Machine Learning and Climate

Spring 2025

Instructor

Alp Kucukelbir <alp@cs.columbia.edu>

Day & time

Mondays 4:10pm – 6:00pm

Location

TBD

Course assistant(s)

(to be approved)

- Nicolas Beltran <nb2838@columbia.edu>

Description

In this course, we will study two aspects of how machine learning (ML) interacts with Earth's climate.

First, we will establish how ML can monitor climate change. We will study how ML can be used to monitor greenhouse gas emissions, forecast how nature is changing, and improve our understanding of climate systems. We will ask questions like: How does monitoring differ from forecasting? What kind of machine learning techniques allow us to draw causal insights?

Second, we will investigate how ML can mitigate climate change. We will focus on industries with large carbon footprints such as power systems, manufacturing, and buildings. We will ask questions like: What are the requirements for applying ML to such problems? How can we evaluate the effectiveness of ML in these domains?

Last, we will consider ML's own impact on the climate. We will focus on the energy and computation that goes into designing, training, and deploying modern ML systems. We will ask questions like: How can we accurately track and account for ML's own energy footprint? What strategies can we employ to minimize it?

By the end of this course, you will learn about modern statistical, generative, and causal ML methods and their applications to the climate. Our focus will be the modeling of real-world phenomena using probability models, with a focus on image processing, time series forecasting, uncertainty quantification, and causality. In addition, you will gain a deeper understanding about the carbon footprint of ML itself and explore how to minimize it.

Prerequisites

This is a graduate-level COMS 6998 course. The class is open to select technical senior undergraduate, masters, and doctoral students; no auditors and no pass/fail.

You **must** be familiar with machine learning and statistics (for example, you took a class where you learned how to do data analysis using modern ML methods). You will be conducting independent data analysis in this course; as such, you **must** be comfortable programming in `python`, working with `git`, and writing in `LaTeX`.

You **should** have a good climate-related dataset in hand. This may be a dataset published alongside a relevant paper, or a dataset that hasn't been used for ML research yet. If you do not have a dataset readily available, you should have a strategy for simulating data for a relevant use case.

Here are a few resources for climate-related projects:

- Climate Change AI
- Kasia Kulma's GitHub Repository
- Carbon Plan
- Carbon Monitor
- Björn Lütjens's GitHub Repository

Structure

This course is based around an project that you will summarize in a final technical paper. Based on the number of students taking the course for credit, you will work either in groups of two or three. You will be expected to present a relevant use case of either ML's application to the climate or a study of ML's own carbon footprint. Your final paper must contain some form of data analysis using ML and will be accompanied by a GitHub repository with `python` code. Each student within a group will receive an individual grade, corresponding to their involvement in the project.

Each class is split into two. In the first part, I will lecture on the topic of the week. In the second part, we will discuss the readings. The discussion will focus on key insights from the readings and how they related to your ongoing projects.

The last class is dedicated to project presentations. Each group will record a short video over 2 slides, summarizing the state of their project so far. This is an opportunity to get feedback and to see what your peers have been working on. The final paper will be due at the end of the semester.

Readings

There is no textbook for this course. For context and background, we will closely follow these sources:

- Sandalow, D. B., McCormick, C., Kucukelbir, A., Friedmann, J., Nagrani, T., Fan, Z., Halff, A. M., d'Aspremont, A., Glatt, R., Méndez Leal, E., & others. (2024). *Artificial*

Intelligence for Climate Change Mitigation Roadmap (Second Edition). ICEF Innovation Roadmap Project. https://transitiondigital.org/ai-climate-roadmap
- Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1–10.

Each week's session has additional readings that provide the foundation necessary for understanding and exploring the ML technique in consideration.

Course Grade

You are graded on completing weekly response to the readings and participation in the class. There are three problem sets. Your course grade will be calculated as follows.

Component	**Percentage**
Weekly responses to the readings	10 %
Participation in class	20 %
Problem sets	20 %
Final paper	50 %

Your weekly responses are graded solely on submitting on time. We will score your responses as −1 for "below expectations", 0 for "meets expectations", and +1 for "exceeds expectations". These scores are simply meant to help you calibrate your engagement with the material.

There are three problem sets for this class, each due three weeks apart. These problem sets assess how well you are following the topics discussed in the sessions; they comprise climate questions about the industries we are studying and technical questions about the ML methods applied to them.

Your final paper will be evaluated for its relevance to the course material, technical correctness, and writing quality. There is no expectation to have a positive result in your data analysis by the end of the class; it is perfectly acceptable to reach a negative conclusion (e.g., such and such technique is not as good as the state of the art in forecasting water usage) as a result of your exploration.

Your final paper should be at most 8 pages long and prepared with the course `LaTeX` template. This is inclusive of images, tables, and bibliography.

Final Project

Each group will work on a project, summarized in a final technical paper. You will showcase and document your work through a private GitHub repository.

Please organize your repository as follows:
- `abstract.md`
- `journal.md`

- `doc/`
- `src/`
- `etc/`

This repository will document your exploration and coding through the semester.

The file `abstract.md` simply contains an abstract of the project. At first, it is an aspirational abstract, one that describes the research program you want to complete. You will refine it through the semester.

The file `journal.md` is a diary of your progress. It contains dated entries with a description of what you are doing, what you found, what you are thinking, and so on. It is mainly a resource for you, but I will glance at it too (at the end of the semester). Please update and commit it at least once per week.

The `doc/` directory contains the LaTeX document that you are writing. We will provide a template for your final paper.

The `src/` directory contains the code you are writing. The data you are analyzing should live here too.

The `etc/` directory contains anything else — materials, notes, photos of whiteboards, and so on — that you want to keep track of.

There should be nothing else in the top level directory of your repository.

Commit often, at least every week to provide an update to your journal. You are graded on the quality of the project and the path that you took to get there.

Course outline begins on the following page

Course Outline

Each week's readings should take you no more than two hours. Suggested readings are suggestions — enjoy them either if you need more background or seek deeper context.

Week 1 | Introduction to climate change mitigation (CCM)

How can we think about mitigating climate change? How do we study climate change? A primer on relevant topics:

- Monitoring emissions
- Climate event attribution
- Earth system modeling
- Biodiversity forecasting

How do different economic industries contribute to climate change? A primer on industries with large carbon footprints:

- Power systems
- Food systems
- Manufacturing
- Transportation
- Buildings

How can machine learning (ML) mitigate these contributions? What are common themes of how ML can make a difference? What are the primary barriers and risks of deploying ML to mitigate climate change?

Required readings

- What is climate change and how can it be mitigated? → (Sandalow et al., 2024, Ch. 2)
- Aligning ML and CCM → (Kaack et al., 2022)

Suggested readings

- Overview on CCM → (Gates, 2021)

Week 2 | Monitoring emissions

Detection and monitoring of methane super-emitters using infrared photography and satellite data. Convolutional neural networks for detection of plume-like structures, and support vector classification to distinguish real methane plumes from retrieval artifacts.

Technical topic | Convolutional neural networks.

Logistics | Problem Set 1 out, due on Week 5.

Required readings

- Greenhouse gas monitoring → (Sandalow et al., 2024, Ch. 12)

- Infrared imagery → (Wang et al., 2020)
- Satellite paper → (Schuit et al., 2023)

Suggested readings

- Convolutional neural networks primer → (Vedaldi, 2019)

Week 3 | Biodiversity forecasting

Comparison of correlative and causal modeling approaches to predict current and future distributions of biodiversity. Projecting future distributions under climate change scenarios. An application to animals in geographies subject to climate change.

> **Technical topic** | Predictive vs. causal models.

Required readings

- Forecasting biodiversity → (Sirén et al., 2022)
- Counterfactual analysis → (Pearl et al., 2016, Ch. 1 and 2)

Suggested readings

- Primer on causal analysis → (Pearl & Mackenzie, 2018)

Week 4 | Climate event attribution

What causes heatwaves? Causal counterfactual theory and how it can be applied to the attribution of weather and climate-related events. Difference between necessary and sufficient causality, and how they can be quantified using probabilities. Illustration on a case study on the 2003 European heatwave.

> **Technical topic** | Counterfactual analysis using data.

Required readings

- Heatwave analysis paper → (Hannart et al., 2016)
- Counterfactual analysis → (Pearl et al., 2016, Ch. 3)

Suggested readings

- Primer on causal analysis → (Pearl & Mackenzie, 2018)

Week 5 | Earth system modeling

Challenges of applying causal inference methods to large-scale complex dynamical systems like the Earth system, where real experiments are often infeasible. Causal hypothesis testing, causal network analysis, exploratory causal driver detection, and causal evaluation of physical models.

Technical topic | Granger causality, structural causal discovery.

Logistics | Problem Set 1 due. Problem Set 2 out, due on Week 8.

Required readings

- Causal inference for Earth systems → (Runge et al., 2019)

Suggested readings

- Primer on causal analysis → (Pearl & Mackenzie, 2018)

Week 6 | Power systems

Power flow optimization. A neural network-based method for solving optimization problems with hard constraints. Application to a variety of convex and non-convex optimization tasks, including an AC optimal power flow problem.

Technical topic | ML approaches to non-convex optimization with constraints.

Required readings

- AC power flow paper → (Donti et al., 2021)
- ML for power systems → (Sandalow et al., 2024, Ch. 3)

Suggested readings

- ML applications within power systems → (Donti & Kolter, 2021)

Week 7 | Manufacturing

ML applications to enable material circularity by adapting to volatility better. An application of Bayesian optimization to reduce the number of experiments. Illustration of a closed-loop optimization to efficiently optimize fast-charging protocols for lithium-ion batteries.

Technical topic | Bayesian optimization.

Required readings

- ML for manufacturing → (Sandalow et al., 2024, Ch. 5)
- ML battery charging protocols → (Attia et al., 2020)

Suggested readings

- Primer on Bayesian optimization → (Shahriari et al., 2015)
- Textbook on Bayesian optimization → (Garnett, 2023)

Week 8 | Advanced manufacturing

Bayesian optimization for tackling complex, expensive-to-evaluate applications to mitigate climate change. Three key application in advanced manufacturing: material discovery, wind farm layout, and optimal renewable control. Importance of benchmarking data and what it takes to advance the field of research in new domains.

Technical topic | Bayesian optimization.

Logistics | Problem Set 2 due. Problem Set 3 out, due on Week 11.

Required readings

- Bayesian optimization for climate application → (Hellan et al., 2023)

Suggested readings

- Primer on Bayesian optimization → (Shahriari et al., 2015)
- Textbook on Bayesian optimization → (Garnett, 2023)

Week 9 | Buildings

ML applications to the life-cycle of buildings, from design to construction, from operations to demolition. Simulation-based reinforcement learning for optimization of energy and emissions in office buildings.

Technical topic | Reinforcement learning.

Required readings

- ML for buildings → (Sandalow et al., 2024, Ch. 8)
- ML in energy efficient buildings → (Goldfeder & Sipple, 2024)

Suggested readings

- Reinforcement learning primer → (Lei, 2021)

Week 10 | Computational material science

Diffusion models for reconstruction of real-world microstructure data. Additional ideas around 2D-to-3D reconstruction and applications to multiscale modeling and structure-property linkages.

Technical topic | Diffusion models.

Required readings

- ML for material science → (Sandalow et al., 2024, Ch. 13)

- Diffusion for new materials → (Düreth et al., 2023)

Suggested readings

- A primer on diffusion models → (Luo, 2022)

Week 11 | The carbon footprint of ML

What is the carbon emissions associated with training machine learning (ML) models? How does the energy source used to power the hardware for training affect emissions? Does higher emissions lead to better model performance? How do we improve transparency and accountability in the field?

Technical topic | Carbon footprint quantification.

Logistics | Problem Set 3 due.

Required readings

- Carbon footprint of ML → (Strubell et al., 2020)
- Carbon footprint of ML → (Lacoste et al., 2019)
- Carbon footprint of ML → (Henderson et al., 2020)
- Carbon footprint of ML → (Luccioni & Hernandez-Garcia, 2023)

Suggested readings

- Cabon footprint of AI → (Sandalow et al., 2024, Ch. 15)

Week 12 | Can ML make climate change worse?

Why is ML progressing so quickly? Are there ways in which ML can make climate change worse? An illustration of risks across a broad set of applications, from manufacturing to disinformation campaigns. A particular focus on large language model usage risks.

Technical topic | Large language models.

Required readings

- Why ML is progessing so quickly? → (Donoho, 2024)
- Can ML make climate change worse? → (Kaack et al., 2022)
- Risks → (Sandalow et al., 2024, Ch. 16.C)
- LLMs → (Sandalow et al., 2024, Ch. 11)

Week 13 | Final presentations

Projects presentations and discussion.

Additional recommended readings

- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., & others. (2019). Tackling climate change with machine learning. *Arxiv Preprint Arxiv:1906.05433*.
- Gentine, P., List, G., Thompson, K., Pardo, T., Li, X., Berg, G., & Bennett, L. (2024). *AI for Climate and Nature: Landscape Assessment*.

Bibliography

Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., Chen, M. H., Cheong, B., Perkins, N., Yang, Z., & others. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, *578*(7795), 397–402.

Donoho, D. (2024). Data science at the singularity. *Harvard Data Science Review*, *6*(1).

Donti, P. L., & Kolter, J. Z. (2021). Machine Learning for Sustainable Energy Systems. *Annual Review of Environment and Resources*, *46*, 719–747.

Donti, P. L., Rolnick, D., & Kolter, J. Z. (2021). DC3: A learning method for optimization with hard constraints. *Arxiv Preprint Arxiv:2104.12225*.

Düreth, C., Seibert, P., Rücker, D., Handford, S., Kästner, M., & Gude, M. (2023). Conditional diffusion-based microstructure reconstruction. *Materials Today Communications*, *35*, 105608.

Garnett, R. (2023). *Bayesian Optimization*. Cambridge University Press.

Gates, B. (2021). *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need*. Knopf Doubleday Publishing Group.

Gentine, P., List, G., Thompson, K., Pardo, T., Li, X., Berg, G., & Bennett, L. (2024). *AI for Climate and Nature: Landscape Assessment*.

Goldfeder, J., & Sipple, J. (2024). *A Calibrated Simulation for Offline Training of Reinforcement Learning Agents to Optimize Energy and Emission in Office Buildings*.

Hannart, A., Pearl, J., Otto, F., Naveau, P., & Ghil, M. (2016). Causal counterfactual theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society*, *97*(1), 99–110.

Hellan, S. P., Lucas, C. G., & Goddard, N. H. (2023). Bayesian Optimisation Against Climate Change: Applications and Benchmarks. *Arxiv Preprint Arxiv:2306.04343*.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, *21*(248), 1–43.

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, 1–10.

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. *Arxiv Preprint Arxiv:1910.09700*.

Lei, C. (2021). Deep reinforcement learning. In *Deep Learning and Practice with MindSpore: Deep Learning and Practice with MindSpore* (pp. 217–243). Springer.

Luccioni, A. S., & Hernandez-Garcia, A. (2023). Counting carbon: A survey of factors influencing the emissions of machine learning. *Arxiv Preprint Arxiv:2302.08476*.

Luo, C. (2022). Understanding diffusion models: A unified perspective. *Arxiv Preprint Arxiv:2208.11970*.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., & others. (2019). Tackling climate change with machine learning. *Arxiv Preprint Arxiv:1906.05433*.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Mar\i, J., & others. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*(1), 1–13.

Sandalow, D. B., McCormick, C., Kucukelbir, A., Friedmann, J., Nagrani, T., Fan, Z., Halff, A. M., d'Aspremont, A., Glatt, R., Méndez Leal, E., & others. (2024). *Artificial Intelligence for Climate Change Mitigation Roadmap (Second Edition)*. ICEF Innovation Roadmap Project. https://transitiondigital.org/ai-climate-roadmap

Schuit, B. J., Maasakkers, J. D., Bijl, P., Mahapatra, G., Berg, A.-W. Van den, Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., Varon, D. J., & others. (2023). Automated detection and monitoring of methane super-emitters using satellite data. *Atmospheric Chemistry and Physics Discussions*, *2023*, 1–47.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148–175.

Sirén, A. P., Sutherland, C. S., Karmalkar, A. V., Duveneck, M. J., & Morelli, T. L. (2022). Forecasting species distributions: Correlation does not equal causation. *Diversity and Distributions*, *28*(4), 756–769.

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(9), 13693–13696.

Vedaldi, A. (2019). *A convolutional neural network primer*. https://www.robots.ox.ac.uk/~vedaldi/assets/teach/2022/c18-notes.pdf

Wang, J., Tchapmi, L. P., Ravikumar, A. P., McGuire, M., Bell, C. S., Zimmerle, D., Savarese, S., & Brandt, A. R. (2020). Machine vision for natural gas methane emissions detection using an infrared camera. *Applied Energy*, *257*, 113998.