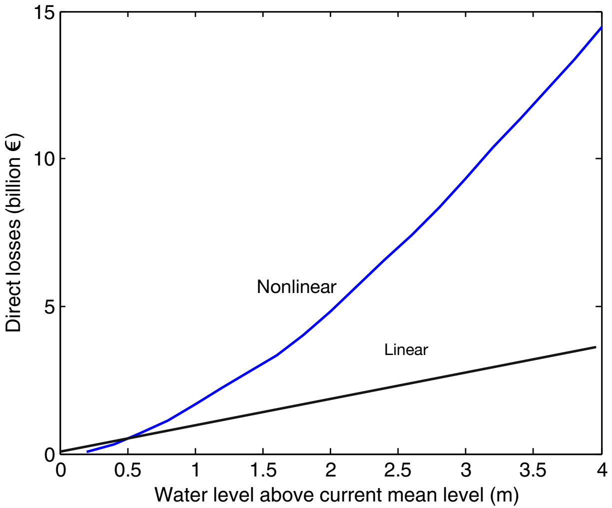
- **Climate**: tells you probability distribution of weather events over an extended period of time (typically around 30 years)
- Average low temperature on January 14 between 2050 and 2080
- Use climate models (based on physics)
- Use historical data to tune climate models
- Although challenging, can project far into the future

Sea Level Rise

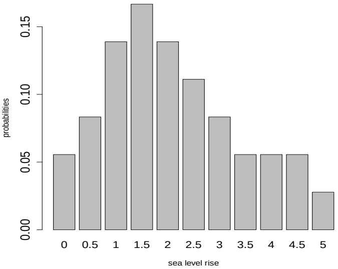
Want analysis to provide probability distribution for future sea level rise

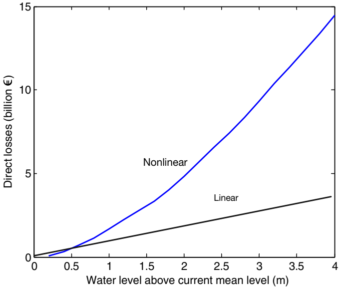


The Cost of Sea Level Rise



Risk Analysis for Sea Level Rise





Tail of the distribution = low probability events. They can be extremely costly so important to estimate them carefully

Probabilities and Risk Analysis

To make decisions based on potential future sea level rise:

- probability distributions associated with future sea level rise
- costs associated with each value of sea level rise

Important to understand the full probability distribution, especially the tails

Decision Making Under Uncertainty

- Decision: How high to build levees on the coast?
- Simpler decision: Should we buy airline ticket insurance?
  - Website offers insurance. Full ticket refund if you have to cancel due to a covered reason
- E.g. your ticket costs \$300. Insurance: \$30
- Is it worth getting?

Airline Ticket Insurance

- You get sick 1 week every year$\Rightarrow$ approximately 0.02 probability that you are sick any given day
- With insurance: spend \$30 whether or not you get sick
- Without insurance: with probability 0.98 your loss is \$0. With probability 0.02 your loss is \$300.
  $\boxed{\text{Expected loss} = 0.98 \times 0 + 0.02 \times 300 = \$6}$
- In the long run: no insurance $\Rightarrow$ lose \$6 on average with insurance lose \$30 on average
- Not a good idea to buy insurance

Insurance Against Sea Level Rise

- How high should we build a levee in New Orleans?
- Instead of two possibilities (flight canceled vs not canceled), many different potential levels of sea level rise: 0 m to 4+ m
- Instead of just two possibilities for actions (buy ticket insurance vs do not), many: do not build a levee, build a levee 0.5 m high, 1.1 m. high,...
- Decision-making is more complicated, but the principles are the same as for airline ticket purchase

Halftime Summary

1. Climate (distribution) versus weather (a realization)
2. Cannot predict weather beyond a few days but reasonable to project climate 50 years in the future
3. To project climate we need to combine physics with data
   - *We need statistical methods* to do this
4. Decision-making under uncertainty
   - *We need statistical methods!* to compute risk

Second Half Predictions

- Answer to the question: How is statistics used to make projections about ice sheets?

Interdisciplinary Research Team

- Dave Pollard, Penn State Institute for Energy and the Environment
  Develop mathematics model and computer code that simulates the ice sheet
- Vivek Srikrishnan, Klaus Keller, Department of Geosciences
  Design model runs, understanding risks, decision-making associated with sea level rise
- Ben Lee, Murali Haran, Department of Statistics
  Develop new statistical methods for combining information from models and data, and producing a projected distribution for sea level rise

Projections

We cannot simply extrapolate based on the past:

- Simple extrapolation based on data alone is risky
- Things are changing over time
- But we can let physics guide us. For example, use our understanding of the physics of the ice sheet to see how the ice sheet will respond to external factors
- Still need to use data: helps us calibrate or “tune” the ice sheet model so it becomes more realistic
- We will build **models** that combine physics + account for randomness + errors

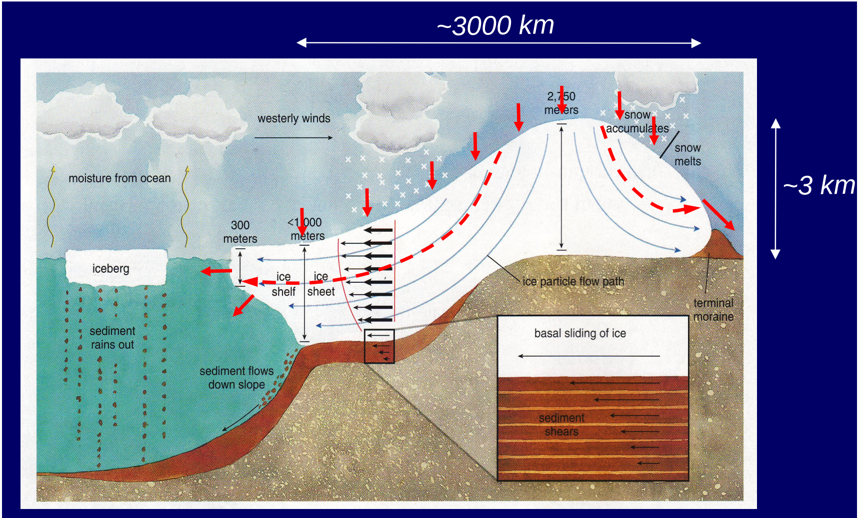
Models

A model is a representation of reality

Mathematical versus Statistical Models

- A mathematical model is a representation of reality in the form of mathematical equations
- A statistical model is a mathematical model that also accounts for two very important features:
  - randomness
  - the fact that the mathematical model is usually not a perfect representation
- Statistical model is a more realistic representation of reality

Model of Ice Sheet Physics



Siegert, 2002, Amer. Sci.

Translated into Mathematics (A Glimpse)

Full Stokes equations

$$
\begin{aligned}
&\frac{\partial}{\partial x}\left(2\eta\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(\eta\frac{\partial u}{\partial y}+\eta\frac{\partial v}{\partial x}\right)\\
&\qquad+\frac{\partial}{\partial z}\left(\eta\frac{\partial u}{\partial z}+\eta\frac{\partial w}{\partial x}\right)-\frac{\partial p}{\partial x}=0,\\
&\frac{\partial}{\partial x}\left(\eta\frac{\partial u}{\partial y}+\eta\frac{\partial v}{\partial x}\right)+\frac{\partial}{\partial y}\left(2\eta\frac{\partial v}{\partial y}\right)\\
&\qquad+\frac{\partial}{\partial z}\left(\eta\frac{\partial v}{\partial z}+\eta\frac{\partial w}{\partial y}\right)-\frac{\partial p}{\partial y}=0,\\
&\frac{\partial}{\partial x}\left(\eta\frac{\partial u}{\partial z}+\eta\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial y}\left(\eta\frac{\partial v}{\partial z}+\eta\frac{\partial w}{\partial y}\right)\\
&\qquad+\frac{\partial}{\partial z}\left(2\eta\frac{\partial w}{\partial z}\right)-\frac{\partial p}{\partial z}=\rho g. \qquad (10)
\end{aligned}
$$

Thoma et al., Geosci. Model Devel., 2014

where

$$\eta := \frac{1}{2}A(\hat{\theta})^{\frac{-1}{n}}\dot{\epsilon}^{\frac{(1-n)}{n}}$$

$$\dot{\epsilon} = \sqrt{\dot{\epsilon}_{xx}^2+\dot{\epsilon}_{yy}^2+\dot{\epsilon}_{xx}\dot{\epsilon}_{yy}+\dot{\epsilon}_{xy}^2+\dot{\epsilon}_{xz}^2+\dot{\epsilon}_{yz}^2}.$$

taken from David Pollard, Penn State Institute of Energy and the Environment

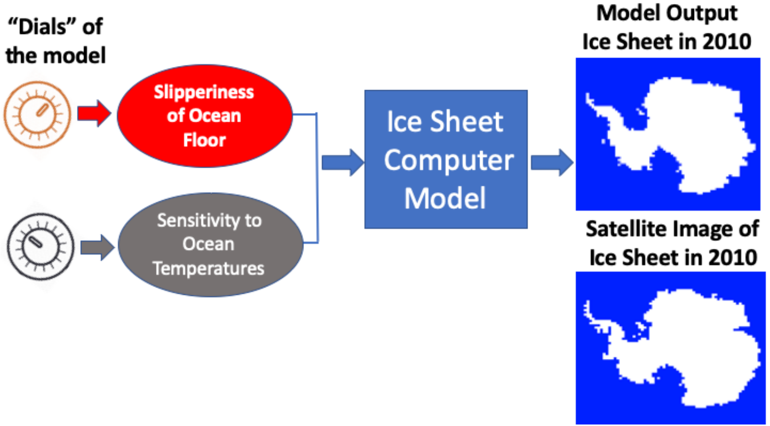
Translated into Computer Code

The mathematical equations are complicated. Need to be solved by computer algorithms

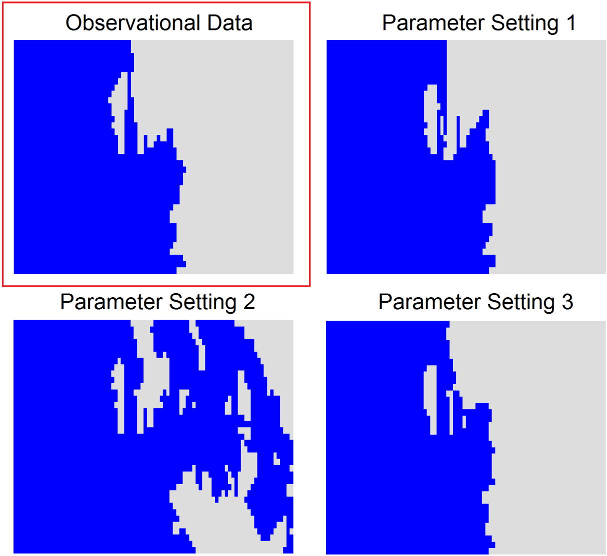
Physics $\rightarrow$ Mathematical Equations $\rightarrow$ Computer Model
(Computer Code)

"Computer experiments" are carried out with these computer models: useful for studying and understanding our climate system, interactions among different components of the climate system, the effects of human activities...

Statistics for Ice Sheets



Satellite Observations versus Model Output



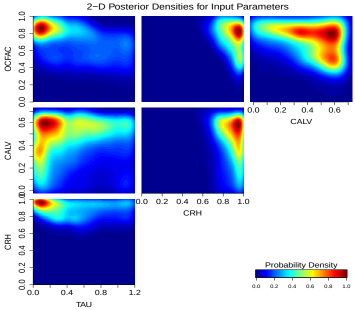
Statistical Methods for This Problem

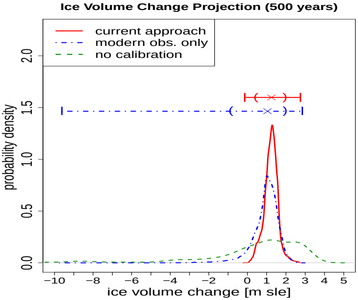
- Remember that we do not want a single "best"' dial (parameter) value. We want to end up with a *distribution* for the values, that is, which values are more likely and which ones are less likely
- For this we use **Bayesian inference**
  prior (parameter distribution) before we see data
  $\times$ **statistical model** (data distribution given parameters)
  = **posterior** (parameters distribution given data) after we see data

Computer Models are (Usually) Slow

- Even on fast computers, it can take hours or days to run the model at even just a few parameter settings.
- Makes it difficult to study how the model behaves for many different settings
- Statistical cleverness to the rescue: construct a statistical approximation called an emulator based on running the original model a relatively small number of times
- Use the fast emulator to study how the model behaves for many more parameter settings

After Lots of Modeling, Programming, Analyzing...





Sensible Science and Decisions Under Uncertainty

- After all this work, are we sure about our projections?
- Of course not!! Ice sheets (the physical world in general) are extremely complex and we cannot know the future exactly
- Projections are made under many assumptions about the model, the data, and future scenarios.
- However:
  - Uncertainty is not the same as not knowing
  - Uncertainty is not a reason for inaction
  - We can make sensible decisions even if we are uncertain
  - Describing uncertainties carefully is central to the scientific enterprise

Statistics is central to science

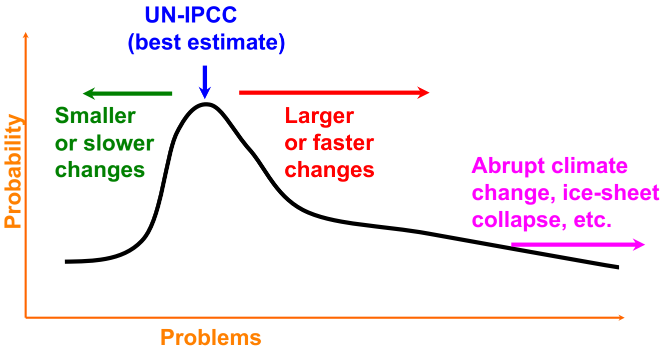
The Last Slide

- With thanks to Richard Alley, Dave Pollard, Dave Hunter, Klaus Keller, Ben Lee, Frances Blanchette

The Last Slide

- With thanks to Richard Alley, Dave Pollard, Dave Hunter, Klaus Keller, Ben Lee, Frances Blanchette
- Questions?

Probability of Various Levels of Future Problems



Interpretation of probability of various levels of future problems

Courtesy Richard Alley