Introduction to Causal Discovery

ATS 681A6, Department of Atmospheric Science
11:00 – 11:50 Tuesday and Thursday, 101 ATS main
2021 Fall Term

Instructor Contact Information

Prof Peter Jan van Leeuwen
Peter.vanleeuwen@colostate.edu
224 ATS West
https://www.atmos.colostate.edu/people/faculty/peter-jan-van-leeuwen/
Office hours: By appointments, just send me an email!

Course Description

Uncovering cause and effect relations is at the very heart of scientific discovery. We often apply causal discovery techniques unconsciously when we analyze an observational data set or when we perform sensitivity experiments with numerical models. However, when the number of variables is large, as is typically the case in atmospheric and oceanographic problems, we quickly loose track, and this course provides techniques to discover cause-and-effect relations in those situations.

After a short treatment of the necessary basic probability and statistics we dive into the elegant and easy to understand description of causal relations via graphical models. Then we discuss the powerful do-framework by Pearl and co-workers that formed the basis of many breakthroughs in as many scientific disciplines. This is followed by what to do if do-calculus cannot be applied, as is the case for many problems in the atmospheric and oceanographic realm. We then discuss a major topic in climate adaptation and mitigation, but also many other topics, counterfactual reasoning, trying to infer what would have happened if we would have intervened in the system. We also discuss how to build physical models from causal discovery, its interaction with machine learning and with dynamical systems theory, leading to some quite interesting and perhaps unexpected insights.

Course Goals

Upon successful completion, students will be able to:

- Explain what causal discovery is and how it can be used to infer cause and effect relations in complex systems
- Analyze numerical models and observation time series for causal relations
- Compare causal results with results obtained by other methods to infer relations between processes
- Compute quantitative causal strengths between different processes in a system, including joint contributions from two or more processes to a target process.
- Design a causal discovery analysis plan for a geoscience application, apply it, and draw science-based conclusions

Course materials

Detailed lecture notes will be available on Canvas, Google Drive, or by email, depending on students accessibility. The instructor does not use a specific textbook. The following recent textbooks provide basic and advanced material that relate to the course:

Pearl and Mackenzie (2018) The book of Why, Penguin Random House UK. *A low-level introduction to causal discovery for the general public.*

Pearl, Glymour, and Jewell (2016) Causal Inference in Statistics: A Primer, Wiley. *A thorough introduction to causal discovery, this book contains about 2/3 of the course material.*

Spirtes, Glymour, and Scheines (2000) Causation, Prediction, and Search, The MIT Press. *A thorough formal treatment of causal discovery, contains about half of the material of this course.*

Peters, Janzing, and Schoelkopf (2017) Elements of Causal Inference: Foundations and Learning Algorithms. *A thorough formal treatment of causal discovery with emphasis of learning causal models from data. It contains a small part of the material of this course.*

Class Participation

Students are expected to attend all classes. Students' participation and engagement are strongly encouraged. All interactions and discussions in the classroom are aimed to provide a supportive and active learning environment for students.

Grading

Grading will be based on 5 assignments (15% each) and a project where the student will apply causal reasoning on his/her own research data set (25%). Assignments will be due at the date and times indicated. No late assignments will be accepted without prior approval. Audits are strongly encouraged to do all assignments.

Statement on Academic Integrity

This course will adhere to the CSU Academic Integrity Policy as found in the General Catalog (http://www.catalog.colostate.edu/FrontPDF/1.6POLICIES1112f.pdf) and the Student Conduct Code (http://www.conflictresolution.colostate.edu/conduct-code). At a minimum, violations will result in a grading penalty in this course and a report to the Office of Conflict Resolution and Student Conduct Services.

Disclaimer

The instructor reserves the right to make modifications to this information throughout the semester.

Preliminary Schedule of Topics, Readings, and Assignments

Lecture	Topics	Deadlines
1	• Logistics • **Session 1: What is causal discovery** and its importance in scientific discovery	
2	• **Session 2: Preliminaries:** probability and statistics, and the difference between statistical and causal reasoning	
3	Session 2 (continued)	
4	• **Session 3: Graphical models:** structure, d-separation, directed acyclic graphs and hypergraphs	
5	Section 3 (continued)	
6	• **Session 4: Effects of interventions I:** adjustment and back-door criterion	Assignment #1 due
7	Session 4 (continued)	
8	• **Session 5: Effects of interventions II:** front-door criterion and atmospheric and oceanographic examples	
9	Session 5 (continued)	
10	• **Session 6: Confounders**	Assignment #2 due
11	Session 6 (continued)	
12	• **Session 7: Non-intervenable systems I:** Wiener and Granger and Transfer Entropy	
13	Session 7 (continued)	
14	• **Session 8: Non-intervenable systems II:** Information theoretic decompositions and confounders	Assignment #3 due
15	Session 8 (continued)	
16	Section 8 (continued)	
17	Section 8 (continued)	

Lecture	Topics	Deadlines
18	• **Session 9: Counterfactual reasoning I:** deterministic and stochastic	Assignment #4 due
19	Session 9 (continued)	
20	• **Session 10: Counterfactual reasoning II:** IPCC, geo-engineering, and other examples from ocean and atmosphere	
21	Session 10 (continued)	
22	• **Session 11: Structural model building**	Assignment #5 due
23	Session 11 (continued)	
24	• **Session 12: Machine learning for causal discovery**	
25	Session 12 (continued)	
26	• **Session 13: Dynamical systems theory for causal discovery**	
27	Session 13 (continued)	
28	• **Session 14: Project presentations**	Project due