

**Summary Memo – Examples of Berkeley Faculty/Staff Using AI and ML for Climate Research**

**Compiled by Bruce Riordan, BCCN Director**

**September 7, 2023**

**PART ONE: Email responses to a query about AI and ML sent to the BCCN members.**

1. **Marianne Cowherd, PhD Candidate, ESPM**

Snowpack measurement from a range of data sources

I'm leading a relevant project right now with team members mostly at Berkeley.

A network of around 900 snow pillows – *in-situ* snow mass measurement instruments – located at snow telemetry (SNOTEL) stations critically inform operations by providing foundational information on Western United States (WUS) snow-water equivalent (SWE). In the 21st Century, warmer temperatures and changing atmospheric circulation will produce unprecedented changes in SWE, likely resulting in alterations in the timing, amount, and spatial patterns of snowpack across many catchments. In the future, the WUS SNOTEL network will be less sensitive to peak SWE. In this work, we explore a parameter space of data and model complexity to produce suggested frameworks for snowpack prediction in a warmer world. We explore the relative importance of model complexity and observational information for constraining SWE estimates in downscaled climate models over the WUS. We show that sufficiently complex data-driven models are likely to maintain SWE estimation skill even in future climate with 1) increased interannual variability and 2) ahistorical spatial and temporal correlations between predictors of peak SWE and the peak SWE field. In this framework, increased model complexity and additional observational datasets can compensate for loss of sensitivity to SWE that the SNOTEL network will experience. Nimble artificial intelligence-based models that incorporate partial, multi-modal SWE information can address these challenges and help ensure that peak SWE estimation in the WUS will be resilient to the future no-analog snowpack conditions.

1. **Bakar Institute of Digital Materials for the Planet (BIDMaP)**

BIDMaP Overview

BIDMaP aims to speed up the development of reticular chemistry and modular structures for achieving cost-efficient, easily deployable ultra-porous metal-organic frameworks (MOFs) and covalent organic frameworks (COFs). These programs will help limit and address the impacts of climate change and extend to downstream technologies like conversion of CO2 to clean fuels, biodegradable polymers, enzymes, and pharmaceuticals. BIDMaP brings together top computation and machine learning experts with chemistry and other physical science researchers to exploit the vast potential these reticular structures have in achieving clean air, clean energy, and clean water.

[BIDMaP members utilize ChatGBT to drastically speed chemical discovery.](https://bidmap.berkeley.edu/news/bidmap-members-utilize-chatgpt-drastically-speed-chemical-discovery)

In a new paper published in ACS Editors' Choice, ChatGPT was trained by BIDMaP faculty and students via precise prompt engineering to efficiently text mine the academic literature on metal-organic frameworks (MOFS). The resulting "ChatGPT Chemistry Assistant" successfully produced highly accurate synthesis condition predictions for over 800 MOFs. Congratulations to Zhileng Zheng, Oufan Zhang, Christian Borgs, Jennifer T. Chayes, and Omar Yaghi for this groundbreaking collaboration!

1. **Corinne Scown, LBNL, Energy Analysis & Environmental Impacts Division**

We have used machine learning to enhance our predictive capabilities for a range of natural and engineered systems relevant to the production of biofuels and biomaterials. Our first project was focused on agricultural systems. We used machine learning to predict how crop yields (in this case, sorghum) could change with climate change across the US by training a model on historical yields and climate data, then using CMIP5 climate data to extrapolate out into the future: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bbb.2087>.

Another example of using real-world data to train machine learning models came out of a collaboration with EBMUD, our local water utility. We were able to work with them through an umbrella NDA negotiated by NAWI leaders and, as a result, they shared operating data on their high-strength waste co-digestion program over its lifetime, along with characterization of individual waste streams that were taken in (and when). We used machine learning to predict biogas yields from different types of organic waste, enabling our local wastewater treatment plant to divert those wastes from methane-emitting landfills: <https://pubs.acs.org/doi/full/10.1021/acssuschemeng.1c04612>. Lastly, we wanted to develop machine learning models for facility designs that have not yet been built, so we used process simulation software to generate synthetic training datasets and then used auto-ML to produce surrogate models that predict cost, energy balances, and mass balances <https://doi.org/10.1016/j.biortech.2022.128528>.

1. **Alex Bayen, Engineering, LBNL and Moffett Field**

Traffic simulations for transportation energy reductions

An interdisciplinary team of industry and academic researchers led by Professor Alexandre Bayen has completed its most ambitious [real-time traffic experiment](https://news.berkeley.edu/2022/11/22/massive-traffic-experiment-pits-machine-learning-against-phantom-jams/) to date. The project was led by the CIRCLES Consortium, an effort led by UC Berkeley and Vanderbilt University, involving collaborators from five universities and multiple government agencies. The experiment tested 100 partially automated vehicles in real traffic with the aim of improving overall traffic flow. Operating out of a massive control center designed to monitor one section of I-14 in Nashville, TN, the researchers used AI to build on existing adaptive cruise control systems to smooth phantom jams collaboratively. Their results show a positive energy impact that could reduce the energy footprint of traffic congestion by up to 10%. “Driving is very intuitive. If there’s a gap in front of you, you accelerate. If someone brakes, you slow down. But it turns out that this very normal reaction can lead to stop-and-go traffic and energy inefficiency,” said Prof. Bayen. “That’s precisely what AI technology is able to fix—it can direct the vehicle to things that are not intuitive to humans, but are overall more efficient.” Short video [here](https://www.youtube.com/watch?v=PA3lyoCZnP0&embeds_referring_euri=https%3A%2F%2Fnews.berkeley.edu%2F&source_ve_path=Mjg2NjY&feature=emb_logo).

1. **Ritwik Gupta, Co-Director, Berkeley AI Research Climate Initiative**

BAIR Climate Overview

The [Berkeley AI Research Climate Initiative](https://www.google.com/url?q=http://ai-climate.berkeley.edu&sa=D&source=editors&ust=1694110960245377&usg=AOvVaw0UrAfXIJMpkUCRzTet1GGW) (BCI) unites AI and climate-related researchers to create and maintain meaningful datasets, benchmarks, and methodologies that bridge these communities. By bringing together leading AI researchers, industry partners, and scientists working on climate change, we seek to foster a community in which fundamental AI advancements are directly motivated by key problems related to climate change and the many problems it creates. BCI collaborations follow an iterative, interdisciplinary practice:

1. Ideate: Form a diverse group of domain experts, industry partners, and AI researchers
2. Connect: Collaborators establish a benchmark, where improvements on this benchmark directly translate to impact on real world problems, while necessarily incentivizing core AI research advancements.
3. Progress: Collaborators research, share, and publish advances on the established benchmark.
4. Revisit: Establish criteria and schedule to update and assess benchmarks and core AI research contributions.

Current Projects outlined [here](https://docs.google.com/document/u/1/d/e/2PACX-1vRUDu8PpYZLkIbyyxgmjyr8U8kMQaHysq8IbdOOracDs3nVdwEvKlwt9VPb05HOBubknPLCO4QKX4E-/pub).

* The Fate of Snow
* AutoFuse: Automatic Fusion for Multimodal Spatiotemporal Data
* Benchmark Dataset: Harmful Algal Blooms
* Animal Re-Identification in Camera Traps
* NumS

Also: We are starting a project this week with Climate Rights International (<https://cri.org/>) on monitoring deforestation caused by nickel mines.

Potential Projects outlined [here](https://docs.google.com/document/u/1/d/e/2PACX-1vRUDu8PpYZLkIbyyxgmjyr8U8kMQaHysq8IbdOOracDs3nVdwEvKlwt9VPb05HOBubknPLCO4QKX4E-/pub).

* Self-Supervised Geometric Deep Learning for Materials Design
* Scientifically constrained machine learning for molecular dynamics simulations
* SCALE: Scale-Aware Multi-task Representations
* Position Paper: We Can Use Meaningful Benchmarks to Progress AI
* Interpretable Multimodal Uncertainty Attribution with Deep Neural Networks
* Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response
* Continuous Representation Learning for Multiscale Geospatial Data
* Deep Learning for Smart Buildings
1. **Alice Agogino, Engineering**

Squishy Robotics for climate

Alice Agogino's lab and spin-off Squishy Robotics have developed a solution to enable the rapid deployment and operation of diverse sensor payloads in remote or dangerous locations with a focus on hazardous material applications.  Our work in public safety is dedicated to providing public safety through improved situation awareness upon demand. Our recent research focuses on reducing climate change through early detection of methane leaks. Methane is the second most common greenhouse gas, accounting for approximately 20 percent of global emissions. In the United States, the oil and gas industry is the largest source of methane. This research has potential applications in early wildfire detection as well.

Capable of being dropped from Small Unmanned Aircraft Systems (sUAS) such as drones as well as other aircraft, these payloads can be rapidly deployed to provide monitoring and sensing that enables faster, better informed, data-driven decisions.  The unique pairing of these ground robots and aerial vehicles ensures rapid deployment to sites of interest and persistent monitoring beyond the battery lifetime of the drone to successfully monitor the selected sites for air deployment..  The robot's payloads can also be customizable to carry a range of sensors and communication devices.

Our lightweight, shape-shifting tensegrity robots are designed to be rapidly deployed by aerial vehicles or thrown to provide rapid situational awareness to first responders. Emergency response and industrial gas monitoring need precise information in a short time period, requiring increased speed in visual and sensor measurements. We are developing novel  sensor technologies for orienting and fusing data from a new generation of methane  sensors carried in a payload on our tensegrity robots originally developed for space exploration.

1. **Adam Arkin, Engineering**

Microbes and the “Knowledge Engine”

My [ArkinLab](https://Arkinlab.bio) studies how microbes and their communities assemble and reciprocally transform the environments and hosts they inhabit and we might predict, control and design their behaviors. Towards that end I run a number of programs aimed at predictive understanding of biological function and activity from genes to ecologies spanning communities driving nutrient cycles in the terrestrial subsurface (see [ENIGMA](https://enigma.lbl.gov)), through understanding the mechanisms and dynamics of viruses of bacteria that we might use to control rising multidrug resistant bacteria, to the design and implementation of integrated sustainable biomanufacturing processes to produce food, pharmaceuticals and building materials 'autotrophically' to (conceptually) support a mission of several astronauts on an extended mission to Mars (see [CUBES](https://cubes.space)).

Each of these programs develop large amounts of genetic, physiological and ecological data that we need to integrate across scales to make predictive models. Driven by the need of programs like these and the recognition that to make these models generalizable requires scientists to be open and work together to share their data, analytical tools, models and results, I also use the Department of Energy's Systems Biology Knowledgebase ([KBase](https://kbase.us)) a flexible open platform that supports diverse communities to integrate, harmonize and model data about microbes, plants and their communities and which, in a relatively few 'clicks' can go from metagenomics sequencing reads to 'causal' models of community function. However, all of these integrate a fair amount of sophisticated statistics and machine learning approaches to bridge the gap between mechanistic knowledge and discovery of new mechanisms and prediction of function.

We are currently building something we call the 'Knowledge Engine' which is designed to support machine learning and artificial intelligence approaches to support research across these large, diverse data sets derived across disparate projects.

1. **Dan Kammen, Energy and Resources Group**

Rooftop solar and justice

We use AI, technically machine learning, in a number of large-scale projects. The biggest is the use of MI to identify solar cells on rooftops on a national sample of all (~ 60 million) rooftops. The partner is [Google's Project Sunroof](https://sunroof.withgoogle.com/) and the core paper (there are several smaller follow-up ones) is [here](https://rael.berkeley.edu/publication/disparities-in-rooftop-photovoltaics-deployment-in-the-united-states-by-race-and-ethnicity/sunter-castellanos-kammen-solar-disparities-by-race-supplementalonlinematerial/), and is part of the Justice40 federal process.

1. **Yang Zhou, LBNL, Climate and Ecosystem Sciences Division**

Atmospheric River Events

I am working on a project with Michael Wehner on the temporal clustering of atmospheric river events — U.S. West Coast Atmospheric River Clusters and Their Key Circulation Patterns (*In submission*)

We applied an unsupervised machine learning clustering technique on the temporal series of landfalling atmospheric rivers (ARs) to identify temporal AR clusters as a short period of back-to-back ARs. We show that the characteristics and impacts vary significantly by cluster density, which is the fraction of AR conditions within a cluster. Focusing on the landfalling ARs over the U.S. West Coast, we found that clusters with high density (such as the 2023 series of nine ARs in three weeks) consist of higher AR categories and higher likelihood for extreme precipitation and severe land surface response. Using reanalysis and model simulation, we showed that the key circulation pattern for AR clusters is mainly attributed to subseasonal variability. Furthermore, we demonstrated that AR clusters with higher density and category will be more frequent in warming climates. Our study highlights the critical role of AR clusters in the planning and development of climate adaptation and resilience.

1. **Jillian Grennan, Berkeley Law**

Proposal: Tackling Climate Change with AI — Wildfires

I have been working on this proposal with researchers from other universities.

Practical Problem: Policymakers and communities do not trust prescribed burns. Even if it is not fire specifically, this could come from utilities shutting down post burn, etc.

Use-Inspired Research Question: To what extent can we build resilience into California’s energy and environmental ecosystems by combining counterfactual burns models, better data and predictions, and especially, ways for incorporating trust and cultural-sensitivity (e.g., through explainable or ethical AI, and/or generative AI for creating educational content and helping to fill out forms for insurance and environmental risk offsets)?

Novel Research Approach: combining energy/infrastructure modeling with new AI tools and a insights from social sciences for gaining trust and greater community buy-in.

1. **Ron Cohen, CDSS, EPS and Chemistry**

We are using AI/ML tools to synthesize models and observations.

[Efficient approaches to interpreting remote sensing observations.](https://spj.science.org/doi/10.34133/2022/9817134)

The two below are analyses of observations that provide insight into decadal trends in urban oxidation rates--those rates govern aerosol and ozone.

[Surface OH](https://pubs.acs.org/doi/10.1021/acs.est.1c05636)

[OH Trends](https://www.pnas.org/doi/10.1073/pnas.2117399119)

Also: The Berkeley Atmospheric Science Center Symposium on March 7 and 8 will be on the topic of AI/ML in climate and earth system prediction, modeling and analysis.

1. **Ali Mesbah, Chemistry**

Plasma-Biowaste

We leverage AI/ML tools towards low temperature plasma-assisted electrification of chemical synthesis. Our work mainly focuses on plasma catalysis and plasma-biowaste processing for ammonia synthesis and organic fertilizer manufacture, respectively. We use AI/ML, along with computational modeling tools and experiments, to improve our fundamental understanding of plasma interactions with complex interfaces in these processes, as well as to enable optimal design and operations of plasma-assisted processed.

1. **Daniel Rodriguez, CED**

Urban Heat

We’ve used AI/Machine Learning for two purposes:

* Impute temperatures at a disaggregate level (~5km) for coastal areas of most of Latin America.
* Classify satellite imagery to identify greenspaces over time in Latin America. Created a longitudinal database of urban greenspaces for 300 plus cities (n> 40 years), and have written a couple of descriptive papers on greenspaces in Latin American cities.
1. **Aditi Krishnapriyan, Engineering and Chemistry**

Extreme weather events

Our research group works on developing new ML methods for the sciences, including problems ranging from spatiotemporal forecasting to molecular generation. Some areas we are looking at include:

* Generative ML models for molecular design, and creating new molecules and materials
with favorable renewable energy properties. For example, [here](https://streaklinks.com/BqOtUCaODgVPC1YptQ4b0HRs/https%3A//arxiv.org/abs/2306.14852).
* ML methods to capture spatiotemporal forecasting over longer time scales (e.g., Navier-
Stokes equations), including dynamical climate events. For example, [here](https://streaklinks.com/BqOtUCWkZFXvH7VhIgy4uGBg/https%3A//arxiv.org/abs/2207.08675).
* ML methods for extreme weather prediction
1. **Bill Collins, LBNL and EPS**

Extreme Events

We are generating huge ensembles of simulations of low-likelihood high-impact climate extremes using NVIDIA's FourCastNet climate-model emulator. (No link just yet -- we don't want to go very public until the first papers are submitted.)

1. **Manuela Girotto, ESPM**

Snowpack and water

Girotto is advisor to [“The Fate of Snow”](https://data.berkeley.edu/news/new-uc-berkeley-initiative-uses-ai-research-solve-climate-problems) – a partnership between the BAIR climate initiative, Lawrence Berkeley Berkeley Lab, Meta AI and the [Center for Western Weather and Water Extremes](https://cw3e.ucsd.edu/) using AI techniques to combine measurements from [aircraft observations of snow](https://www.airbornesnowobservatories.com/)

 and a multitude of openly available weather and satellite data sources.

**PART TWO: Intel on AI and ML use for climate from other sources**

These names came from an AI and ML search on our [BCCN Climate Map](https://bccn.berkeley.edu/).

Laurel Larsen is using ML on hydrology research

Julianne Miller - watershed decision-support

Haruko Wainwright - detailed ecosystem models

David Anthoff - economic impact of climate change on ag, health, energy

Max Auffhammer - economic damages from climate change

Trevor Kennan - carbon cycling

Whendee Silver - carbon sequestration on working lands

David Zilberman - economics of renewable energy

James Ben Brown - economic impacts of CC on agriculture

Jeff Long and Joseph Gonzalez — ML for MOF development

Many of the above came from this excellent talk by Jennifer C in 2020.

[**Tackling Climate Change with Machine Learning**](https://slideslive.com/38947095/tackling-climate-change-with-ml?time=37935s)**(Chayes 2020 conference video keynote) shows examples of Berkeley/LBNL climate**